

2026년 연구보고서

국가통계 개념의 체계적 관리를 위한 GSIM 활용 방안 연구

2026. 9.



국가데이터처
국가데이터연구원

연구보고서 2026-01

국가통계 개념의 체계적 관리를 위한 GSIM 활용 방안 연구

김능경 · 배미경



Ministry of Data and Statistics
Data and Statistics Research
Institute

발간사

국가데이터·통계는 사회와 경제의 변화를 정확히 이해하고 미래를 준비하기 위한 핵심 자산입니다. 최근 인공지능(AI)을 비롯한 디지털 기술의 발전과 데이터 활용 환경의 변화는 국가데이터·통계의 생산과 관리, 활용 방식에도 새로운 혁신을 요구하고 있습니다.

국가데이터연구원은 이러한 변화에 능동적으로 대응하여 국가데이터·통계의 정확성과 활용성을 높이고, 데이터에 기반한 정책 수립을 지원하기 위한 연구를 지속적으로 수행해 왔습니다.

2026년 연구보고서에는 국가데이터·통계의 체계적 관리와 인공지능(AI) 기반 통계 생산 혁신, 그리고 우리 사회와 경제의 주요 변화를 심층적으로 분석한 연구 성과를 담았습니다.

첫째, 국가데이터·통계를 체계적으로 관리하고 통계생산의 효율성을 높이기 위한 연구를 강화하였습니다. 국가데이터·통계에서 사용되는 개념과 정보를 일관성 있게 관리하기 위해 국제표준인 GSIM을 활용할 수 있는 방안을 연구하였으며, 또한, 사망 원인통계의 원사인 분류와 기업통계등록부 정비과정의 산업분류에 인공지능(AI) 기술 적용 가능성을 연구하여, 전문적인 판단과 많은 시간이 필요한 통계분류 업무의 정확성과 효율성을 높이기 위한 방안을 제시하였습니다.

둘째, 다양한 데이터를 연계·활용하여 청년층의 경제적 이동과 삶의 변화를 입체적으로 파악하고자 하였습니다. 통계등록부를 활용하여 청년층의 경제적 자립과 지역 이동의 특성을 이해할 수 있는 소득 변화와 이동성을 분석하였습니다. 이러한 연구는 청년정책과 지역정책을 보다 정교하게 설계하기 위한 기초자료로 활용될 수 있을 것입니다.

셋째, 국민의 삶과 지속가능한 발전을 종합적으로 측정하기 위한 연구를 수행하였습니다. 주요국의 삶의 질 지표와 지속가능발전목표(SDG) 지표체계를 비교·분석하여 각국의 지표 구성과 활용방식을 살펴보고, 우리나라 지표체계의 발전 방향과 정책적 활용도를 높이기 위한 시사점을 도출하였습니다. 이는 사회 발전을 경제적 성과뿐만 아니라 삶의 질과 지속가능성의 관점에서 폭넓게 진단하는 데 기여할 것으로 기대합니다.

이번 연구들은 국가데이터·통계의 생산과 관리에 새로운 기술과 국제표준을 적극적으로 접목하는 한편, 다양한 데이터를 활용하여 사회·경제 현상을 심층적으로 이해하려는 노력의 결과입니다.

국가데이터연구원은 앞으로도 최신 기술과 다양한 데이터를 적극 활용하여 국가데이터·통계의 전문성과 신뢰성을 높이고, 시대적 변화에 선제적으로 대응할 수 있는 연구를 이어가겠습니다.

본 연구보고서가 통계 생산자와 이용자 모두에게 실질적인 도움이 되고, 데이터에 기반한 정책 수립과 의사결정에 유용한 자료로 활용되기를 기대합니다.

많은 관심과 성원을 부탁드립니다.

2026년 9월

국가데이터연구원 연구기획실장



목 차

제1장 서론	1
제2장 GSIM에 대한 이해	4
제1절 개발 배경 및 다른 표준과의 관계	4
제2절 GSIM 이해를 위한 검토 자료	9
제3절 그룹별 정보 클래스 세부 내용	12
제4절 소결	30
제3장 해외 사례	32
제1절 캐나다 통계청	32
제2절 노르웨이 통계청	49
제3절 해외 사례 비교	59
제4장 국내 활용 방안 검토	66
제1절 개념 정보 관리 거버넌스	66
제2절 정보 클래스별 관리 방안	67
제3절 실무 적용 고려사항	77
제5장 결론 및 시사점	86
참고문헌	89
부 록	91
Abstract	132

요 약

적합한 통계 활용을 위해서는 통계가 어떤 대상과 기준, 방법으로 작성되었는지 설명하는 체계적 개념 정보가 필수적이다. 본 연구는 UNECE가 승인한 통계정보 모델인 GSIM(Generic Statistical Information Model)을 검토하고 캐나다와 노르웨이의 사례를 분석하여 국가통계 개념정보 관리체계 방안을 제시하였다.

GSIM은 통계 생산 전 과정을 포괄하며, 정보 클래스를 역할에 따라 개념, 구조, 비즈니스, 교환, 기초의 5개 그룹으로 구분한다. 특히 설계와 구현을 분리하는 변수 연쇄(Variable Cascade) 구조를 통해 공통 정보의 재사용성을 높이고 통계 간 일관성을 확보한다. 해외 국가의 GSIM 적용 방식에는 차이가 있었는데, 캐나다는 “표준에 관한 정책”을 통해 표준의 제정과 재사용을 강조하는 반면, 노르웨이는 변수, 통계분류 및 코드목록 등을 기능별 시스템에서 관리하고 최종 통계표 조회까지 연계하는 방식이 특징적이었다. 다만, Conceptual Variable(개념변수)에 대한 설명이 추상적인 개념에만 머무르지 않고 판정기준이나 도출방법과 같은 구체적인 내용을 포함하는 사례는 공통으로 확인할 수 있었다.

국내 적용을 위해서는 Unit Type, Variable(변수), Code List(코드목록)를 주요 관리 대상으로 설정하고, 변수를 표준 변수와 통계 프로그램별 변수로 이원화하여 관리할 것을 제안한다. 또한 표준 관리부서와 통계 프로그램 담당자의 역할을 구분하고, 표준의 제정·개정·승인 등의 관리 절차를 확립할 필요가 있다.

향후 다양한 통계로 실증 범위를 확대하는 한편, 통계 프로그램별 개념·변수·코드목록을 구조화하여 확보한 후 공통성과 활용성이 높은 정보부터 단계적으로 표준화하는 접근을 고려할 수 있다. GSIM 기반의 국가통계 개념 정보 관리체계가 안정적으로 정착되어 통계 간 비교가능성과 재사용성을 획기적으로 향상하고 통계·데이터의 연계·융합을 촉진하는 기반이 되기를 기대한다.

주요 용어: GSIM, Unit Type, Conceptual Variable, Code List

제 1 장

서 론

적합한 통계 활용을 위해서는 통계데이터가 갖는 고유한 의미와 맥락에 대한 명확한 이해가 선행되어야 한다. 수많은 통계자료 속에서 이용자의 목적에 가장 부합하는 데이터를 식별하고 올바르게 해석하기 위해서는, 해당 통계가 어떤 대상과 기준, 방법으로 작성되었는지 설명해 주는 체계적인 정보가 필수적이다. 이는 생성형 인공지능이 통계표를 오차 없이 읽고 정확히 추론하기 위한 기계 가독성 측면에서도 마찬가지로 요구된다.

동일하거나 유사한 명칭의 개념이라도 조사대상, 측정 단위 및 방법, 범주 구성 등에 따라 서로 다른 의미가 있을 수 있고, 유사한 개념이라도 변수(용어) 명칭은 서로 다를 수 있다. 예컨대, 국가통계포털에서 경제활동인구조사와 지역별고용조사는 ‘취업자’, 사업체노동력조사는 ‘종사자’, 일자리행정통계는 ‘일자리’라는 다른 항목명으로 통계표를 제공하고 있다. 모두 노동시장과 관련된 통계이지만 측정 대상과 개념에서는 차이가 있다.

<그림 1> 노동시장 규모 관련 통계표

경제활동인구조사	경제활동인구조사 (계절조정)	지역별고용조사	사업체노동력조사	일자리행정통계
2024.10 취업자 (천명)	2024.10 취업자 (천명)	2024.2/2 취업자 (천명)	2024.10 종사자_전체 (명)	2024 일자리 (만개)
28,847	28,613	28,847	20,332,990	2,671.4

출처: 국가데이터처, 국가통계포털 [웹사이트].

국내 노동시장 규모를 알고 싶은 통계 이용자는 관련 통계를 폭넓게 식별할 수 있어야 하며, 이들 통계 중에서 이용자의 이용 목적에 가장 부합한 것이 무엇인지 판단할 수 있어야 한다. 이는 통계 이용자에게만 국한되는 과제가 아니다. 신규 정책에 대응하여 통계를 새로 기획·작성하는 통계작성기관 역시 기존 통계와의 개념적 차이와 연계 가능성을 명확히 파악할 수 있어야 중복 생산을 방지하고 통계 간 비교·연계 및 융복합 활용성을 높일 수 있다.

<그림 2> 용어조회 서비스 및 용어사전 검색 사례 예시(검색어: '취업자')

조사용어 총 78 건 (취업자)(논산시청년통계 2025년) 조사 대상 기간에 소득, 이익, 봉급, 임금 등의 수입을 목적으로 1시간 이상 일한 자 (취업자)(충청남도논산시기본통계 2023년) 조사 대상 기간에 소득, 이익, 봉급, 임금 등의 수입을 목적으로 1시간 이상 일한 자 18시간미만 취업자 중 추가취업희망자(경제활동인구조사 2009년) 18시간 미만 취업자 중 경제적 이유(일거리가 없거나 사업부진)로 18시간 미만 일하였으면서 추가 취업을 희망하는 자 ◦ (취업자) (광양시청년통계 2017년) 조사대상 기간 중에 수입을 목적으로 1시간 이상 일한자와 18시간 이상 일한 무급가족 종사자, 일시 휴직자

출처: 국가데이터처, 통계설명자료서비스 용어조회; 통계용어사전 [웹사이트].

그러나 <그림 2>에서 보듯이 현행 통계설명자료 서비스가 제공하는 용어 조회 서비스는 별도의 상세 작성 기준 없이 개별 통계작성기관의 여건과 판단에 따라 작성된 용어 해설에 기반하고 있어서 통계 간의 공통점과 차이점을 체계적으로 식별하고 비교하는데 한계가 있다. 통계용어사전의 경우 국가 차원에서 정제된 정의와 해설을 통해 공통된 개념을 제공한다는 장점이 있으나, 개념이 적용되는 조사 단위유형 등에 대한 정보가 충분히 구조화되어 있지 않고, 관련 통계도 대표적인 통계 중심으로 제시되어 있어서 해당 개념을 사용하는 전체 통계와의 연계 분석에는 제약이 따른다.

김민규 외(2025)에서는 동일 현상을 동일 의미로 관측·해석·비교할 수 있도록 변수 개념을 공통 기준에 맞추어 변환·표준화하는 개념 조화(semantic harmonization)의 중요성을 강조한다. 실례로 포르투갈 은행에서는 400개 이상의 변수와 7,000개 이상의 코드를 통합해 조화된 코드목록 데이터 사전을 구축하여, 개별 개념을 표준 개념으로 매핑하고, 표준화된 코드 기반의 집계 테이블 생성까지 데이터의 개념적 일관성을 확보하는 작업이 상당 부분 진행되었다.

UNECE에서는 통계 생산 과정의 정보 관리의 표준화를 위해 GSIM(Generic Statistical Information Model)을 개발하였고, 해외 주요 통계기관에서도 이를 도입하고 있다. 본 연구는 국가통계 전반에 적용할 수 있는 통계 개념의 정의와 기술 체계에 대한 공통 기준 마련을 목표로, 국제표준인 GSIM 모델과 이를 도입한 해외 사례를 면밀하게 분석하여 국내 통계 환경에 적합한 활용 방안을 모색한다.

제 2 장

GSIM에 대한 이해

제1절 개발 배경 및 다른 표준과의 관계

1. 개발 배경

GSIM은 통계기관 내외부 의사소통 증진 및 표준화 구성요소의 재사용 지원을 목적으로 개발되었다. 전 세계 통계기관들이 통계분류체계 적용, 데이터세트 구축, 보고서 발간 등 유사한 활동을 수행하며 그 과정에서 생산·활용되는 정보 또한 본질적으로 동일한데, 이러한 정보를 정의하는 표준체계의 부재¹⁾하여 국가나 기관 내·외부 간 유기적인 협업과 정보 공유가 저해된다는 문제의식이 그 시작점이었다.

2010년 6월 OECD 통계위원회 산하 비공식 워킹그룹 회의에서 통계 생산 프로세스 구동에 필요한 정보를 정의하기 위한 일관된 참조 모델의 필요성이 제기되었고, HLG-MOS²⁾(공식 통계 현대화를 위한 고위급 그룹)은 2011년 6월 GSIM 0.1 버전을 시작으로 2012년³⁾ 두 차례의 스프린트와 한 차례의 워크숍 등 집중 검토 과정을 거쳐

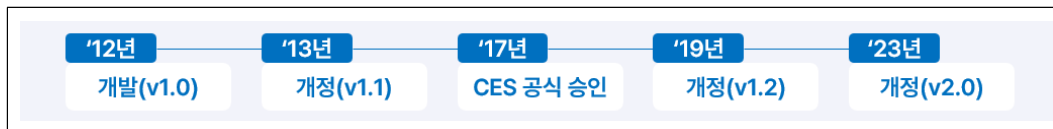
- 1) 일례로 캐나다 통계청의 ‘survey’는 UNECE에서는 ‘survey instance’로, 호주 통계청은 ‘collection cycle’로 지칭하며, 일반 조사기관에서는 ‘study’로 사용함 (출처: UNECE(2012), GSIM Specification v1.0.)
- 2) HLG-MOS는 공식 통계 현대화를 추진하기 위해 CES가 설립하였으며, 2026년 현재 13개 국가 및 국제기구의 수석 통계관으로 구성되어 있음. CES는 UNECE와 UN 통계위원회라는 두 개의 상위기구를 두고 있으며, 최종적으로는 UN총회의 조정을 받음 (출처: UNECE(2016), CES Members’ Guidebook.)



2012년 12월 GSIM 1.0 버전을 공식적으로 발간했다. 국제표준으로서의 승인은 2017년 CES(유럽 통계기관장 협의체)에서 다른 ModernStats 모델들과 함께 이뤄졌다.

이후 국제 통계 커뮤니티의 지속적인 협력을 통해 2023년 12월 GSIM 2.0 최신 버전이 발간되었고, 본 연구는 이에 기반한다.

<그림 3> GSIM 개정 연혁



출처: UNECE Statwiki, Generic Statistical Information Model [웹사이트].

2. 국제 표준 모델

GSIM은 통계 정보에 대한 국제 표준 모델이다. HLG-MOS의 후원 하에 공식 통계의 현대화를 지원하기 위해 개발된 모델을 통칭하여 ModernStats 모델이라고 하는데, 2017년 CES 승인을 받은 ModernStats 모델은 총 4종으로 GSIM을 비롯한 GSBPM, GAMSO, CSPA가 해당한다. 이들 모델은 UN 핸드북⁴⁾에서도 SDMX, SDMX and SDGs, DDI 등⁵⁾과 함께 국가통계청 차원의 표준으로 소개하고 있어, GSIM의 국제 표준으로서의 지위를 재확인할 수 있다.

GSBPM(Generic Statistical Business Process Model)은 요구사항 정의 → 설계 → 구축 → 수집 → 처리 → 분석 → 배포 → 평가를 포함하는 통계 생산 프로세스 단계를 설명하는 표준 용어를 제공한다. GAMSO(Generic Activity Model for Statistical Organizations)는 통계기관 내에서 발생하는 활동 중에서 통계 생산과 직접적인 관련이 없거나 기업 또는 전략적 수준에서 관리되는 일반적인 활동(예: 전략 및 리더십, 역량 개발 및 기업 지원 분야 활동)을 설명하고 정의한다. CSPA(Common Statistical Production Architecture)는 흔히 ‘플러그 앤 플레이’ 방식이라고 하며, 통계기관 내부 및 기관 간에 공유 및 재사용할 수 있는 소프트웨어 서비스 구축에 대한 지침을 제공하고 공통 인프라 및 서비스 개발을 위한 국제 협력 이니셔티브를 지원한다.

3) 1차 스프린트는 슬로베니아(2.20~3.2), 2차 스프린트는 한국(4.16~27), 워크숍은 네덜란드(9.17~21)

4) UN(2025). Handbook on Management and Organization of National Statistical Systems.

5) SDMX(Statistical Data and Metadata eXchange)는 통계 데이터와 메타데이터를 효율적으로 교환하고 공유하기 위한 국제표준, SDGs(Sustainable Development Goals)는 2015년 UN이 채택한 2030년까지 달성하기로 결의한 지속가능발전목표, DDI(Data Documentation Initiative)는 사회·행동·경제·보건 분야 연구 데이터에 대한 메타데이터를 문서화하는 표준임

<표 1> ModernStats 4종 모델 개발/개정 버전 이력(2008~2017년)

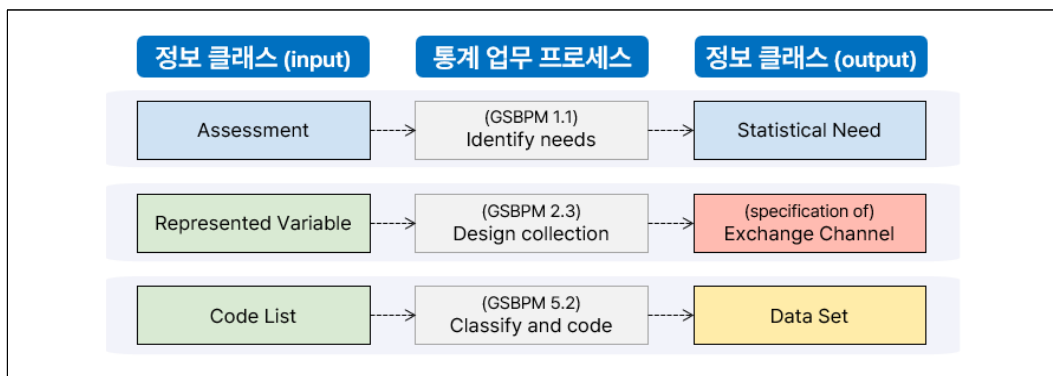
모델명	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017 (CES 승인)
GSBPM	<u>1.0</u> / 2.0 / 3.1	4.0				5.0				
GAMSO			<u>1.0</u>							1.1
GSIM				0.1	0.3 / 0.4 / 0.8 / <u>1.0</u>	1.1				
CSPA						0.1		<u>1.1</u> / 1.5		

출처: UNECE HLG-MOS Supporting Standards Group(2021), ModernStats Models Governance Guidance.

3. GSIM 성격 및 다른 표준과의 관계

GSIM은 통계 설계 및 생산 전 과정에서 투입과 산출로 사용될 수 있는 정보 클래스의 속성과 관계를 표준화되고 일관성 있게 기술한 참조 프레임워크이다. 예컨대, <그림 4>과 같이 평가 결과에서 식별 과정을 통해 통계적 수요를 도출한다고 할 때, ‘식별 과정’은 통계 활동이며, 여기에 투입 및 산출로 사용되는 ‘평가’ 및 ‘통계적 수요’는 정보 클래스에 해당한다. GSIM은 GSBPM과 같은 모든 통계 비즈니스 프로세스를 지원하며(설계원칙 ①), 이러한 정보 클래스에 합의된 이름을 부여하며 정의하고, 필요한 속성 및 다른 정보 클래스와의 관계를 나타낸다(설계원칙 ⑤). 다만, 인적 자원, 재무 또는 법률 기능과 같은 조직 내 활동과 관련된 정보 클래스는 통계 생산에 직접 사용되는 경우를 제외하고는 GSIM의 범위에 포함되지 않는다.

<그림 4> 통계 업무프로세스 단계의 투입/산출되는 정보 클래스 예시



출처: UNECE(2022), Linking GSBPM and GSIM 재구성.

<표 2> GSIM 설계 원칙

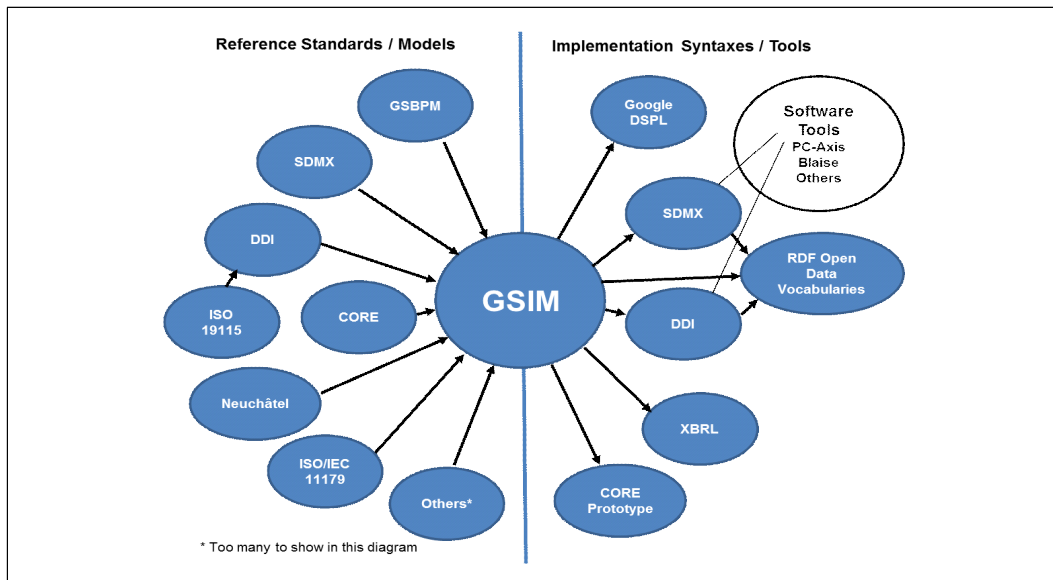
원칙	내용
0. Change control (변경 관리)	◦ GSIM 개정하거나 현지화할 때, GSIM 설계 원칙 준수
1. Complete coverage (완전 포괄성)	◦ GSBPM을 지원하며, 통계 생산 전 과정을 포괄 - GSBPM 사용 기관은 GSIM을 사용하도록 권장
2. Independence (독립성)	◦ 다른 프레임워크나 표준과 독립적으로 사용 가능 - 단, 기존 표준·모델에 대한 지식을 바탕으로 개발
3. Stakeholder appeal (이해관계자 친화적)	◦ 모든 이해관계자가 이해하기 쉽도록 직관성 필요 - 이해관계자에는 최고 경영진, 주제·방법론·기술 전문가뿐 아니라 타 통계기관, 개발자, 솔루션 업체 등 모두 해당
4. Support for products (통계산출물 지원)	◦ 설계, 문서화, 생산, 유지관리 등 통계산출물의 모든 측면 지원
5. Separation of design and production (설계와 생산 분리)	◦ 설계 단계와 생산 단계의 명확한 분리를 지원 - 방법론, 자료처리, 구성요소 등 재사용 지원
6. Usable for current and Innovative approaches (혁신지원)	◦ 현행 및 새로운 통계 생산 방식을 모두 지원 - 전통적, 혁신적 방식에 적용가능한 광범위한 지침 필요
7. Linking processes (프로세스 연계)	◦ 모든 단계의 투입 및 산출 정보를 명확히 정의
8. Common language (공통 언어)	◦ 정보 클래스에 대한 공통된 이해의 기반 제공 - 기존 용어를 매핑할 수 있는 참조 프레임워크 제공으로, 시스템 간 상호 운용성 증대
9. Layered Information (계층적 정보)	◦ 문서화에 계층적(요약~상세) 접근 방식 사용 - 목적 및 이해관계자를 고려하여 계층 구성
10. Agreed level of detail (합의된 상세 수준)	◦ 주요 이해관계자가 동의한 수준까지의 정보 클래스 포함 - 범용성(대내외 적용), 상세화(모듈화 통계생산) 균형 필요 - 상위 수준 합의 후, 경험 누적 후 상세화 합의 시도 가능
11. Robustness, adaptability and extensibility (견고성, 적응성, 확장성)	◦ 통계기관에서 채택할 수 있을 정도로 안정적이되, 사용자 요구사항을 충족할 수 있도록 지속적인 검토와 적응 필요 - 테스트 유인책 및 지속적인 검토체계 구축 필요
12. Simple presentation (간결한 표현)	◦ 클래스와 그 관계는 간결하게 표현되어야 함 - UML 다이어그램 등 정보 모델링 전문가 위주 표기 지양
13. Optimal reuse (최적의 재사용)	◦ 기존 용어와 정의를 최적으로 재활용 - 개발 효율성과 수용성 제고
14. Platform Independence (플랫폼 독립성)	◦ 특정 IT 환경이나 도구에 의존하지 않음 - 모든 사용자에게 유용하고, 시간이 지나도 견고함 유지
15. Formal specification (형식화된 명세)	◦ 속성·관계의 명세 등 정보 클래스를 적절하게 정의·분류 - 참조 프레임워크로서, 일관된 표준 또는 사양 개발을 지원

출처: UNECE(2024), GSIM 2.0 User Guide.

GSIM은 모델 구현에 사용되는 표준이나 기술에 대해서는 가정하지 않지만, 개발과 정에서는 참조 프레임워크로서의 구현 가능성 확보를 위해 다양한 표준과 기술을 참고하였다(설계원칙 ②, ⑭). 2012년 2차 스프린트 회의 자료에서도 이러한 설계원칙을 엿볼 수 있는데, <부록 2>의 메타데이터 표준 및 모델 타임라인에서 통계적 개념은 OECD glossary, Eurostat CODED와 SDMX MCV(Metadata Common Vocabulary)을, 기술 표준은 DDI, SDMX, ISO-11179, Dublin Core, SKOS 등을, 방법론적인 가이드라인은 Metanet, Neuchâtel, GSBPM 등을 정리하고 있다.

특히 개념 그룹 중 개념, 변수, 단위유형 등에 있어서는 GSIM과 DDI-CDI(Cross Domain Integration)의 호환성이 높다. 다만 GSIM은 데이터의 물리적 표현에 대한 세부 사항을 다루지 않지만 DDI-CDI에서는 개념적 값과 그 표현인 인스턴스 값을 더 구체적으로 설명한다는 차이가 있다⁶⁾. 한편, 개념 그룹 중 통계분류체계와 관련된 정보 클래스는 Neuchâtel 용어 모델을 많이 참고하였다.⁷⁾

<그림 5> GSIM과 다른 표준/모델과의 관계



출처: UNECE(2012), GSIM Specification v1.0.

6) DDI Alliance(2025). DDI-Cross Domain Intergration: Specification Overview.

7) UNECE(2013.04) NEUCHÂTEL Terminology Model - Classification database object types and their Attributes.

제2절 GSIM 이해를 위한 검토 자료

GSIM에 대한 설명서 등의 보고서는 이용 목적 및 이해관계자를 고려하여 여러 가지 상세화 수준으로 작성되어 있다(설계원칙 ⑨). 따라서 아래와 같은 다양한 수준별 문서를 활용하여 GSIM을 이해하고자 하였다.

첫 번째, Communication Paper(2022.1.)에서는 GSIM이 가져다주는 이점을 다양한 관점에서 설명하며 다른 모델·표준과의 관계를 요약하여 설명한다.

두 번째, GSIM 2.0 User Guide(2024.3.)에서는 정보 클래스별 역할 및 다른 정보 클래스와의 관계를 설명한다. 다만 본문에 제공되는 다이어그램은 정보 클래스 간의 관계를 일반 이용자를 고려하여 일반 연결선과 관계명으로 단순화하여 표현하고 있다.

세 번째, GitHub에서는 정보 클래스에 대한 용어집을 제공한다. 정의, 설명문, 유어(있는 경우)와 함께, 필요한 속성이 있는 경우에는 속성명, 속성 설명, 카디널리티(Cardinality), 속성값 유형을 제공한다. 카디널리티는 다중성을 표시하는 차수로서, ‘(최솟값..최대값)’의 형식으로 표현한다. 예컨대, <그림 6>의 경우 (1..1)이므로 Conceptual Domain에 대해 항상 하나의 Sentinel 속성값을 가져야 하는 것으로 해석된다.

<그림 6> GSIM 용어집 조회화면 (일부)

Conceptual Domain				
Definition				
Class	Group	Definition	Explanatory Text	Synonyms
Conceptual Domain	Concept	set of valid Concepts	The Concepts can be described by either enumeration or by an expression.	
Attributes				
Name	Description	Cardinality	Value Type	
Sentinel	If true, the domain is sentinel (i.e. values used to represent a state in the processing life-cycle e.g. missing data), otherwise the domain is substantive (i.e. values used to represent an observation of some Unit of interest).	1..1	Boolean	

* Attributes inherited from super-type(s) are not included here

출처: UNECE, GSIM Version 2.0 - 클래스 정의 [웹사이트] (2026.1. 조회).

네 번째, GitHub에서 그룹별로 그룹 내 모든 정보 클래스 간의 관계를 나타내는 그룹 다이어그램을 제공한다. User Guide와 달리 UML(Unified Modeling Language) 표기법에 따른 연결선과 카디널리티를 제공한다. 설계원칙⑫에서 이와 같은 정보 모델링 전문가 위주의 표기를 지양한다고 하지만, 이를 통해 정보 클래스의 역할, 관계 및 제약 조건 등 GSIM 모델을 더 면밀하게 이해할 수 있다.⁸⁾

다섯 번째, HLG-MOS 지원표준그룹 TF에서 2019년부터 2021년까지 개발한 Linking GSBPM and GSIM task team report(2022.1.)를 검토하였다. 본 문서에서는 <그림 7>과 같이 GSIM 정보 클래스를 GSBPM 하위 프로세스의 투입 및 산출로 구조화하는 템플릿을 제공한다. 프로세스 입력에 해당하는 핵심 입력, 매개변수 입력, 프로세스 지원 입력과 프로세스 출력에 해당하는 핵심 출력 등을 통해 프로세스 과정에서의 정보 클래스의 역할과 상태 변화를 더욱 입체적으로 이해할 수 있다.

<그림 8> GSBPM 하위 프로세스별 GSIM 정보 클래스 연계 (일부)

Specification level ^{4,1}		
Process Input Specification ^{4,2}	Process Step Design ^{4,3}	Process Output Specification ^{4,3}
Core Input type^{4,4} <i>Business Cases</i> containing proposed solutions and approved decisions regarding <i>Concept to be measured^{4,4}</i> ^{4,4} Parameter Input type^{4,4} ^{4,4} Process Support Input type^{4,4} Standards, policies or guidelines to be followed on how to design <i>Variables</i> to be collected and write metadata description ^{4,4} Repository or inventory of ^{4,4} • <i>Variable, Unit Type</i> (e.g., persons, households, enterprises), <i>Measurement Type</i> (e.g., currencies, capita)^{4,4} • <i>Value Domain, Measurement Unit^{4,4}</i> • <i>Statistical Classification, Code List^{4,4}</i> that are agreed and approved in the <i>Organisation</i> to be used as reference ^{4,4} <i>Represented Variables, Process Methods, Statistical Classification</i> or any other referential metadata available from previous <i>Statistical Program Cycles</i> or similar <i>Statistical Programs</i> that can be used as reference ^{4,4} ^{4,2}	Process Method^{4,4} Review <i>Business Case</i> regarding <i>Concepts</i> and <i>Variables^{4,4}</i> ^{4,4} Consult repository or inventory of <i>Variables, Statistical Classifications, etc.</i> , relevant standards, guidelines and any available information from previous <i>Statistical Program Cycles</i> or similar <i>Statistical Programs^{4,4}</i> ^{4,4} Design <i>Represented Variables (Variables, Value Domains)^{4,4}</i> ^{4,4} Design <i>Statistical Classifications*</i> if needed ^{4,4} ^{4,4} Design <i>Process Methods</i> to be used for derivation and calculation of <i>Represented Variables^{4,4}</i> ^{4,4} Prepare metadata descriptions of <i>Represented Variables, Process Methods, etc.^{4,4}</i> ^{4,4} ^{4,4} ^{4,4} * Version local to the current business process, the maintenance of <i>Statistical Classification</i> that is used across the <i>Organization</i> takes place as a separate business process. ^{4,2}	Core Output type^{4,4} <i>Statistical Program Designs</i> that containing ^{4,4} • <i>Represented Variables (Variables, Value Domains)</i> to be used for collection and to be derived^{4,4} • <i>Process Methods</i> specifying how to derive and calculate the <i>Represented Variables^{4,4}</i> • <i>Statistical Classifications</i>: description of the statistical classification (geospatial) structure and content.^{4,4} • <i>Referential Metadata Set</i> describing any relevant information about <i>Process Methods, Statistical Classification, etc.^{4,4}</i> ^{4,4} Process Metric type^{4,4} Quality measures regarding designing such as: ^{4,4} • Percentage of <i>Represented Variables</i> and classifications that have metadata descriptions^{4,4} • Percentage of definitions and classifications associated to (key) <i>Represented Variables</i> and <i>Populations</i> that follow international and national standards^{4,4} Quality measures of the <i>Process Step</i> such as: ^{4,4} • Time spent to complete the <i>Process Step</i> (derived from <i>Process Execution Log^{4,4}</i>)^{4,4} • Cost spent to complete the <i>Process Step^{4,4}</i>

출처: UNECE(2022), Linking GSBPM and GSIM.

여섯 번째, DDI-Cross Domain Integration Specification Overview(2025) 문서를 검토했다. DDI Alliance는 GSIM 개발 과정에서 호환성 확보를 위해 협력 관계를 유지해 왔다. DDI-CDI의 핵심 구조 중 하나인 변수 연쇄(Variable Cascade)가 GSIM에 기반을 두고 있음을 명시하면서 상세한 설명과 예시를 수록하고 있어서 GSIM의 변수 재사용성 구조를 구체적으로 이해하는데 유용한 자료이다. 또한 속성과 특성(characteristic), 내연적 정의와 외연적 정의 등 기본적인 용어에 대해서도 이해할 수 있도록 부록에 수록하고 있다.

<그림 9> DDI-CDI에서 설명하는 변수 연쇄 예시 (일부)

Concept			Represented Variable (Inherits from Conceptual Variable)		
ID	Name	Definition	Name	Marital Status	Age
1	Marital status	Category of current marital arrangement	Universe	Deer Hunters	Gun Shops
2	Age	Whole number of years of operation	Substantive Value Domain	<s, Single> <m, Married> <d, Divorced> <w, Widowed>	Count (of years) represented with 2-digit Arabic numeral
Conceptual Variable (Links to Concept)			Unit of Measure	N/A	years
			Intended Datatype	Nominal	Quantitative discrete
			Instance Variable (Inherits from Represented Variable)		
			Name	Marital Status	Age1
			Population	US Deer Hunters in 2019	US Gun Shops in 2019
			Sentinel Value Domain	<1, Missing> <2, Refused>	<1, Missing> <2, Refused>
			Function	demographic	establishment
			Physical Datatype	1-character	2-integer

출처: DDI Alliance(2025), DDI-CDI Specification Overview.

그밖에 GSIM 2.0 웨비나(2025.9.16.)와 GSIM wiki(statwiki.unece.org/spaces/gsim/overview)에서 조회할 수 있는 개정 과정에서의 의견 등을 다각적으로 검토하여 참조 메타데이터 모델로서의 GSIM을 제대로 이해하고자 노력하였다.

제3절 그룹별 정보 클래스 세부 내용

정보 클래스는 역할에 따라, 기초, 비즈니스, 교환, 구조, 개념의 5개 그룹으로 구분하고, 다이어그램에서도 그룹별 고유의 색상으로 표현한다. 비즈니스 그룹과 교환 그룹은 통계 업무 프로세스 과정에서 사용되며, 개념 그룹과 구조 그룹은 기저(foundational) 정보로 사용되고, 기초 그룹은 기술적인(technical) 정보를 담고 있다. 본 연구의 주요 초점은 개념 그룹이지만, 개념 그룹을 이해하기 위해 다른 그룹의 정보 클래스에 대한 이해도 일부 필요하다. 따라서, 기초, 비즈니스, 교환, 구조 그룹을 간단히 설명한 후, 개념 그룹에 대해 상세히 다루고자 한다.

<표 4> GSIM 그룹별 정보 클래스

그룹	역할	정보 클래스 개수*	다이어그램 색상
기초 그룹 (Base Group)	◦ 식별, 버전 관리 등 - 주체, 역할, 이력, 식별가능 아티팩트 등	12개	
비즈니스 그룹 (Business Group)	◦ 통계 프로그램 설계 및 수행 - 통계 프로그램, 업무 프로세스 등	32개	
교환 그룹 (Exchange Group)	◦ 데이터 수집 및 배포 - 조사표, 산출물, 정보 이용자 등	21개	
구조 그룹 (Structures Group)	◦ 데이터 구성 방법 - 데이터세트 구조, 구성요소 등	23개	
개념 그룹 (Concepts Group)	◦ 데이터의 의미 및 표현 - 개념, 변수, 단위유형, 분류체계 등	39개	

* 정보 클래스 개수는 용어집에 수록된 정보 클래스의 개수로, User Guide에서 설명하지 않는 Sentinel Value Domain과 같은 정보 클래스를 모두 포함한 개수임.

출처: UNECE, GSIM Version 2.0 - 클래스 정의 [웹사이트] (2026.1. 조회).

1. 기초 그룹(Base Group)

기초 그룹에서 Identifiable Artefact⁹⁾은 GSIM의 최상위 추상 클래스로서, 관리대상이 되는 대부분의 정보 클래스가 Identifiable Artefact의 속성¹⁰⁾을 상속받아 고유하게 식별되고 관리된다. 따라서 별도의 언급이 없더라도, 정보 클래스별로 필수(1..1) 속성인 Id (고유식별자)와 Name(명칭)은 수반되어야 하며, 선택(0..1) 속성인 Local ID(예: 특정 데

9) Identifiable Artefact 클래스 정의: 식별, 명명 및 기타 문서화에 필요한 기본 속성과 연관 관계를 포함하는 추상 클래스 (abstract class that comprises the basic attributes and associations needed for identification, naming and other documentation)

- (해설) 관리 세부 정보, 역할 담당 에이전트, 변경 이벤트, 데이터, 프로세스 입력(및 하위 유형), 프로세스 출력(및 하위 유형)을 제외한 모든 GSIM 클래스는 식별가능한 아티팩트의 하위 유형이다. SDMX에서 '식별가능한 아티팩트'는 '객체에 식별 정보를 제공할 수 있는 구조를 포함하는 구성 요소'로 정의된다.

10) Identifiable Arefact 클래스 속성:

Name	Description	차수	값유형
Description	정보 클래스에 대한 설명	0..1	다국어Text
Id	정보 클래스의 고유 식별자 (owner agency에서 부여)	1..1	String
Local ID	특정 로컬 맥락에서의 고유 식별자 (예: 특정 데이터세트 컬럼명)	0..1	String
Name	정보 클래스의 이름 (동이어·다국어 명칭이 여러 개 가능)	1..1	다국어Text
Version	버전 표시자 (owner agency에서 부여)	0..1	String
Version Date	버전 생성 날짜	0..1	Date
Version Rationale	버전 생성 이유	0..1	String

이터세트에서의 컬럼명), Description(설명), Version(버전), Version Date(버전생성일자), Version Rationale(버전 작성 사유) 등도 함께 관리되는 것이 필요할 것으로 생각된다.

2. 비즈니스 그룹(Business Group)

비즈니스 그룹은 통계 생산 과정에서 발생하는 수요의 도출, 사업 기획 및 설계, 프로세스의 정의와 실제 실행에 이르는 비즈니스 활동 전반을 다루는 정보 클래스 집합이다.

통계 생산의 시작점은 Statistical Need(통계 수요)이다. ‘고령층 고용통계를 세분화해 달라’와 같은 정책 수요가 해당한다. 이러한 통계 수요는 다듬어지지 않은 요청 상태일 뿐이므로 이를 개념적으로 구조화한 것이 Change Definition(변경 정의)이다. Change Definition은 완전히 새로운 개념을 만들어 내는 것이 아니라, 이미 있는 개념을 참조하여(is informed by) 변경해야 할 사항을 명확히 정의하는 식으로 이루어진다. 앞의 고령층 세분화 요구 예시에 대해 ‘연령구간 표현변수의 Value Domain 세분화’라는 식으로 변경사항을 정의할 수 있다.

Business Case(기획서)는 이러한 변경 정의에 기반하여, 예산, 실현가능성 등을 검토 사항을 담은 기획 문서이다. 기획서에 따라 하나 이상의 Statistical Programme(통계 프로그램)이 추진된다. Statistical Programme은 특정 Universe를 대상으로 한 통계생산활동 일체를 말하는 것으로 경제활동인구조사, 사업체노동력조사, 인구총조사 등이 해당된다. 조사가 반복적으로 수행되는 경우 2026년 6월 경제활동인구조사, 2026년 7월 경제활동인구조사 등은 각각 하나의 Statistical Programme Cycle(통계 프로그램 주기)¹¹⁾에 해당한다.

Statistical Programme이 목적을 달성하려면 구체적인 수행 방법을 기술하는 Statistical Programme Design(조사설계서)¹²⁾이 필요하다. 이 설계에는 수행하는 활동(‘무엇’)을 나타내는 Business Function(업무 기능)을 사용한다. 이 Business Function을 실제로 수행하는 구조화되고 반복 가능한 활동(‘어떻게’)은 Business Process(업무 프로세스)이다. 예를 들어, 자료수집, 자료처리, 공표 등은 Business Function이며, 이를 일정한 순서로 수행하

11) Statistical Programme Cycle 클래스 속성:

Name	Description	차수	값유형
Reference Period End	참조 기간의 종료일자	1..1	Date
Reference Period Start	참조 기간의 시작일자	1..1	Date

12) Statistical Programme Design 클래스 속성:

Name	Description	차수	값유형
Conceptual Framework	통계 프로그램(예: SNA)의 개념적 틀에 대한 설명	0..*	String
Status	예: 신규제안, 개발 중, 진행 중, 완료, 취소, 다른 기관으로 이전	1..1	통제어휘집

는 월별 생산 절차는 Business Process가 된다. 같은 Business Function이라도 조직마다, 기술 환경마다 서로 다른 Business Process로 구현될 수 있다는 점에서 이 둘의 구분이 재사용성의 핵심이 된다.

Business Process는 다시 Process Step(프로세스 단계)이라는 더 작은 개별 작업 단위로 나뉘고, Process Design(프로세스 설계)은 이 Process Step들을 어떤 순서로 배열하고 각 단계에서 어떤 정보 클래스가 입력되고 출력되어야 하는지를 Process Input Specification(프로세스 입력 사양)과 Process Output Specification(프로세스 출력 사양)으로 정의한다. 프로세스 입출력 변환 과정에서 Process Method(프로세스 방법론)가 적용된다. Process Method는 표본추출 방법, 분산추정 방법, 계절조정 방법 등 특정 작업을 수행하기 위한 방법론을 정의하는 클래스이다.

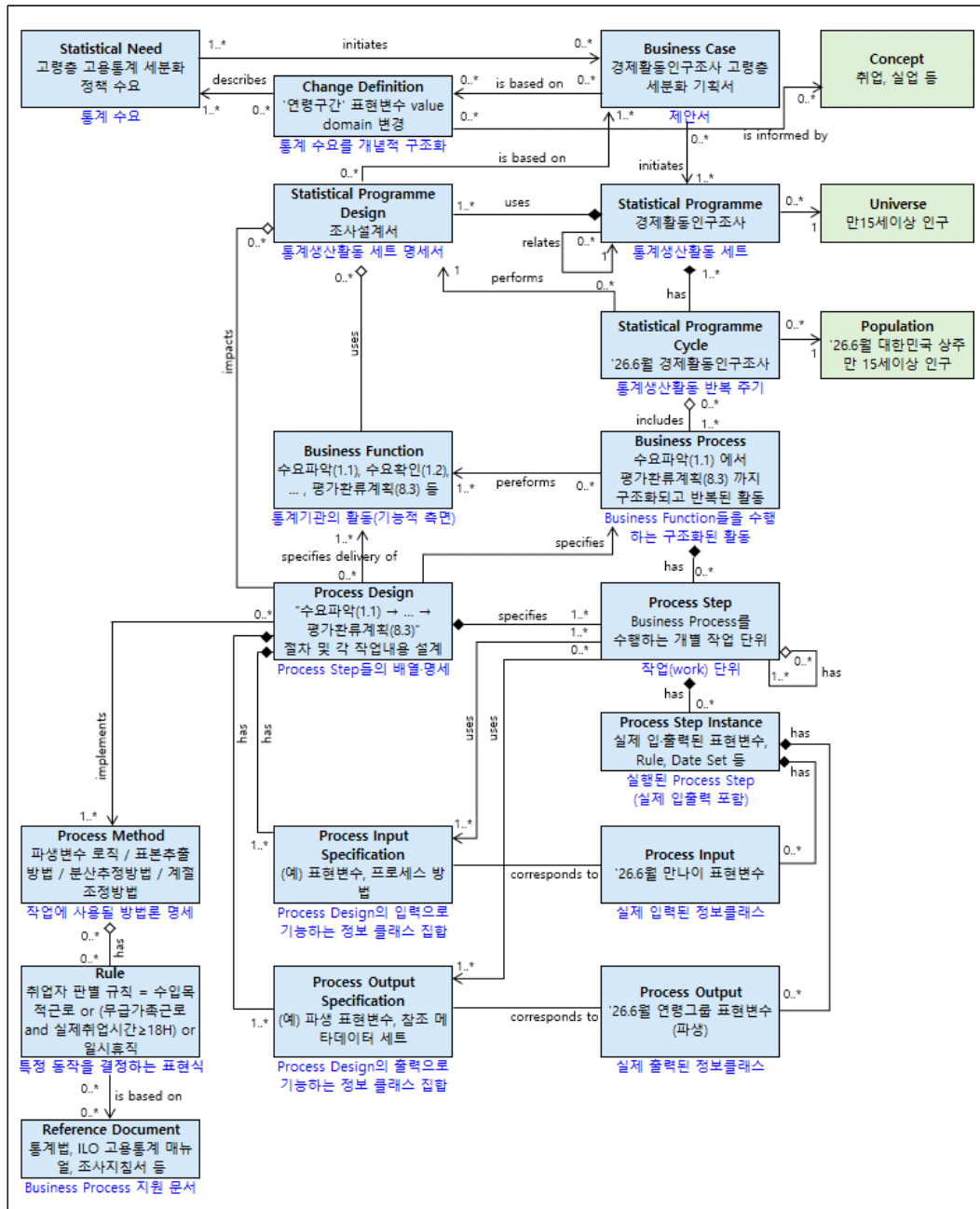
Process Method의 한 유형인 Rule(규칙)¹³⁾은 ‘특정 동작을 결정하기 위한 수학적 또는 논리적 표현’으로 정의된다. 규칙의 형태와 목적, 특성, 표현 방식은 매우 다양하다. Rule의 속성 중 Rule Type은 통제어휘집(ControlledVocabulary)¹⁴⁾ 값을 가지며, GSIM 용어집은 그 예시로 Input, Comparison, Imputation, Edit, Derivation, Recode 등을 제시하고 있다. 이 중에서 Derivation, Recode는 Concept 그룹의 Represented Variable(표현변수) 생성·변환과 직접적으로 관련되는 규칙으로 볼 수 있다.

13) Rule 클래스 속성:

Name	Description	차수	값유형
Algorithm	알고리즘으로 표현된 규칙	0..1	String
Command Code	시스템이 명령을 처리하기 위한 구조화된 정보	0..*	String
Expression	실행되는 규칙의 표현	0..1	String
Is System Executable	시스템 실행용 형식 지정 or 단순 문서화 수준	0..1	Boolean
Rule Type	예: Input, Comparison, Imputation, Edit, Derivation, Recode 등	0..1	통제어휘집

- 14) GSIM 용어집에서 값 유형이 통제어휘집(ControlledVocabulary)인 정보 클래스 속성을 국내에 적용할 경우, 허용가능한 값 목록을 사전에 정의할 필요가 있음을 시사함

<그림 10> 비즈니스 그룹 주요 정보 클래스



* 예시가 같은 흐름에서 서로 연계되는 것이 아님을 유의

** 비즈니스 그룹에 대한 정보 클래스 전체에 대한 재배치 도식은 <부록> 참고

출처: UNECE, GSIM Version 2.0 - 그룹 다이어그램 [웹사이트] (2026.7. 조회) 재구성.

3. 교환 그룹(Exchange Group)

교환 그룹은 통계 생산 기관과 외부 환경 간에 이루어지는 정보의 수집·획득 및 배포를 설명하는 정보 클래스 집합이다. 개념 그룹이 정의한 개념이 실제로 어떤 문항·도구를 통해 수집되고, 어떤 형태로 이용자에게 전달되는지, 그 정보 교환의 주체는 누구인지 등을 포함한다. 개념 그룹과의 접점은 Question 및 Instance Question으로, 개념 그룹의 Conceptual Variable 및 Instance Variable과 각각 관련된다.

<표 5> 교환 그룹 정보 클래스 목록

구분	정보 클래스 (총 18개)	내용
주체, 근거	Information Consumer	정보 소비자 (예: 정책부처, 연구자, 일반국민)
	Information Provider	정보 제공자 (예: 응답가구, 행정DB 제공 기관)
	Provision Agreement	정보교환 근거 (예: 이용약관, 입수 협약)
교환 도구	Exchange Specification	교환 설계서
	Exchange Instrument	교환 도구
	Data Harvest	교환 도구 유형 - 기계 교환 (예: Open API)
	Dissemination Instrument	교환 도구 유형 - 배포 도구 (예: 웹서비스, 책자)
	Questionnaire	교환 도구 유형 - 조사표
수집	Questionnaire Specification	조사표 설계서
	Questionnaire Component	조사표 구성
	Questionnaire Logic	조사표 문항순서 제어
	Question Block	질문문항 및 안내문구 묶음
	Question	질문문항
	Statement	조사표 내 안내문구
수집 구현	Instance Question Block	Question Block 실제 사용
	Instance Question	Question 실제 사용
	Instance Statement	Statement 실제 사용
배포	Output Specification	Product 안에서 Information Set의 배치·표현 설계서
	Product	소비자 배포용 패키지화된 콘텐츠 (예: 보도자료, KOSIS 통계표)
	Presentation	데이터 및 참조메타데이터 표현 방식*
기타	Register	통계 등록부 및 행정 등록부

* Presentation은 표, 그래프, 구조화된 데이터 파일 등 다양한 형태가 가능하다. 예를 들어, 표, 표준 기반 데이터 파일(e.g. SDMX), PDF, 표준 형식의 구조적 메타데이터(e.g. DDI 3.1 XML), 산출물 및 서비스 목록(e.g. a product catalogue or a web services description language (WSDL) file), 통계분류 및 개념변수 설명을 조회할 수 있는 웹 페이지 등 (출처: UNECE, GSIM Version 2.0 - Presentation 클래스 설명문)

출처: UNECE, GSIM Version 2.0 - 클래스 정의 [웹사이트] (2026.1. 조회).

4. 구조 그룹(Structure Group)

구조 그룹은 통계 비즈니스 프로세스 전반에 걸쳐 정보를 구조화하는 데 사용되는 정보 클래스 집합이다. 데이터와 참조 메타데이터를 포괄하는 모든 정보에 대해서 다루고 있다.

구조 그룹을 구성하는 핵심 정보 클래스는 데이터 구조의 유형과 구성요소를 정의하는 클래스들로 체계화된다.

우선, 데이터 구조의 유형은 Unit Data Structure(단위 데이터 구조)와 Dimensional Data Structure(차원 데이터 구조)로 구분된다. 전자는 마이크로데이터와 같이 하나의 레코드가 하나의 관측 단위(Unit)에 대응되는 구조이고, 후자는 통계표와 같이 특정 차원을 기준으로 집계·재구성한 구조이다. 예를 들어 경제활동인구조사에서 개인의 취업여부는 Conceptual Variable로 정의될 수 있으며, 이를 표현하기 위한 Represented Variable과 Value Domain이 결정된다. 이후 실제 조사자료에서는 이러한 표현변수가 Unit Data Structure에 포함되어 하나의 변수인 Instance Variable로 저장되고, 개별 응답값은 Data Point로 기록된다. 공표 단계에서는 동일한 개념이 Dimensional Data Structure를 이용하여 성별, 연령, 지역 등의 차원(Dimension) 형태의 통계표로 재구성된다. 이처럼 구조 그룹은 동일한 개념이 단위 자료와 집계 자료에서 서로 다른 구조로 표현될 수 있도록 지원하며, 개념과 데이터 구조를 분리함으로써 동일한 개념을 다양한 데이터세트와 통계표에서 재사용할 수 있다.

다음으로 Data Structure Component(데이터 구조 구성요소)는 3가지로 구분되는데, Identifier Component(식별자 구성요소)는 데이터 행(Row)이나 셀(Cell)을 유일하게 식별하기 위한 변수이고, Measure Component(측정값 구성요소)는 통계 수치에 해당하는 변수, Attribute Component(속성 구성요소)는 데이터를 해석하는 데 필요한 정보 중 식별자 및 측정값을 제외한 정보로 측정단위, 잠정 여부, 주석 등을 생각할 수 있다. 개념 그룹이 무엇을 측정하는지에 대한 의미를 다루고, 구조 그룹은 그렇게 측정된 값들이 데이터세트 안에서 어떻게 배치되는지에 대한 구조를 다룬다는 점에서 역할이 분명하게 구분된다. 다만 개념 그룹에 해당하는 Represented Variable이 Data Structure Component라는 역할을 부여받는 점에 주목할 필요가 있다.¹⁵⁾ 이러한 식별자, 측정값, 속성이라는 역할은 Represented Variable이 원래부터 갖고 있는 성질이 아니라 특정 Data Structure 안에 놓였을 때 그 맥락에서 부여받는 것이다. 예컨대 ‘성별’이라는 Represented Variable

15) Data Structure Component 클래스는 ‘role of the Represented Variable in the context of a Data Structure’로 정의되며, UNECE(2024), GSIM 2.0 User Guide.에서도 ‘A Data Structure describes the structure of a Data Set by means of Data Structure Components (Identifier Components, Measure Components and Attribute Components). These are all Represented Variables with specific roles.’이라고 설명하고 있다.

은 ‘성별 취업자 수’ 통계표에서는 식별자 역할을 부여받지만, 성별 자체가 관심 측정값인 통계표에서는 측정값 역할을 할 수도 있다.

<표 6> 구조 그룹 정보 클래스 목록

구분	정보 클래스 (총 23개)				
정보 구조	Information Structure				
	하위 유형	Data Structure			
		하위 유형	Unit Data Structure (단위 데이터 구조)		
			Dimensional Data Structure (차원 데이터 구조)		
		구성 요소	Data Structure Component		
			하위 유형	Identifier Component (식별자 구성요소)	
				Measure Component (측정값 구성요소)	
		Attribute Component (속성 구성요소)			
Referential Metadata Structure					
정보 세트	Information Set				
	하위 유형	Data Set			
		구성 요소	Data Record		
			Data Point		
		Referential Metadata Set			
정보 자원 (=정보 세트의 집합)	Information Resource				
	하위 유형	Data Resource			
		Referential Metadata Resource			
데이터 구조 관련	Logical Record				
	Record Relationship				
참조 메타데이터 관련	Referential Metadata Subject				
	Referential Metadata Subject Item				
	Referential Metadata Content Item				
	Referential Metadata Attribute				

출처: UNECE, GSIM Version 2.0 — 클래스 정의 [웹사이트] (2026.1. 조회).

5. 개념 그룹(Concept Group)

<표 7> 개념 그룹 정보 클래스 목록

구분	정보 클래스 (총 39개)		
개념 상위	Subject Field		
	Concept System		
개념	Concept		
	하위유형	Unit Type	(단위의 추상화)
		Universe	(맥락 반영, 시공간 조건 없음)
		Population	(Universe + 시공간 조건)
		Category	(개념을 외연적 정의)
변수	Conceptual Variable		(개념+단위유형)
	하위유형	Represented Variable	(개념+단위유형+표현)
		유형 Instance Variable	(실제 데이터세트에 구현)
개념 도메인	Conceptual Domain		
	하위유형	Described Conceptual Domain	(서술형)
		Enumerated Conceptual Domain	(열거형)
값 도메인	Value Domain		
	하위유형1	Described Value Domain	(서술형)
		Enumerated Value Domain	(열거형)
	하위유형2	Substantive Value Domain	(실질값)
		Sentinel Value Domain	(결측처리값)
명칭	Designation		
	하위유형	Code	(범주에 대한 Designation)
노드 세트	Node Set		
	하위유형	Category Set	
		Code List	
		Statistical Classification	
	구성요소	Level	(필수 아님)
	Correspondence Table		
	하위요소	Map	(필수)
	Classification Family		
	Classification Series		
	Classification Index		
	Classification Index Entry		
노드	Node		
	하위유형	Category Item	
		Code Item	
		Classification Item	
측정단위	Measurement Type		
	Measurement Unit		
기타	Unit		
	Datum		

출처: UNECE, GSIM Version 2.0 - 클래스 정의 [웹사이트] (2026.1. 조회).

1) 재사용성을 고려한 정보 클래스 분리

GSIM은 설계 수준과 구현 수준을 분리하여(설계원칙 ⑤), 실제로 구현된 실체를 Instance 클래스로 정의한다. 교환 그룹에서는 Conceptual Variable에 대한 Question(질문)과 특정 조사표에서 사용된 질문인 Instance Question(인스턴스 질문)이 분리되어 있고, 개념 그룹에서는 측정값을 어떻게 표현할지까지 정의된 Represented Variable(표현변수)과 실제 데이터세트에 표현변수가 사용된 Instance Variable(인스턴스 변수)로 분리되어 있다. 비즈니스 그룹에서도 Process Step과 각 단계에서의 투입/산출되는 정보 클래스의 사양과 그 실체가 분리되어 있다. GSIM의 이런 설계는 공통 사양을 분리함으로써 정보 클래스의 재사용성을 높이고, 공통 정보의 변경이 있을 때 일괄적인 반영 등 관리 효율성을 제고함으로써 상호운용성과 일관성을 확보할 수 있다는 장점이 있다.

<표 8> 설계와 구현이 분리되어 있는 정보 클래스

그룹	정보 클래스 (설계)	정보 클래스 (구현)	정의
교환	Question	Instance Question	use of a Question in a particular Questionnaire
	Question Block	Instance Question Block	use of a Question Block in a particular Questionnaire
	Statement	Instance Statement	use of a Statement in a particular Questionnaire or more specifically a Questionnaire Component
비즈니스	Process Step	Process Step Instance	executed step in a Business Process specifying the actual inputs to and outputs from an occurrence of a Process Step
	Process Input Specification	Process Input	instance of an information class supplied to a Process Step Instance
	Process Output Specification	Process Output	instance of an information class produced by a Process Step Instance
개념	Represented Variable	Instance Variable	use of a Represented Variable within a Data Set

출처: UNECE, GSIM Version 2.0 - 클래스 정의 [웹사이트] (2026.1. 조회).

이러한 분리는 추상적 수준의 Concept(개념)이 Conceptual Variable(개념변수)과 Represented Variable(표현변수)을 거쳐 마침내 물리적 실체인 Instance Variable(인스턴스 변수)로 구체화되는 변수 연쇄(Variable Cascade)¹⁶⁾에서도 극명하게 나타난다. DDI-CDI Specification(2025)에서는 이들 정보 클래스가 실제 변수라기 보다는 데이터세트의 변수를 설명하기 위한 재사용 및 분리가능한 부품이라고 설명한다. 개념이라는 추상적인

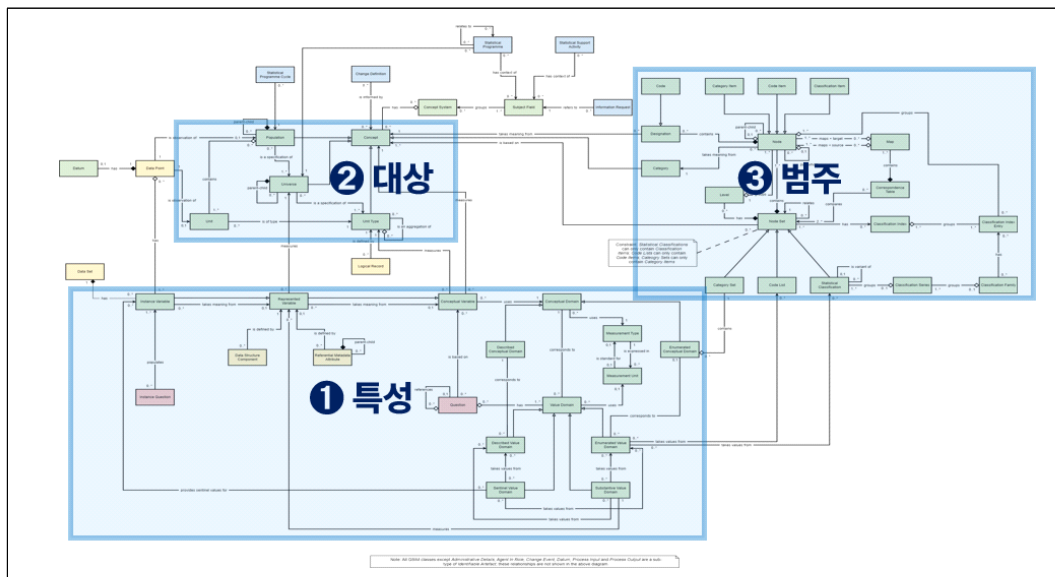
16) ‘Variable Cascade’라는 표현은 GSIM 2.0 웨비나(‘25.9.), DDI-CDI Specification(‘25.1.)에서 사용되었음

수준에서 점차 다양한 역할로 구체화되는 과정과 그 과정에서 작동되는 장치 및 속성 변화를 고려하며 개념 그룹의 정보 클래스를 이해할 필요가 있다.

2) ‘개념’이란?

개념 그룹의 가장 중심에 있는 Concept(개념) 정보 클래스는 ‘특성으로 구별되는 사
고의 단위(unit of thought differentiated by characteristics)’로 정의된다. 정보 클래스 속성
은 Definition(1..*)으로, 개념에 대한 정의 속성은 반드시 필요하다.¹⁷⁾ GSIM 2.0 User
Guide(2024)에서는 Concept이 통계 수명주기 전반에 걸쳐 다양한 방식으로 사용되는데,
❶ 측정할 특성으로서의 역할, ❷ 정보가 다루고자 하는 대상(객체, 사건, 엔터티)의 집
합(단위유형, 모집단, 유니버스 등)으로서의 역할, ❸ 개념의 세부 정의를 위한 범주
서의 역할을 수행한다고 설명한다.

<그림 11> 개념 그룹 다이어그램



* ❶~❸ 묶음은 저자가 표시한 것으로, Concept 상위의 Concept System, Subject Field와 모든 역할과 독립적인 Datum은 역할 묶음에 미포함

출처: UNECE, GSIM Version 2.0 - 그룹 다이어그램 [웹사이트] (2025.12. 조회).

17) Concept 클래스 속성:

Name	Description	차수	값유형
Definition	관련 개념들과 구별되는 서술적 진술을 통해 개념을 표현하는 것	1..*	MultilingualText

- Definition 속성의 카디널리티가 (1..*)이므로 개념에 대한 정의가 여러 개 존재하는 것을 허용하는 것으로 해석할 수 있음. (단, 값유형이 MultilingualText라는 것은 여러 가지 언어로 작성될 수 있음을 의미하는 것이므로, 속성의 카디널리티 해석과는 독립적임)

GSIM과 호환되는 다른 표준에서도 유사하다. ISO 11179-1에서는 Concept은 ‘고유한 특성들의 조합으로 생성된 지식 단위’이며, 특정 자연어에 국한되는 것은 아니지만 사회·문화적 배경의 영향을 받는다고 설명하고 있다. 또한, Concept과 그 지칭(designation)을 혼동하지 않도록 주의해야 한다고 강조한다. 예컨대, ‘소득’이라는 단어는 모든 독자에게 동일한 개념으로 전달되지 않을 수 있으며, 더 중요한 것은 ‘소득’이라는 개별 사례가 모두 동일한 개념을 지칭하는 게 아닐 수도 있다고 설명한다. ‘노동’, ‘근로’, ‘일’, ‘Labour’, ‘Work’ 등도 마찬가지로 사례가 될 것으로 보인다.

DDI-CDI Specification(2025)에서는 변수들이 ‘최상위 수준의 공통 개념’을 공유한다고 설명한다. 예컨대, 개별 조사에서의 ‘혼인상태’ 변수는 제각각일 수 있지만, 사회 전반에서 통용하는 ‘혼인상태’라는 본질은 동일하다는 것이다. 따라서 이 수준에서의 개념은 모든 가능한 변수 변형을 포괄해야 하므로 매우 일반적이라고 설명한다. 한편, 동일 개념이 변수 및 단위 유형뿐 아니라 범주로도 자주 사용된다는 설명에서, ‘성별’ 개념변수의 범주인 ‘남성’이 데이터 변환을 통해 ‘남성 여부’라는 개념변수로 정의되더라도 두 변수의 개념적 본질은 유지되므로 서로 비교 가능하다고 설명한다.

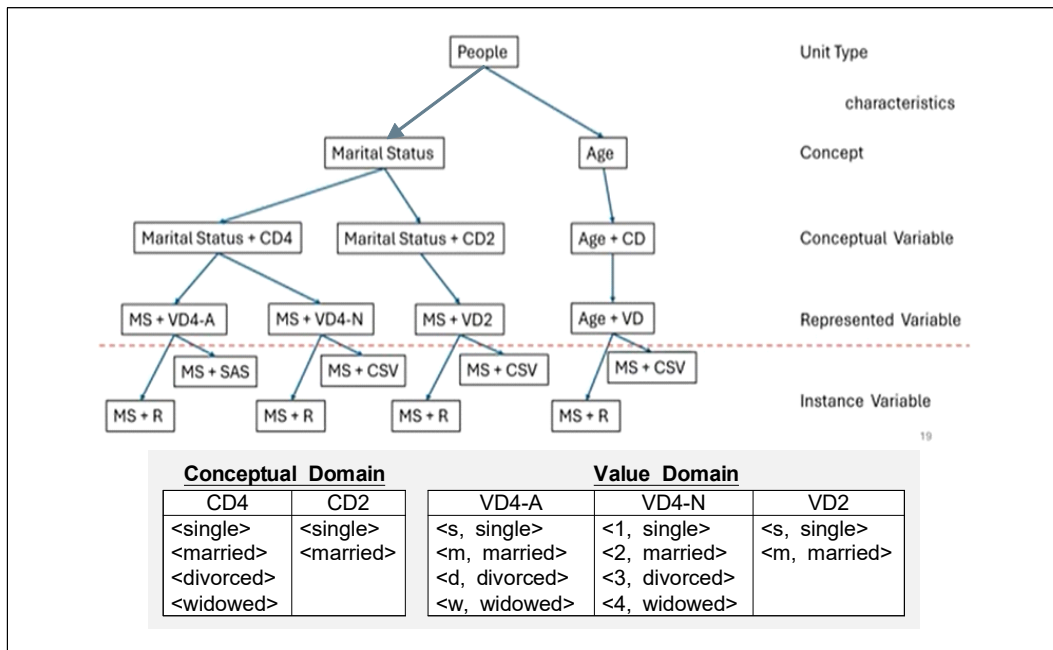
<그림 12> 국제표준별 Concept 정보 클래스의 정의 및 활용시 유의점

GSIM 용어집
• 클래스 정의: 특성(characteristic)으로 구별되는 사고(thought)의 단위 • 클래스 속성: Definition(1..*)(관련 개념들과 구별되는 서술적 진술을 통해 개념을 표현하는 것)
GSIM User Guide
• 개념은 통계 수명주기 전반에 걸쳐 다양한 방식으로 사용됨 ① 측정할 특성, ② 정보의 대상(단위유형, 모집단, 유니버스, 이벤트, 엔터티 등), ③ 개념의 세부 정의를 위한 범주
ISO 11179-1
• 고유한 특성들의 조합으로 생성된 지식 단위 - 특정 자연어에 국한되는 것은 아니나, 사회·문화적 배경의 영향을 받음 / 종종 다른 범주화로 이어짐 / 표현방식과는 독립적 • 개념과 그 지칭(designation)을 혼동하지 않도록 주의 → (예) ‘소득’이라는 단어는 모든 독자에게 동일한 개념으로 전달되지 않을 수 있으며, 더 중요한 것은 ‘소득’이라는 개별 사례가 모두 동일한 개념을 지칭하는 게 아닐 수 있음
DDI-CDI Specification
• 변수들은 최상위 수준의 공통 개념을 공유함 → (예) 개별 조사의 ‘혼인 상태’ 변수는 제각각일 수 있지만, 사회 전반의 ‘혼인 상태’라는 본질은 동일함. 따라서 이 단계의 개념은 모든 가능한 변수 변형을 포괄해야 하므로 매우 일반적임 • 개념은 변수 및 단위유형뿐 아니라, 범주로도 자주 사용됨 → (예) ‘성별’ 개념변수의 범주인 ‘남성’이 데이터 변환을 통해 ‘남성 여부’라는 개념변수로 정의되더라도 두 변수의 개념적 본질은 유지되므로 서로 비교 가능함

2-1) 특성으로서의 개념

Conceptual Variable(개념변수, 이하 CV로도 표현)은 Concept과 Unit Type(단위 유형)을 결합하여 측정할 특성을 정의한다¹⁸⁾. Concept가 ‘무엇을 의미하는지’에 초점을 둔다면, Conceptual Variable은 ‘어떤 개념적 특성을 측정할지’에 초점을 둔다. Conceptual Domain(개념 도메인, 이하 CD로도 표현)은 개념변수가 참조하는 유효한 개념들의 집합이다. 개념변수(0..*)는 항상 개념 도메인(1) 1개를 참조하며, 같은 개념이라도 개념 도메인에 따라 개념변수가 달라질 수 있다. <그림 13>에서 보듯이, 개인(people)이라는 동일한 Unit Type을 측정하더라도 참조하는 개념 도메인이 {미혼, 유배우, 이혼, 사별} 또는 {미혼, 기혼}으로 달라짐에 따라 다른 개념변수로 정의되는 것을 확인할 수 있다.

<그림 13> Variable Cascade & Value Domains



출처: UNECE Statistics, GSIM 2.0 Webinar 2025 [YouTube] (2025.9. 게시).

Represented Variable(표현변수, 이하 RV로도 표현)은 개념변수의 속성을 상속받으면서 그 측정값에 대한 표현방식이 더 구체화된 정보 클래스이다. 따라서 표현변수는 항상 1개의 Substantive Value Domain(실질적 값 도메인)을 참고하게 된다. 실질적 값 도메인은 주제와 관련된 실질값을 포함하는 값 도메인이고, 이와 상반되는 Sentinel Value

18) 다만 실제 통계기관에서 제공하고 있는 개념변수에 대한 정보는 추상적 수준에만 머무르는 것만은 아니며, 실제 측정 및 판정에 필요한 구체적인 정보가 함께 기술될 수 있음을 제3절에서 확인함

Domain(센티널 값 도메인)은 주제와 상관없이 수집·처리와 관련하여 데이터의 누락 사유(예: 응답거부, 알수없음, 마스킹 등)를 나타내는 특수값¹⁹⁾을 포함한다. 값 도메인은 맥락에 따른 유형 외에도, 표현 방식에 따라 Enumerated Value Domain(열거형 값 도메인), Described Value Domain(서술형 값 도메인)으로 구분할 수 있다. 열거형은 성별 코드, 표준산업분류체계와 같이 값을 나열하는 방식이고, 서술형은 {0 이상의 정수}와 같이 표현식으로 정의하는 방식이다. 이들의 상위 유형인 Value Domain(값 도메인)은 개념변수에 허용가능한 값들의 집합으로, 데이터유형(명목, 서열, 구간 등)과 측정단위(예: 년, 톤, 건수) 정보를 포함할²⁰⁾ 수 있다.

<표 9> Value Domain의 하위 유형

구분 기준	정보 클래스	정의	예시
주제 관련성	Substantive Value Domain	주제와 관련된 실질값을 포함하는 값 도메인	▪ {0~100 사이의 정수} ▪ {1^{남자}, 2^{여자}}
	Sentinel Value Domain	주제와 상관없이 처리와 관련된 특수값을 포함하는 값 도메인	▪ {M^{누락}, R^{거부}, E^{오류}}
표현방식	Enumerated Value Domain	범주 및 관련 코드목록으로 표현된 값 도메인	▪ {1^{남자}, 2^{여자}}
	Described Value Domain	표현식으로 정의된 값 도메인	▪ {0~100 사이의 정수}

Instance Variable(인스턴스 변수)은 표현변수가 데이터세트에서 실제 구현된 것으로, 데이터세트마다 새로운 인스턴스 변수가 생성된다. DB의 컬럼에 해당하며, 어플리케이션(R, SAS 등) 데이터유형 및 센티널 값 도메인을 포함한다.

Concept보다 상위 클래스인 Concept System(개념 체계)은 개념들 간의 관계에 의해

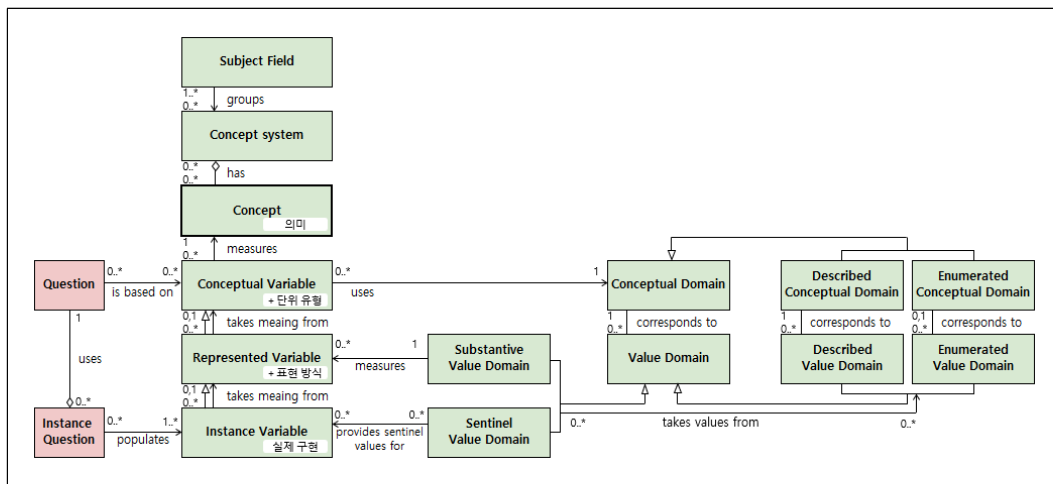
19) HLG-MOS Work Programme for 2026: ‘센티널 값은 데이터 값의 수집 및 처리와 관련되어 해당 값이 누락된 이유(예: 응답자가 거부함)를 나타내는 값을 의미합니다(알 수 없음; 기밀 유지상의 이유로 생략됨; 불완전한 데이터 처리 단계를 위한 자리 표시자). 본 연구의 목적은 센티널 값을 정의하는 공통 범주 집합 개발이 실현 가능하고 유용한지 탐색하고, 가능하다면 결측값과 거부된 값의 다양한 인코딩을 포함하는 결측값 범주를 생성하는 것입니다. 따라서 본 연구는 다음과 같은 목적을 가집니다. 상호 운용성(예: SDMX 및 DDI)을 향상시키고 여러 기관 간의 일관성을 개선하여 사용자가 센티널 값을 더욱 정확하게 해석할 수 있도록 지원합니다. 2026년에는 자원 제약이 예상되므로 사무국은 이 활동에 최소한의 지원만 제공할 예정입니다.’

20) GSIM 2.0 User Guide(66번 단락): The Value Domain includes data type and unit of measurement information. The data type contains information on the allowed computations one may perform on the Datum (nominal-, ordinal-, interval-data, etc.), while the unit of measurement (Years, Tonnes, Count, Dollars, etc.) refines the measurement of the Value Domain. For example, gender codes lead to nominal statistical data, whereas age values in years lead to interval data.

구조화된 개념들의 집합이다. GSIM 용어집에서는 ‘남성’, ‘여성’, ‘기타’ 개념을 묶은 ‘성별 개념’ 및 ISIC(국제표준산업분류)를 예시로 설명한다.

Concept System보다 상위에 있는 Subject Field(주제 영역)는 일련의 개념(Concepts)과 그 명칭(Designations)이 사용되는 지식 분야로 정의된다. ISO 11179-1에서는 주제 영역이 세분될 경우, 세분된 결과 역시 각각 하나의 주제 영역이 된다고 설명한다.²¹⁾

<그림 14> 특성으로서의 개념 역할과 관련된 정보 클래스 도식



출처: UNECE, GSIM Version 2.0 - 그룹 다이어그램 [웹사이트] (2025.12 조회) 재구성.

2-2) 대상(Unit Type, Population, Universe)으로서의 개념

개념의 두 번째 역할은 정보가 다루고자 하는 대상(객체, 사건, 엔터티)들의 집합을 설명하는 Unit Type, Population, Universe로서의 역할이다.²²⁾ 본 연구에서는 편의상 ‘대

21) ISO 11179-1, 3.2.2 domain (유의어: subject field): 특정 지식 영역(field of special knowledge)을 의미하며, domain이 세분될 때 세분된 결과 역시 각각 하나의 domain이 된다. [SOURCE: ISO 1087:2019, 3.1.4]

22) GSIM 2.0 User Guide(58번 단락): At an abstract level, a Concept is defined in GSIM as a ‘unit of thought differentiated by characteristics’. Concepts are used in different ways throughout the statistical lifecycle, and each different role of a Concept is described using a different information class (which are subtypes of Concept). A Concept can be used in these situations:

- As a characteristic to be measured (e.g., ‘gender’, ‘value of production’)
- As a Unit Type, a Population, or a Universe to describe a set of objects, events, or entities that information is about (e.g., the Population of adults in Netherlands or the Universe of adults with a university degree, based on the Unit Type or persons)
- As a Category to further define details about a Concept (e.g., ‘Male’ and ‘Female’ for the Concept ‘gender’). Codes can be linked to a Category via a Node (i.e., a Code Item or Classification Item), for

상으로서의 개념'으로 약칭한다. Unit Type(단위 유형)은 Unit(단위)을 그룹화하는 단일 특성이 되는 추상적 개념으로 구분한 것이다²³⁾. Unit(단위)는 정보수집 대상이자 궁극적으로 통계자료가 작성되는 개체이다. Universe(유니버스)는 단위 유형에 주제적 특성에 따른 맥락을 부여한 것인데, 시·공간적 요소는 포함하지 않는다. Population(모집단)은 유니버스에 시·공간적 조건을 부여한 것이다. User Guide(2024)에서는 목표 모집단, 조사 모집단, 표본추출틀, 분석 모집단 등 프로세스 단계에 따라 여러 종류의 Population이 있다고 설명한다. 앞서 변수 연쇄처럼, Unit Type에서, Universe 및 Population으로 점점 구체화되는 계층적 구체화가 시현된다.

<표 10> Unit Type, Universe, Population, Unit 정보 클래스 예시

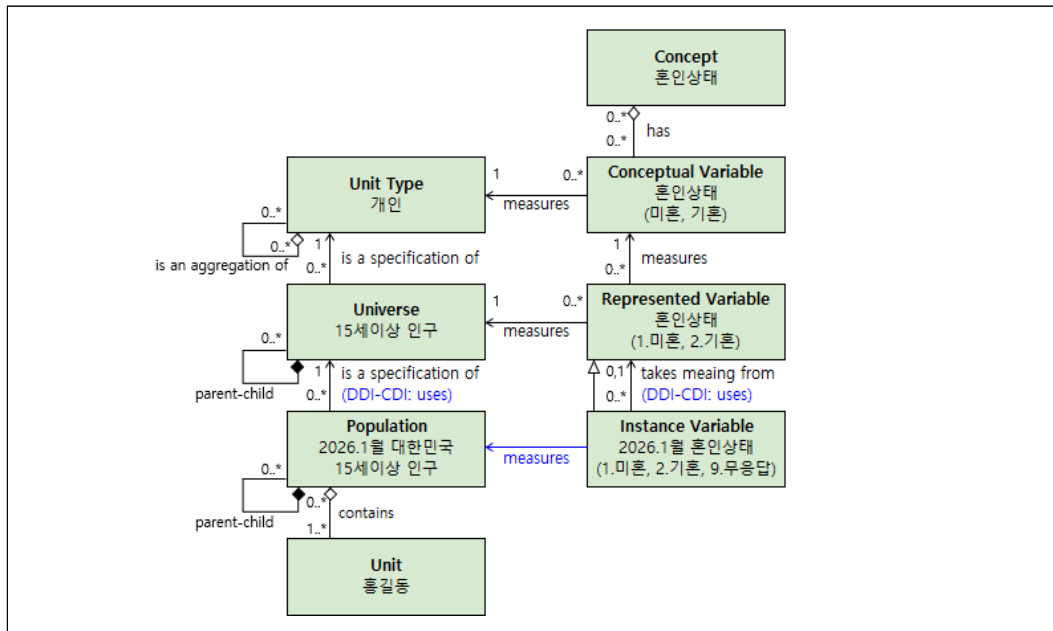
Unit Type	Universe	Population	Unit
• 개인 • 사업체 • 지자체	• 대학학위 소지자 • 15세이상 취업자 • 취업자가 아닌 15세이상 인구	• '25.11.1. 한국 국민 • '26년말 미국 제조업체 • '23.1.1. 전 세계 대학교	• 홍길동, 이순신, ... • 구글, 아마존, ... • 서울대, 충남대, ...

Unit Type, Universe 및 Population은 모두 자기 자신을 참조하는 관계가 있으나, 그 관계의 성격은 서로 다르다. Unit Type은 빈 마름모로 표현되는 'is an aggregation of' 관계가 있다. 예컨대, '가구'라는 단위 유형이 '개인' 단위 유형의 집합으로 정의될 수 있으며, 이때 '개인'은 '가구'와 독립적인 단위 유형으로 존재할 수 있다. 반면, Universe는 채워진 마름모(composition)로 표현되는 'parent-child' 관계가 있다. 예컨대, '만 15세 이상 인구'라는 상위 유니버스에 조건을 추가하여 '취업자'라는 하위 유니버스를 정의할 수 있으며, 이때 '취업자' 유니버스는 상위 유니버스와 무관하게 독립적으로 재정의될 수 없다. Population도 Universe와 마찬가지로 강한 parent-child 관계를 가질 수 있다.

use within a Code List or Statistical Classification).

- 23) Unit Type 용어집 해설: 단위 유형은 시간이나 지리적 범위를 명시하지 않고 단일 특성을 기준으로 단위들의 집합 또는 그룹을 설명하는 데 사용된다. 예를 들어, '사람'이라는 단위 유형은 "사람"이라는 특성을 가진 단위들을 하나의 그룹으로 묶은 것이다. 이는 통계 자료 배포에 사용되는 단위 유형 뿐만 아니라 통계 처리 과정 전반에 걸쳐 적용된다. 예를 들어, 행정자료를 사용할 때 재정 단위(fiscal unit)를 활용할 수 있다.

<그림 15> 대상으로서의 개념 역할과 관련된 정보 클래스 도식



* 파란색 연결선은 GSIM 다이어그램에는 명시되지 않았으나, DDI-CDI 모델 사양서에서 설명하는 '변수 계층구조'와 '단위 계층구조'의 수평성을 반영

출처: UNECE, GSIM Version 2.0 - 그룹 다이어그램 [웹사이트] (2025.12 조회); DDI Alliance(2025), DDI-CDI Specification Overview를 참고하여 재구성

2-3) 범주로서의 개념

Category(범주)는 특성을 외연적으로 정의하고 측정하는 역할을 하는 개념이다. 예컨대, 삼각형을 정의할 때, '삼각형이란 정삼각형, 이등변삼각형, 직각삼각형, 둔각삼각형 등을 통칭한다.'하고 정의하는 방식이 외연적 정의이고, '삼각형이란 동일 직선 위에 있지 않은 세 개의 점을 선분으로 연결한 평면도형이다.'라고 정의하는 방식은 내포적 정의에 해당한다.

Node(노드)²⁴⁾는 Category와 관련 속성의 결합으로, 총 3가지 하위 유형이 있다. Category Item(범주 항목)은 표현(코드) 없이 범주의 의미만 포함하며, Code Item(코드 항목)은 표현(코드)와 함께 범주의 의미를 포함하고, Classification Item(분류 항목)은 특정 레벨에서 범주와 표현을 결합한다. Designation(명칭)은 개념을 표기하는 기호적 명칭이며 Code(코드)는 그 하위 유형으로 범주에 대한 Designation에 해당한다.

24) Node 클래스 속성:

Name	Description	차수	값유형
Aggregation Type	노드 간의 parent/child 관계가 part/whole인지, super/sub인지 정의	0..1	String

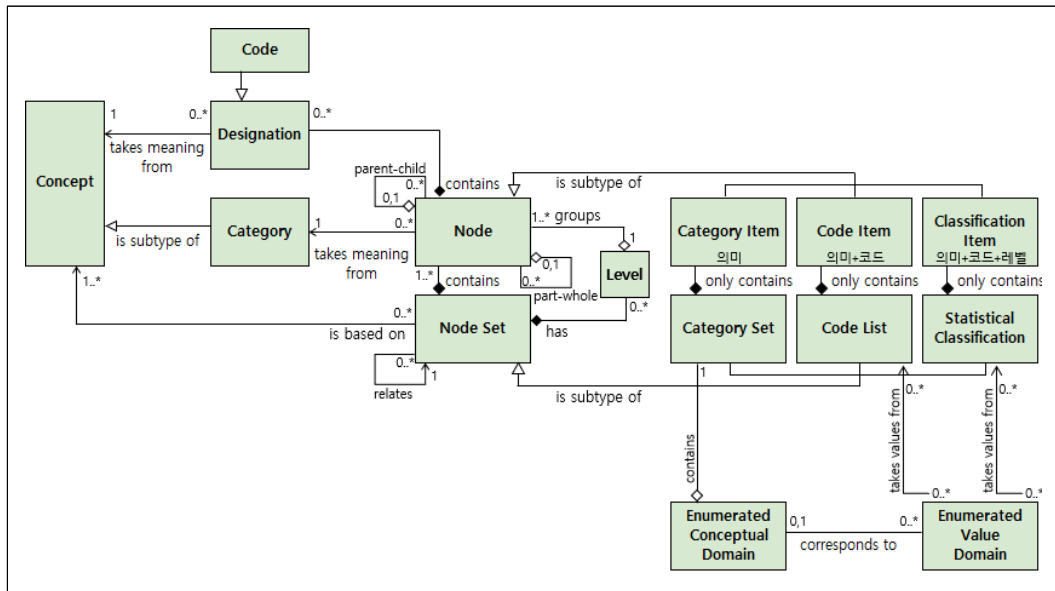
Node Set(노드세트)는 이러한 노드들의 집합으로, 마찬가지로 3가지 하위 유형을 가진다. Category Set(범주세트)는 범주항목을 통해 범주들을 그룹화하는 노드세트 유형이다. Code List(코드목록)은 코드항목을 통해 범주와 그 코드 쌍을 그룹화하는 노드세트 유형으로, 열거형 값 도메인에서 사용할 수 있는 허용값 집합을 사전 정의하여 제공하는 역할²⁵⁾을 한다. Statistical Classification(통계분류체계)은 동일·유사한 특성을 공유하는 범주들을 상호 배타적이고(중첩 없고) 포괄적으로(누락 없이) 계층화한 집합으로, 한국표준산업분류, 한국표준교육분류 등이 해당한다.

두 통계분류체계를 비교했을 때, 원천 통계분류체계의 분류항목과 대상 통계분류체계의 해당 분류항목 간의 관계를 나타내는 표현(expression)이 Map(맵) 정보 클래스이며, 두 통계분류체계의 분류항목 간에 설정된 Map들의 집합은 Correspondence Table(대응표)에 해당한다. 대응표와 Map의 사용은 통계분류체계에 국한되지 않고 모든 유형의 노드와 노드세트로 확장할 수 있다. 즉, 대응표를 통해 코드목록, 범주세트의 항목 간에도 매핑이 가능하다.

통계분류체계와 관련된 정보 클래스는 Classification Family, Classification Series, Classification Index, Classification Index Entry 등이 있다. 국내 통계분류체계는 이미 체계적으로 관리되고 있어서, 이와 관련된 정보 클래스는 본 연구에서 상세히 다루지 않는다.

25) Code List 클래스 설명문: 노드세트에서 상속받은 ‘관련됨(related to)’ 관계를 통해, 유사한 코드목록들을 그룹화될 수 있다. 코드목록은 열거형 값 도메인에서 사용할 수 있는 허용값들의 집합을 사전에 정의하여 제공한다. (Similar Code Lists can be grouped together (via the relates to relationship inherited from Node Set). A Code List provides a predefined set of permissible values for an Enumerated Value Domain.)

<그림 16> 범주로서의 개념 역할과 관련된 정보 클래스 도식



* Correspondence Table, Map 및 Classification Family, Classification Series, Classification Index, Classification Index Entry 정보 클래스는 재구성 도식에 미포함

출처: 출처: UNECE, GSIM Version 2.0 - 그룹 다이어그램 [웹사이트] (2025.12 조회) 재구성.

제4절 소결

본 장에서는 GSIM 2.0의 개발 배경과 다른 국제표준과의 관계를 살펴보고, 기초·비즈니스·교환·구조·개념의 5개 그룹으로 구성된 정보 클래스 체계를 검토하였다. 특히 본 연구의 주된 관심 대상인 개념 그룹은 39개의 정보 클래스로 구성되며, Concept의 역할이 ① 측정할 특성, ② 대상으로서의 집합 범위, ③ 외연적 정의를 위한 범주라는 세 가지로 구체화됨을 확인하였다.

특히, 변수 연쇄(Variable Cascade) 구조는 설계와 구현을 분리함으로써 개념의 재사용성과 조사별 차별성을 동시에 확보하는 GSIM의 핵심 설계 원리임을 확인하였다. 이는 국가통계 개념 정보 관리를 위해 어떤 정보를 공통 기준(참조 저장소)으로 관리하고, 어떤 정보를 개별 통계 프로그램의 몫으로 남겨둘 것인지를 판단하는 이론적 근거가 된다.

이러한 정보모델 구조에 대한 이해를 토대로, 제3장에서는 캐나다 통계청과 노르웨이 통계청에서 이 구조를 실제로 어떻게 적용하고 있는지 살펴보고, 제4장에서는 이를 국내 통계 환경에 어떻게 적용할 수 있을지를 검토한다.

제 3 장

해외 사례

제1절 캐나다 통계청

1. 표준에 관한 정책

캐나다는 통계 프로그램에서 사용되는 모집단, 통계 단위, 개념, 변수 및 분류에 대해 표준 명칭과 정의를 사용하기 위해 2004년 7월 14일 ‘표준에 관한 정책(Policy on standards)’을 수립²⁶⁾하였다. 각 표준은 1) 필요시 표준이 다루는 개념을 설명할 수 있어야 하고, 2) 적용되는 통계 단위를 식별할 수 있어야 하며, 3) 표준에 포함되는 각 변수의 이름과 정의를 제공해야 하고, 4) 각 변수에 대한 데이터 편찬(compilation) 및 배포에 사용되는 분류체계를 제공해야 한다. 표준의 적용 의무 수준은 <표 11>와 같이 3가지로 구분하며, <그림 18>과 같이 표준마다 적용 의무 수준을 함께 제공하고 있다.

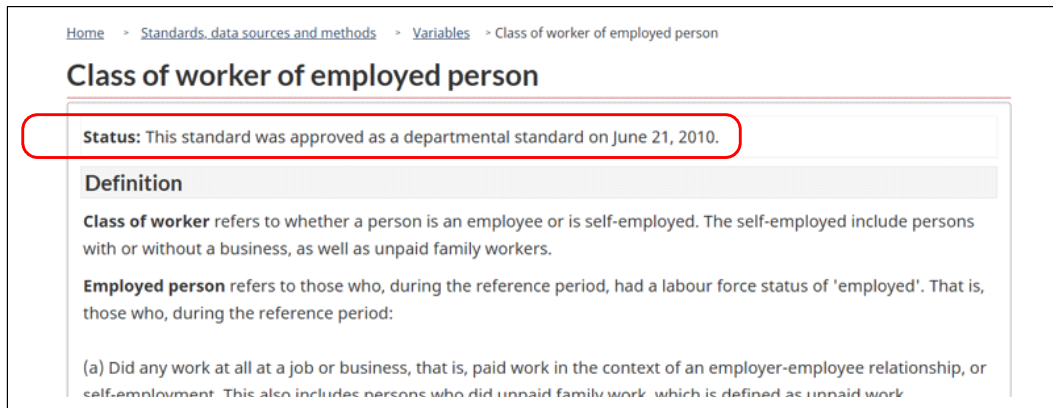
<표 11> 적용 의무 수준에 따른 표준 종류

표준 종류	의무 수준
부처 표준 (departmental standard)	정책위원회(Policy Committee)의 승인을 받은 표준으로, 본 정책의 조항에 따라 명시적으로 면제를 받은 경우를 제외하고는 적용이 강제됨
권고 표준 (recommended standard)	방법·표준 위원회(Methods and Standards Committee)에서 권고 표준으로 인정한 표준으로, 일정 기간의 시범 적용 기간을 둘 수도 있으며, 해당 기간 경과 이후 부처 표준으로 지정될 수 있음
프로그램별 표준 (program-specific standard)	통계 프로그램에 의해 채택되어 표준부서(Standards Division)에 등록된 표준으로, 시간 경과에 따른 시계열 일관성을 확보하기 위한 표준

출처: Statistics Canada, Policy on standards (revised 2004) [웹사이트].

26) 웹 페이지에 제공하는 본문의 최종 수정일은 2025년 10월 31일임.

<그림 18> 캐나다 표준 의무 수준 표기 사례



출처: Statistics Canada, Surveys and statistical programs [웹사이트].

캐나다 통계청은 메타데이터의 적절한 관리를 보장하기 위해 성숙하고 효과적인 거버넌스 및 관리구조를 갖추고 있다. 핵심 주체는 표준부서, 방법·표준 위원회, 프로그램 영역의 세 기관이다. 표준부서(Standards Division)는 통계 표준과 메타데이터의 개발·관리·확산을 담당하는 실무 총괄 부서이다. 방법·표준 위원회(Methods and Standards Committee)는 국장급으로 구성되고 통계청 차장이 의장을 맡는 의사결정 및 자문기구로, 중앙 메타데이터(Corporate Metadata)를 총괄 관리하고, 통계 개념, 변수, 분류를 기관 표준으로 승인하며, 필요한 경우 기관 표준 적용 예외도 승인한다. 또한, 표준 정책 준수 현황을 정기적으로 보고받고 필요시 정책 및 관련 표준에 대한 검토를 시작한다. 마지막으로, 프로그램 영역(Program Areas)은 통계 생산 및 조사를 수행하는 현업 부서로, 표준을 준수하고 메타데이터를 직접 작성하고 문서화할 책임이 있다.

2. 캐나다 Standards, data source and methods 조회 서비스

캐나다 통계청은 IMDB(Integrated Metadata Base)를 중심으로 기관 전체 메타데이터를 통합 관리하고 있으며, 조사별 메타데이터뿐 아니라 개념, 변수, 방법론 등 다양한 메타데이터를 통합 관리하고 있다. IMDB는 외부 이용자뿐 아니라 통계 생산 내부에서도 재사용되는 Corporate Metadata Repository이다. 이러한 중앙 집중식 메타데이터 저장소를 구축함으로써 예상하지 못한 효과를 포함한 다양한 편익을 비용 효율적으로 달성할 수 있었으며, 캐나다 통계청의 통계 프로그램 전반에 걸쳐 메타데이터의 품질과 활용성을 향상시키는 효과가 있었다고 설명한다(Cloutier and Born, 2010).

외부 이용자는 캐나다 통계청 누리집 ‘Reference’ 메뉴 하위의 ‘Standards, data sources and methods’ 메뉴에서 조회할 수 있다. 표준 개념과 개별 변수를 조회할 수 있는 메뉴 경로는 상이하다. 표준 개념은 ‘Find classifications, variables and statistical units’ 소메뉴에서 일괄 조회할 수 있으며, Statistical Unit(통계 단위) 33개, Classification(분류) 226개, Variable(변수) 162개가 등록되어 있다. 반면, 개별 변수는 ‘Surveys and statistical programs’ 소메뉴에서 개별 통계 프로그램을 찾아 선택한 후에 조회할 수 있다. 최신 참조기간을 기준으로 조회되므로 <그림 20>에서 보이는 ‘Other reference periods’ 탭을 선택하면 해당 조사 주기에 정의된 변수 목록을 조회할 수 있다.

<그림 19> 캐나다 표준 개념 조회 화면

Home > Reference > Standards, data sources and methods

Classifications, variables and statistical units

Browse our central repository of standard classifications, variables and statistical units.

According to Statistics Canada's [Policy on standards](#), a standard must include a statement regarding the degree to which its application is compulsory. More details can be found at [Is your standard compulsory?](#)

Filter options

Use filters to change the focus of your results in the following list.

Data type **Subject**

All types

Filter items Showing 1 to 18 of 18 entries (filtered from 425 total entries) | Show entries

Title	Subject	Type
Absences from work of employed person	Labour	Variable
Class of worker - variant on employees and self-employed	Labour	Classification
Class of worker of employed person	Labour	Variable
Commuting destination of employed person	Labour	Variable
Commuting duration of employed person	Labour	Variable
Commuting vehicle occupancy of employed person	Labour	Variable
Distance (straight-line) from home to work of employed person	Labour	Variable
Employed person	Population and demography; Labour	Statistical unit
Employed Status	Labour	Classification
Industry of employed person	Labour	Variable

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

<그림 20> 캐나다 통계 프로그램별 변수 조회 화면

Home > Standards, data sources and methods > Surveys and statistical programs

Labour Force Survey (LFS)

Labour Force Survey (LFS)	Variable(s)	Summary of changes	Other reference periods
Variables for April 2026			
• Absences from work of employed person, category • Absences from work of employed person, length • Absences from work of employed person, ratio • Absences from work of employed person, type • Activity index of economy, category • Activity index of economy, index • Activity prior to unemployment of person, category • Actual hours worked of employed person, category • Actual hours worked of employed persons, number • Age at retirement of person, measure			

출처: Statistics Canada, Surveys and statistical programs [웹사이트].

1) Subject

캐나다 통계청은 총 25개의 주제별로 표준 정보를 제공한다. 통계표 서비스에서 제공하는 일부 주제에 대해서는 제공되는 표준 정보가 없다. 특이할 점은 ‘Employed person’이라는 통계 단위와 같이 인구, 노동 주제에 모두 관련되는 경우는 해당 주제에 모두 포함되는 것으로 나타내고 있다.²⁷⁾

27) 2개 이상의 주제에 속하는 분류, 통계단위, 변수 내역

구분	Subject	표준 명칭
분류 (9개)	아동청소년; 가족가구	Child presence / Stepfamily status
	건설; 주거	Period of construction
	가족가구; 사회공동체	Marital status
	주거; 소득	Shelter cost to income ratio
	다양성; 인구	Population group
	국제무역; 경제계정	Variant of NAPCS Canada 2017 Version 1.0 - Merchandise import and export accounts
	노동; 고령인구	Retirement
	사회공동체; 가족가구	Marital status - short title variant for dissemination
	가족가구; 주거	Owner household / Private household / Renter household
통계 단위 (21개)	주거; 건설	Building / Building unit / Collective dwelling / Dwelling
	인구; 가족가구	Couple / Couple family / Couple family with children / Economic family / Household / Lone-parent family / Person not in a census family / Person not in economic family
	인구; 다양성	Foreign resident / Immigrant / Non-permanent resident
	인구; 노동	Employed person / Experienced labour force person / Person with recent work experience
	아동청소년; 가족가구	Child presence in census family 등 2개
변수 (7개)	주거; 소득	Presence of mortgage payments of owner household 등 4개
	노동; 고령인구	Retirement of person aged 55 or over

<표 12> 캐나다 표준이 제공되는 Subject 목록

연번	Subject
1	Population and demography (인구, 인구통계)
2	Children and youth (아동, 청소년)
3	Older adults and population aging (고령인구, 고령화)
4	Families, households and marital status (가족, 가구, 혼인상태)
5	Immigration and ethnocultural diversity (이민, 민족문화적 다양성)
6	Indigenous peoples (원주민)
7	Society and community (사회, 공동체)
8	Housing (주거)
9	Languages (언어)
10	Education, training and learning (교육, 훈련, 학습)
11	Labour (노동)
12	Health (건강)
13	Income, pensions, spending and wealth (소득, 연금, 지출, 자산)
14	Prices and price indexes (가격, 가격지수)
15	Economic accounts (경제 계정)
16	Digital economy and society (디지털 경제, 사회)
17	Business and consumer services and culture (기업 서비스, 소비자 서비스, 문화)
18	Agriculture and food (농업, 식품)
19	Manufacturing (제조업)
20	Construction (건설)
21	Energy (에너지)
22	Business performance and ownership (기업활동, 소유)
23	Science and technology (과학, 기술)
24	International trade (국제무역)
25	Environment (환경)

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

2) Statistical unit

캐나다 통계청은 총 33개 표준 Statistical unit(통계 단위)를 지정하고 있다. ‘Person’, ‘Person 15 year or over’와 같은 사례를 보면, 캐나다 통계청의 Statistical unit은 GSIM의 Unit Type에만 국한되기보다 맥락이 부여된 정보 클래스인 Universe까지 포괄하는 것으로 보인다. 한편, 33개 통계 단위 중에서 단일 주제에 해당하는 경우는 12개에 불과하고, 나머지 21개는 2개 주제에 복합적으로 해당하는 것으로 제공하고 있다.

<표 13> 캐나다 표준 Statistical unit

Subject		개수	Statistical unit
계		33	
단 일	Population and demography	7	· Census family · Female · Geographic area · Person · Person 15 years or over · Person aged 55 or over · Stepfamily
	Business performance and ownership	4	· Company · Enterprise · Establishment · Location
	Housing	1	· Private dwelling
복 합	Population and demography ; Families, households and marital status	8	· Couple · Couple family · Couple family with children · Economic family · Household · Lone-parent family · Person not in a census family · Person not in economic family
	Population and demography ; Labour	3	· Employed person · Experienced labour force person · Person with recent work experience
	Population and demography ; Immigration and ethnocultural diversity	3	· Foreign resident · Immigrant · Non-permanent resident
	Families, households and marital status ; Housing	3	· Owner household · Private household · Renter household
	Housing ; Construction	4	· Building · Building unit · Collective dwelling · Dwelling

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

통계 단위별로 의무 수준 및 제공 정보는 일부 차이가 있다. 대부분은 부처 표준(departmental standard)으로 지정되어 있으나, ‘Building’은 권고 표준(recommended standard)으로 지정되어 있다. 통계 단위에 대한 정보는 대부분 정의, 이전 버전과의 관계, 또는 용법 정도를 제공하고 있으나, ‘Building’처럼 관련 국제 기준이 있는 경우에는 함께 제공한다.

<표 14> Statistical Unit 사례별 제공 정보 비교(Person, Building)

Statistical Unit		Person	Building
제 공 정 보	Status	Departmental standard	Recommended standard
	Definition	O	O
	Usage	X	O
	Conformity to relevant internationally recognized standards	X	O
	Relation to previous version	O	O

<그림 21> 캐나다 Statistical Unit 조회 화면 예시(Person, Building)

Person
Status: This standard was approved as a departmental standard on October 22, 2007.
Definition
Person refers to an individual and is the unit of analysis for most social statistics programs.
Relation to previous version
Person October 22, 2007 to current This is the current standard.
Building
Status: This standard was approved as a recommended standard on July 17, 2018.
Definition
Building refers to a roofed independent free-standing permanent structure usually enclosed within external walls or dividing walls that extend from the foundations to the roof and comprises one or more rooms or other space. A building may be used or intended for residential, commercial, industrial or institutional purposes, including the provision of services. A building can be entered by persons or animals and is suitable or intended for protecting them and objects. For technical reasons, "building" also includes a separately usable underground construction.
Usage
A building should have its own unique civic address. If such an address is not available, a combination of the building name, spatial coordinates, a picture and/or a detailed unique description can be used to identify the building.
Conformity to relevant internationally recognized standards
This standard is compatible with the definition of building presented in the United Nations Principles and Recommendations for Population and Housing Censuses Revision 3, 2017. This standard is also compatible with the definition of building presented in the Eurostat Classification of types of constructions, 1997.
Relation to previous version
Building July 17, 2018 to current This is the current standard. Building January 01, 2018 to July 16, 2018 This version was replaced on July 17, 2018.

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

3) Classification

표준 Classification(분류)은 총 226개가 등록되어 있다. 표준 분류가 많이 지정된 주제(단일 주제 기준)는 노동 34개, 교육·훈련·학습 28개, 가족·가구·혼인상태 26개, 인구·인구통계 23개 순이다. <표 15>에서 노동 단일 주제에 대한 34개 표준 Classification을 보면, 캐나다 통계청에서는 GSIM의 Code List(코드목록)와 Statistical Classification(통계 분류체계)을 통칭하여 Classification으로 제공하고 있는 것으로 보인다.

<표 15> 캐나다 표준 Classification (노동 단일 주제)

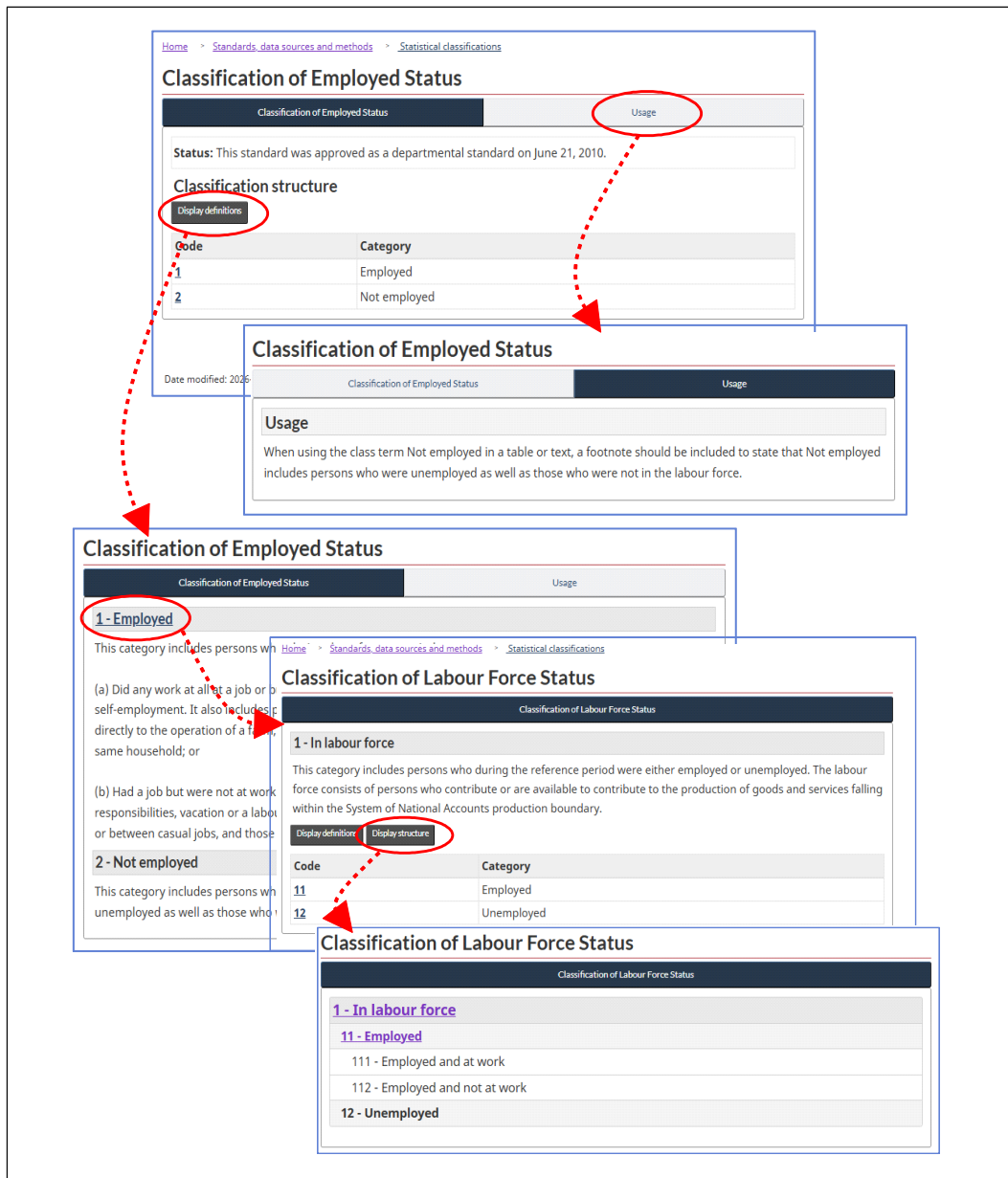
연번	분류명	
1	Class of Worker	종사상 지위
2	Class of worker - variant on employees and self-employed	종사상 지위 - 직원과 자영업자의 변형
3	Commuting destination	통근 목적지
4	Commuting vehicle occupancy	통근 차량 탑승자
5	Duration of commute	통근 소요 시간
6	Employed Status	고용 상태
7	Full-time and Part-Time Work Hours	정규직 및 파트타임 근무 시간
8	Full-time or part-time weeks worked during the reference year	기준 연도 동안 근무한 정규직 또는 파트타임 주 수
9	Hours of activity per week	주당 활동 시간
10	Labour Force Status	노동력 현황
11	Labour Force Status - Collapsed classification	노동력 현황 - 통합 분류
12	Main reason for absence from work (main reason for time lost)	결근 사유 (근무 시간 손실의 주요 원인)
13	Main reason for not working the full year	연중 내내 근무하지 않는 주된 이유
14	Main reason for work interruptions	업무 중단의 주요 원인
15	Main reason for working mostly part time	주로 파트타임으로 일하는 주된 이유
16	Military service status	군 복무 상태
17	Mode of commuting	통근 방식
18	Place of work status	근무지 현황
19	Primary activity status	주요 활동 상태
20	Recipient of unpaid service	무상 서비스 수혜자
21	Source(s) of financial compensation received during absence from work	결근 동안 받은 재정적 보상의 출처
22	Time arriving at work	출근시간(직장 도착)
23	Time leaving for work	출근시간(집에서 출발)
24	Unpaid care	무급 간병
25	Usual Weekly Work Hours	통상 주간 근무 시간
26	Usual Weekly Work Hours - Collapsed classification	통상 주간 근무 시간 - 분류 간소화
27	North American Industry Classification System (NAICS) Canada 2022 Version 1.0	북미산업분류시스템(NAICS) 캐나다 2022 v1.0
28	Variant of the NAICS 2022 Version 1.0 for Labour force	북미산업분류시스템(NAICS) 2022 v1.0의 노동력 부문 변형
29	National Occupational Classification (NOC) 2021 Version 1.0	국가직업분류(NOC) 2021 v1.0
30	Variant of the NOC 2021 Version 1.0 for Analysis by TEER (Training, Education, Experience and Responsibility) categories	TEER(훈련·교육·경험·책임) 범주별 분석을 위한 국가직업분류(NOC) 2021 v1.0 변형
31	Variant of NOC 2016 Version 1.0 - Highly aggregated data	국가직업분류(NOC) 2016 v1.0 변형 - 대분류
32	Weeks worked during the reference year	기준 연도 동안 근무한 주 수
33	Work activity during the reference year	기준 연도 동안의 근로 활동
34	Work interruption in the reference period	기준 기간 중 업무 중단

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

Classification마다 제공하는 정보량은 차이가 있다. Usage가 등록된 경우에는 탭이 활성화되고, 범주에 대한 정의가 필요한 경우에는 [Display definitions] 버튼이 호출된다.

범주가 단일 레벨이 아닌 경우에는 [Display structure] 버튼이 호출된다. 하위 범주가 있는 경우, 전체 코드목록이 한 번에 보이진 않고, 계속된 클릭을 통해 최하위 범주까지 도달해야 한다. 범주별 정의가 제공되는 점에서는 유용하나, 전체 코드 구조를 한 번에 조회하기에는 어려운 정보 제공 방식이다.

<그림 22> 캐나다 Classification 조회 화면 예시



출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

위에서 사례로 든 ‘Classification of Labour Force Status’의 각 레벨별 코드와 범주별 정의를 보면, 범주 정의가 조작적 정의에 가까운 것을 확인할 수 있다.

<그림 23> 캐나다 ‘Labour force status’ 코드 및 범주 정의

1 - In labour force This category includes persons who during the reference period were either employed or unemployed. The labour force consists of persons who contribute or are available to contribute to the production of goods and services falling within the System of National Accounts production boundary.
11 - Employed This category includes persons who, during the reference period: (a) Did any work at all at a job or business, that is, paid work in the context of an employer-employee relationship, or self-employment. It also includes persons who did unpaid family work, which is defined as unpaid work contributing directly to the operation of a farm, business or professional practice owned and operated by a related member of the same household; or (b) Had a job but were not at work due to factors such as their own illness or disability, personal or family responsibilities, vacation or a labour dispute. This category excludes persons not at work because they were on layoff or between casual jobs, and those who did not then have a job (even if they had a job to start at a future date).
111 - Employed and at work This category includes employed persons who did any work at all at a job or business during the reference period.
112 - Employed and not at work This category includes employed persons who had a job but were not at work during the reference period due to factors such as their own illness or disability, personal or family responsibilities, vacation, or a labour dispute. This category excludes persons who were not at work during the reference period because they were on layoff or between casual jobs, and those who did not then have a job (even if they had a job to start at a future date).
12 - Unemployed This category includes persons who during the reference period: - Were without work but had looked for work in the past four weeks ending with the reference period and were available for work; or - Were on temporary layoff due to business conditions and were available for work; or - Were without work, had a job to start within four weeks from the reference period and were available for work. Usage Note for Unemployed: The Labour Force Survey uses more questions to determine temporary layoff status and availability for work than other Statistics Canada surveys since the main objective of the survey is to classify the working-age population into three groups (employed, unemployed and not in the labour force) and to report on them on a monthly basis.
2 - Not in labour force This category refers to persons who were neither employed nor unemployed during the reference period. This includes persons who, during the reference period were either unable to work or unavailable for work. It also includes persons who were without work and who had neither actively looked for work in the past four weeks nor had a job to start within four weeks of the reference period.

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

아래 <표 16>에서는 Classification의 세 가지 특징을 확인할 수 있다.

먼저, 축약(collapsed) 버전이 표준에 포함된다. 사례2는 사례1의 서로 다른 레벨의 범주(‘11’, ‘12’, ‘2’)를 단일 위계(‘1’, ‘2’, ‘3’)로 축약했다. 사례5는 사례4의 7개 범주를 통합하여 2개 범주로 축약했다.

두 번째, 축약되었더라도 최종 개념 도메인 맥락에 따라 완전히 다른 통계분류명의 표준으로 지정되는 사례도 있다. 사례3의 범주 정의 및 Usage에 따르면 ‘Unemployed’는 ‘Not in labour force’까지 포함하는 범주이다. 즉, 범주 도메인을 조정하면서, 사례1과 사례2는 노동력상태 통계분류, 사례3은 취업상태 통계분류로 통계분류명을 지정했다.

마지막으로, 범주의 정의 제공은 필수는 아니다. 사례4와 사례5의 ‘평소 주간 취업시간’과 같이 범주 명칭만으로 설명이 가능한 경우에는 정의가 제공되지 않는다.

<표 16> 캐나다 통계청 Classification 사례별 제공 정보 비교

사례번호		사례1		사례2		사례3		사례4		사례5	
Classification		Classification of Labour Force Status		Collapsed ¹⁾ Classification of Labour Force Status		Classification of Employed Status		Classification of Usual Weekly Work Hours		Collapsed ²⁾ Classification of Usual Weekly Work Hours	
제 공 정 보	Status	Departmental standard									
	Classification structure	코드	범주					코드	범주	코드	범주
		1	In labour force	코드	범주	코드	범주	1	Less than 15h	1	Less than 30h
		11	Employed					2	15 to less than 30h	2	30h more
		111	Employed and at work	1	Employed	1	Employed	3	15 to less than 30h		
		112	Employed and not at work					4	15 to less than 30h		
		12	Unemployed	2	Unemployed			5	40h		
		2	Not in labour force	3	Not in labour force	2	Not employed	6	More than 40h to less than 50h		
							7	50h or more			
	Category definitions	O		O		O		X		X	
Usage	X		X		O ³⁾		O ⁴⁾		O ⁵⁾		

1),2) 축약 분류

3) When using the class term Not employed in a table or text, a **footnote should be included** to state that Not employed includes persons who were unemployed as well as those who were not in the labour force.

4),5) When usual weekly hours are determined for a person's main or only job, this information can be used to determine whether the person works full-time or part-time. Less than 30 hours per week is considered to be part-time and 30 hours or more is considered to be full-time.

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

4) Variable

표준 Variable(변수)은 총 162개로, GSIM의 Conceptual Variable 정보 클래스 설계를 그대로 반영하여 ‘개념 of 통계단위’와 같이 변수명을 부여한다.

<표 17> 캐나다 표준 Variable (노동 영역)

연번	변수명
1	Labour force status of person
2	Class of worker of employed person
3	Class of worker of experienced labour force person
4	Class of worker of person with recent work experience
5	Industry of employed person
6	Industry of person with recent work experience
7	Occupation of employed person
8	Occupation of experienced labour force person
9	Occupation of person with recent work experience
10	Place of work status of employed person
11	Location of workplace of employed person
12	Usual work hours of employed person
13	Multiple modes of commuting of employed person
14	Main mode of commuting of employed person
15	Commuting vehicle occupancy of employed person
16	Commuting destination of employed person
17	Commuting duration of employed person
18	Time arriving at work of employed person
19	Time leaving for work of employed person
20	Distance (straight-line) from home to work of employed person
21	Absences from work of employed person
22	Military service status of person
23	Unpaid care provided by person
24	Unpaid housework by person
25	Work activity during the reference year of person 15 years or over
26	Main reason for not working the full year of person 15 years or over
27	Main reason for working mostly part time of person 15 years or over
28	Work interruptions of person

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

Variable(변수)에 대해서는 표준의무수준(Status), 정의(Definition), 용법(Usage), 도출 방법(Derivation)/측정산식(Measurements), 관련 국제표준 준수(Conformity to relevant

Internationally recognized standards), 분류(Classifications)²⁸⁾, 추가 정보, 버전 이력 등의 정보를 제공한다.

다만, 표준 변수와 달리 통계 프로그램 변수에서는 변수명, 정의, 관련 분류 링크만 제공한다는 점에서 차이가 있으며, 그 세부 내용도 약간 차이가 있다. 우선 통계 프로그램 변수명은 측정유형이 명시된 ‘개념 of 통계단위, 측정유형’의 형태로 GSIM에서의 Represented Variable에 해당하는 것으로 이해할 수 있다. 또한 통계 프로그램 변수의 분류 링크에서는 해당 변수에서 사용하는 Classification 정보만 제공된다.

<표 18> 캐나다 표준 변수와 통계 프로그램 변수 사례 비교

변수 구분	변수명	항목명
	항목 내용	
표준 변수	Labour force status of person	Classifications
	• Employment in the calendar year prior to overdose January 01, 2023 to current • Job loss in the year prior to overdose January 01, 2023 to current • Number of years employed in the 5 years prior to overdose January 01, 2023 to current • Labour force status, NEET variant January 01, 2020 to current • Classification of labour force status, non-student and student September 01, 2019 to current • Classification of single and multiple jobholders January 01, 2016 to current • Classification of Employed Status June 21, 2010 to current • Classification of Labour Force Status June 21, 2010 to current • Collapsed Classification of Labour Force Status June 21, 2010 to current	
통계 프로그램 변수	Labour force status of person, category	Using Classifications
	• Labour Force Status • Labour force status, NEET variant	

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

<그림 24>은 ‘개인의 노동력 상태’ 표준 변수에 대한 정의, 용법, 도출방법이다. 정의에서는 보다 추상적인 수준에서만 개념을 정의하고²⁹⁾(단, ‘개인의 노동력 상태’ 변수이므로 범주를 이용하여 외연적으로 정의하고 있음), 용법에서 참조기간이나 유니버스를 설명한다. 이를 측정하기 위해 어떤 정보들을 수집하고, 이로부터 노동력 상태를 분류하는 조작적 정의에 해당하는 내용은 ‘도출방법’에서 설명하고 있다.

28) 표준으로 지정되지 않은 일반 통계분류를 포함함.

29) 단, ‘개인의 노동력 상태’ 변수이므로 범주를 이용하여 외연적으로 정의하고 있음.

<그림 24> 캐나다 'Labour force status of person' 변수 설명 일부

Labour force status of person

Status: This standard was approved as a departmental standard on June 21, 2010.

Definition

Labour force status refers to whether a person was employed, unemployed or not in the labour force during the reference period. The labour force consists of persons who contribute or are available to contribute to the production of goods and services falling within the System of National Accounts production boundary.

Person refers to an individual and is the unit of analysis for most social statistics programs.

Usage

'Labour force status' is used to describe the currently active population and to produce measures of employment and unemployment based on current economic activity. Because the intention is to determine if a person is employed or unemployed at a specific moment in time, a relatively short reference period such as a week is used.

Information on labour force status is typically collected for persons 15 years or over and excludes institutional residents. The population included in the study should be specified.

Derivation

Labour force status of a person is derived. It is derived from the following information:

1. Whether the person worked during the reference period;
2. Whether the person was absent from work during the reference period;
3. The person's reason for absence from work during the reference period;
4. Whether the person looked for work in the four weeks ending with the reference period;
5. Whether the person has a job to start within four weeks from the reference period;
6. The person's availability for work; and
7. The reason the person is not available for work.

Persons in the labour force are those who were either employed or unemployed during the reference period.

Employed persons include those who during the reference period:

1. Had paid work either as an employee or were self-employed;
2. Did unpaid family work, unpaid family work is defined as work contributing directly to a farm, business or professional practice owned and operated by a related member of the household;
3. Had a job but were absent from work due to illness or disability, personal or family responsibilities, vacation or labour dispute;
4. Were not at work because of their work schedule (e.g. shift work); or
5. Were not working because they were self-employed and had no work available.

Persons not at work because they are on layoff or are between casual jobs or because their job starts at a future date are not considered employed.

Unemployed persons include those who during the reference period:

1. Were without work but had looked for work in the past four weeks ending with the reference period and were available for work;
2. Were on temporary layoff due to business conditions and were available for work; or
3. Were without work, had a job to start within four weeks of the reference period and were available for work.

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

3. 그 밖의 사례

1) 마이크로데이터에서 변수별 Universe 제공

캐나다 노동력조사 공공용 마이크로데이터(Public Use Microdata File, PUMF)는 마이크로데이터 코드북에서 변수별로 Universe를 제공하며, 다른 변수명을 이용한 조건식이 아니라 ‘전체 응답자’, ‘15~29세 응답자’, ‘취업자’와 같이 개념어 형태로 작성한다.

<그림 25> 캐나다 변수별 Universe 제공 사례(2025.12월 노동력조사 PUMF)

1	Variable	Label	Universe	Note (English)
77	educ	Highest educational attainment	All respondents	
78		0 0 to 8 years		
79		1 Some high school		
80		2 High school graduate		
81		3 Some postsecondary		
82		4 Postsecondary certificate or diploma		
83		5 Bachelor's degree		
84		6 Above bachelor's degree		
85	mjh	Single or multiple jobholder	Currently employed only	
86		1 Single jobholder, including job changers		
87		2 Multiple jobholder		
88	blank	Not applicable		

<a> 코드북에서의 Universe 제공 방식

UNIVERSE	변수 수
전체	60
All respondents	13
Currently employed only	7
Currently employed or worked within the past 12 months	4
Currently employed, absent from work	2
Currently employed, employees	5
Currently employed, part-time usual work hours at their main or only job was below 30 per week	1
Currently unemployed	2
Currently unemployed, job searchers	7
Currently unemployed, temporary layoffs	1
Employees and self-employed - incorporated only	1
Employees at work only	5
Not currently employed	1
Not currently employed but worked at some time in the past	1
Not currently employed but worked in the past 12 months	1
Not currently employed but worked within the previous twelve months	3
Unemployed temporary layoffs and job searchers only	1
Temporary layoffs, job searchers and future starts, and those who wanted work but did not look because believes no suitable work available	1
Persons not in the labour force who say they wanted a job in the reference week	1
Respondents aged 15 to 29 years	1
Respondents aged 15 to 64 years	1
Parents only	1

 Universe 내용

출처: Statistics Canada(2026), Labour Force Survey (LFS) PUMF Codebook.

다만, <그림 26>의 국민가구조사 공공용 마이크로데이터 문서에서는 ‘종사상 지위 (과생변수)’에 대한 Universe를 ‘Persons in private households aged 15 years and over, who worked at some time since January 1, 2010’과 같이 참조기간을 포함하여 서술형으로 상세하고 있다. 캐나다 노동력조사 사례와 비교했을 때, 조건식을 이용한 서술은 아니지만 작성방식에는 소폭 차이가 있다. 두 사례의 작성시점이 각각 2025년과 2015년으로 동일하지는 않지만, Universe 작성 지침을 강하게 규정하는 것은 아니었을 것으로 짐작된다.

<그림 26> 캐나다 변수별 Universe 제공 사례(2011 국민가구조사 PUMF)

Variables Description				
Dataset contains 133 variable(s)				
File : NHS-99-004-XWE-E-2011-pumf-individuals_F1				
# COW: Labour: Class of worker (derived)				
Information	[Type= discrete] [Format=numeric] [Range= 1-9] [Missing=*/8/9]			
Statistics [NW/ W]	[Valid=518078 / 19047723.553] [Invalid=368934 / 13804599.447]			
Definition	Applies to persons aged 15 years and over, in private households, who worked at some point in time since January 1, 2010.			
Universe	Reported for: Persons in private households aged 15 years and over, who worked at some time since January 1, 2010			
Notes	Employee: This category includes persons who worked for others in the context of an employer -employee relationship. This includes persons who worked for wages, salaries, commissions, tips, piece -rates, or payments 'in kind' (payments in goods or services rather than money). Self-employed: This category includes persons whose job consisted mainly of operating a business, farm or professional practice, alone or in partnership. This includes: operating a farm, whether the land is rented or owned; working on a freelance or contract basis to do a job (e.g., architects, private duty nurses); operating a direct distributorship selling and delivering products such as cosmetics, newspapers, brushes and cleaning products; and fishing with own equipment or with equipment in which the person has a share. The business can be incorporated or unincorporated. Self-employed persons include those with and those without paid help. Unpaid family workers: They are persons who work without pay in a business, farm or professional practice owned and operated by another family member living in the same dwelling.			
Value	Label	Cases	Weighted	Percentage (Weighted)
1	Employee	463983	17029887.0	89.4%
2	Unpaid family worker	1270	46156.1	0.2%
3	Self-employed, without paid help, incorporated	8032	298906.5	1.6%
4	Self-employed, with paid help, incorporated	11193	416395.1	2.2%
5	Self-employed, without paid help, unincorporated	24966	937448.7	4.9%
6	Self-employed, with paid help, unincorporated	8634	318930.1	1.7%
8	Not available	2683	97208.8	
9	Not applicable	366251	13707390.7	
Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.				

* Information에서 자료 타입, 형식, 범위 등의 측정 정보를 제공하며, 범주별 Label에서 ‘Not available’, ‘Not applicable’과 같이 센티널 값을 제공하고 있음

출처: Statistics Canada(2015), National Household Survey (NHS) 2011 PUMF.

2) 마이크로데이터에서 Variable Group 제공

캐나다 2011년 국민가구조사 공공용 마이크로데이터에서는 총 133개 변수를 유사한 특성으로 그룹화하여 총 17개의 ‘Variable Groups’³⁰⁾로 제공한다. 모든 통계에 공통으로 적용되는 표준은 아니지만, 해당 통계 프로그램 내에서는 Concept System과 유사한 역할로 생각된다. <부록>과 같이 미국에서도 Frequently-used Variables라는 유사한 사례를 확인할 수 있다.

<그림 27> 캐나다 Variable Group 제공 사례(2011 국민가구조사 공공용 MD)

Variables Group(s)							
Dataset contains 17 group(s)							
Group Identifier							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	PPSORT	Unique record identifier	discrete	numeric-6.0	887012	0	-
Group Geography							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	CMA	Census metropolitan area of current residence (2011)	discrete	numeric-3.0	887012	0	-
2	PR	Province or territory of current residence (2011)	discrete	numeric-2.0	887012	0	-
Group Demography							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	AGEGRP	Age groups	discrete	numeric-2.0	882287	4725	-
2	MARSTH	Marital status (de facto)	discrete	numeric-1.0	887012	0	-
3	SEX	Sex	discrete	numeric-1.0	887012	0	-
(중략)							
Group Labour							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	COW	Labour: Class of worker (derived)	discrete	numeric-1.0	518078	368934	-
2	FPTWK	Labour: Full-time or part-time weeks worked in 2010	discrete	numeric-1.0	504028	382984	-
3	LFACT	Labour: Labour force status	discrete	numeric-2.0	734772	152240	-
4	LSTWRK	Labour: When last worked for pay or in self-employment	discrete	numeric-1.0	734772	152240	-
5	NAICS	Labour: Industry sectors (based on the NAICS 2007)	discrete	numeric-2.0	513559	373453	-
6	NOC11	Labour: Occupation major groups (based on the NOC 2011)	discrete	numeric-2.0	507205	379807	-
7	NOC5	Labour: Occupation broad categories (based on the NOC-S 2006)	discrete	numeric-2.0	507205	379807	-
8	WKSWRK	Labour: Weeks worked in 2010	discrete	numeric-1.0	520761	366251	-
9	WRKACT	Labour: Work activity in 2010	discrete	numeric-2.0	734772	152240	-

출처: Statistics Canada(2015), National Household Survey (NHS) 2011 PUMF.

30) 그룹명(괄호 안은 변수 개수): Identifier(1) / Geography(2) / Households(3) / Families and Family composition(3) / Place of Birth and Generation Status(9) / Aboriginal Peoples(3) / Ethnic Origin(4) / Language(18) / Education(7) / Mobility and Migration(5) / Journey to Work(7) / Labour(9) / Income(33) / Dwellings: Housing and shelter costs(11) / Religion(1) / Weighting(9)

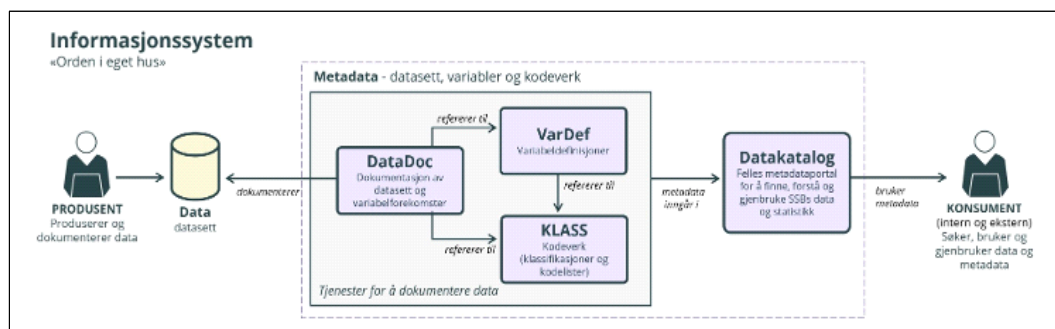
제2절 노르웨이 통계청

1. 노르웨이 Metadata 조회시스템

노르웨이 통계청은 데이터를 올바르게 사용하기 위해 수년 동안 ‘데이터 관련 메타데이터’를 문서화하는 시스템³¹⁾을 운영해 왔다. 다른 통계청의 경험상 하나의 메타데이터 시스템이 너무 크고 복잡했으며 Datadok이라는 기존 시스템을 계속 유지해야 했기 때문에, 단일 시스템을 구축하는 대신 기능별로 특화된 3개의 개별 시스템을 연동하는 전략을 선택했다.

Datadok에서는 변수, 변수 유형, 소수점 자릿수 등 데이터세트의 문서화에 중점을 두는데, 현재 메타데이터 포털³²⁾에서는 외부 이용자가 직접 검색할 수 있는 서비스를 제공하지는 않는다. Vardok에서는 정의 및 소유권 등 변수를 문서화하고, Klass에서는 통계분류 및 코드목록을 포함한 코드 세트를 관리한다. 시스템 간 통합을 통해 Vardok의 ‘산업’ 변수는 Klass의 ‘표준산업분류’와 연결되고, Vardok은 변수 정의 교환을 위해 Statbank 및 MetaDB와 연결되며, Datadok의 변수 또한 Vardok의 해당 변수 정의를 참조할 수 있다.

<그림 28> 노르웨이 통계청 메타데이터 관리 및 카탈로그 제공 흐름도

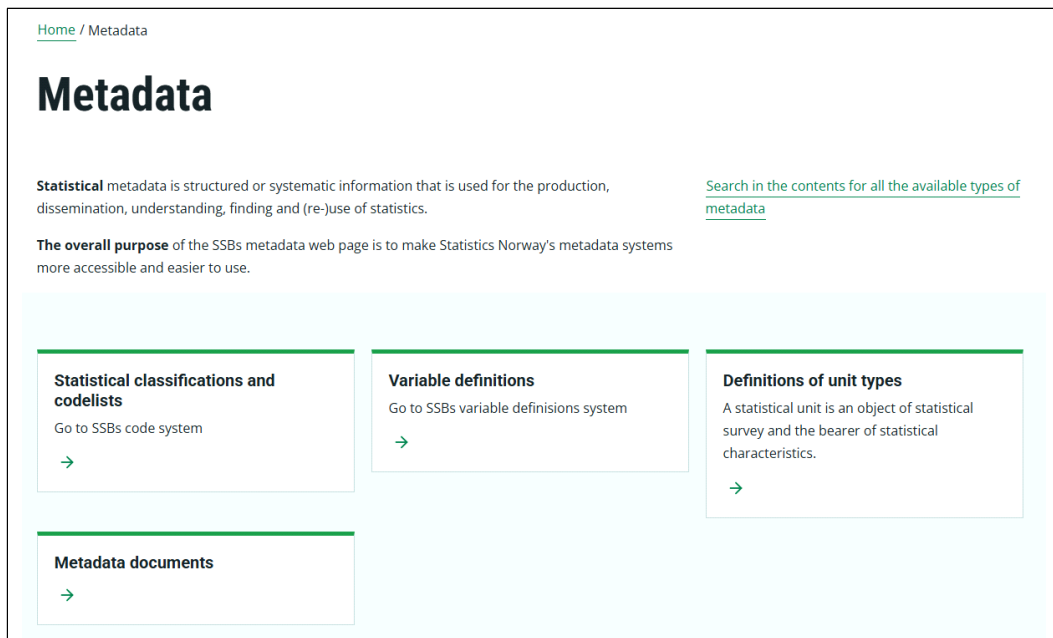


출처: Statistics Norway, Metadata – Dapla-manualen [웹사이트].

- 31) 메타데이터(데이터를 생성·검색·이해·재사용하는데 필요한 정보) 중에서 데이터세트 변수의 의미, 다른 데이터세트 변수와의 비교가능성 등 데이터세트에 포함된 내용을 이해하기 위해 밀접한 관련이 있는 ‘데이터 관련 메타데이터’에 초점을 둬(출처: <https://manual.dapla.ssb.no/statistikkere/metadata.html>)
- 32) 단, 현행 메타데이터 시스템에서 신규 데이터 포털(SSB Dataportabl)로 마이그레이션 진행 중이며, 신규 포털에서는 Datadok 조회 서비스도 포함되는 것으로 보이나 조회된 결과 문서가 노르웨이어로 작성되어 있어서 세부 내용을 이해하기 어려움

이렇게 연동된 메타데이터 시스템이 데이터 카탈로그로 연결되어 일반 이용자는 통합 메타데이터 포털을 통해 변수 및 분류 정보를 조회할 수 있다. 노르웨이 통계청은 <그림 29>과 같이 메타데이터란 ‘통계의 생산, 배포, 이해, 검색 및 (재)사용에 사용되는 구조화되거나 체계적인 정보’라고 소개하면서, 1) 통계분류 및 코드목록, 2) 변수 정의, 3) 단위유형 정의, 4) 메타데이터 문서의 총 4가지 메타데이터 조회 기능을 제공한다.

<그림 29> 노르웨이 통계청 메타데이터 조회화면(메인)



출처: Statistics Norway, Metadata [웹사이트].

1) 통계분류 및 코드목록(Statistical classifications and codelists: Klass)

노르웨이 통계청은 GSIM 노드세트의 하위 유형 중에서 통계분류와 코드목록을 엄격하게 분리한다. 통계분류(Statistical classification)는 공식적인 것으로 상호 배타적이고 완전하게 구성하며 ‘Classification of 개념’으로 명칭을 부여한다. 코드목록(Codelist)은 공식적이지 않으며 특정 통계에만 적용될 수 있는 것으로 정의하며 ‘Codelist of 개념’으로 명칭을 부여한다. 예를 들어, ‘Classification of age’는 121개 범주가 상호 배타적으로 구성되나, 코드목록은 중복 및 재코딩하여 해당 목적에 맞게 다수의 연령 코드목록이 구성되어 있다.

<표 19> 노르웨이 age 관련 통계분류 및 코드목록

구분	통계분류/코드목록명	주제	통계단위 ¹⁾	level	code	name
통계 분류	Classification of age (코드 121개)	인구	Person	1	000	0 years
				1	001	1 year
				1	∴	
				1	119	119 years
				1	120	120 years
코드 목록	Codelist for functional age and age groups (코드 166개)	인구	Person	1	000	0 years
				1	001	1 year
				1	∴	
				1	119	119 years
				1	120	120 years
				1	F000-002	0-2 years
				1	F000-003	0-3 years
				1	F000-005	0-5 years
				1	∴	
				1	F000-120	Total
	Codelist for IPLOS age groups (코드 7개)	건강	Person	1	F018-120	over 18 years
				1	∴	
				1	F00_17	0-17 years
				1	F00_120	total
				1	F18_49	18-49 years
				1	F50_66	50-66 years
	Codelist for enterprises by age group (코드 5개)	기업· 회계	Enterprise	1	F67_79	67-79 years
				1	F80_89	80-89 years
				1	F90_120	90 years and older
				1	01	0-2 years old
				1	02	3-5 years old
	Codelist for Age groups for teachers for use in KOSTRA (코드 5개)	교육	Education/ course/ issue	1	03	6-9 years old
				1	04	10-14 years old
				1	05	15 years old and older
				2	0	All age groups
				2	01	40 years and younger
				2	02	41 to 49 years
				2	03	50 to 59 years
				2	04	60 years and older

1) 노르웨이는 시스템 메뉴별로 'Statistical Unit' 과 'Unit Type' 표기가 혼용되어 있음

출처: Statistics Norway, Metadata [웹사이트].

따라서 통계분류 151개에 비해 코드목록은 596개로 훨씬 더 많다. 24개 주제 영역 중 농림어업·수렵, 지역 및 교통·관광을 제외하고는 주제 영역별 통계분류는 10개 미만이다. 다만 인구 주제에서는 통계분류가 6개로 코드목록 2개보다 많다.

<표 20> 노르웨이 주제별 통계분류 및 코드목록 제공 건수

Subject area	계	통계분류	코드목록
전체	747	151	596
노동 (Labour market and earnings)	67	1	66
금융 (Banking and financial markets)	3	-	3
인구 (Population)	8	6	2
건설·주택·부동산 (Construction, housing and property)	17	6	11
에너지·제조업 (Energy and manufacturing)	41	4	37
보건 (Health)	89	6	83
소득·지출 (Income and consumption)	8	-	8
이민·이민자 (Immigration and immigrants)	2	2	-
농림어업·수렵 (Agriculture, forestry, hunting and fishing)	47	24	23
문화·여가 (Culture and recreation)	23	-	23
국민계정·경기 (National accounts and business cycles)	3	1	2
자연·환경 (Nature and environment)	64	7	57
공공 부문 (Public sector)	53	1	52
물가·물가지수 (Prices and price indices)	13	6	7
지역 (Region)	64	36	28
사회·복지·범죄 (Social conditions, welfare and crime)	40	7	33
기술·혁신 (Technology and innovation)	10	2	8
교통·관광 (Transport and tourism)	36	15	21
교육 (Education)	82	5	77
대외 경제 (External economy)	13	8	5
선거 (Elections)	2	2	-
도소매·서비스 (Wholesale and retail trade and service activities)	2	2	-
기업·계정 (Establishments, enterprises and accounts)	26	6	20
통계청 내부용 (For internal use Statistics Norway)	34	4	30

출처: Statistics Norway, Metadata [웹사이트].

<표 21>에서는 주제별로 어떠한 통계분류 및 코드목록을 지정하고 있는지 살펴볼 수 있다. ‘통계청 내부용’ 주제가 별도로 구축되어 있다는 점이 인상적인데, SIMS 분류, GSBPM 분류, 예/아니오 코드목록, 단위유형 코드목록 등 각종 통계 생산을 위한 기초적인 통계분류 및 코드목록을 포함하고 있다. 특이한 점은 {1. Male, 2. Female}로 구성된 ‘Classification of sex’은 인구 주제, {0. Both genders, 1. Men, 2. Women}으로 구성된 ‘Codelist of gender’는 인구 주제가 아닌 통계청 내부용 주제에 포함되어 있다.

<표 21> 노르웨이 주제별 통계분류 및 코드목록 예시

주제	통계분류 및 코드목록	
노동	통계 분류 (1)	· Classification of occupations (직업 분류)
	코드 목록 (66)	· Codelist for Apprentices (견습생 코드목록) · Codelist for Applying for work outside the labour force - LFS (구직 신청 코드목록) · Codelist for educational status (교육상태 목록) ·
인구	통계 분류 (6)	· Classification of age (연령 분류) · Classification of citizenship (시민권 분류) · Classification of families (가족 분류) · Classification of households (가구 분류) · Classification of marital status (혼인상태 분류) · <u>Classification of sex</u> (성별 분류)
	코드 목록 (2)	· Codelist for functional age and age groups (기능적 연령 및 연령그룹 코드목록) · Codelist for citizenship (시민권(형용사형) 코드목록)
기업 · 계정	통계 분류 (6)	· Classification of codes for PRODCOM (광제조업 품목 분류) · Classification of enterprises by size group (기업규모별 분류) · Classification of Institutional sector (기관 부문 분류) · Classification of Legal form (법적 형태 분류) · Classification of Standard Industrial Classification (표준산업 분류) · Classification of Variant of SIC - Environmental accounts (표준산업분류 변형 - 환경계정)
	코드 목록 (20)	· Codelist for employed persons (sts) (취업자 코드목록) · Codelist for enterprises by age group (기업 설립 연수 그룹에 대한 코드목록) · Codelist for industrial classification in the business demography (16 categories) ·
통계청 내부용	통계 분류 (4)	· Classification of months (월별 분류) · Classification of organisational units (조직단위 분류) · Classification of Single Integrated Metadata Structure (SIMS) (SIMS 분류) · Classification of the Generic Statistical Business Process Model (GSBPM 분류)
	코드 목록 (30)	· Codelist for classification method (단위 분류방법에 대한 코드목록) · <u>Codelist for gender</u> (성별('Both genders' 포함) 코드목록) · Codelist for measurement types and measurement units (측정유형 및 측정단위 코드목록) · Codelist for No/Yes (아니오(0)/예(1) 코드목록) · Codelist for Yes/No (1,2) (예(1)/아니오(2) 코드목록) · Codelist for Seasonal Adjustment (계절조정 코드목록) · Codelist for sentinel values (센티널 값 코드목록) · Codelist for unit types (단위유형 코드목록) ·

출처: Statistics Norway, Metadata [웹사이트].

통계분류/코드목록 상세 조회화면에서는 통계단위(Statistical unit)를 확인할 수 있고, 전체 코드를 한 번에 조회할 수 있다. 다만, 각 노드에 대한 정의는 본 서비스에서는 제

공하지 않는다. 또한 GSIM의 Base 그룹 Identifiable Artefact에 해당하는 버전 정보, Business 그룹의 Owner, Legal base, Concept 그룹의 대응표 등 상세 사항을 구조화된 틀로 제공하고 있다.

<그림 30> 노르웨이 통계분류 및 코드목록 상세 조회화면

Statistical unit: **Person**

Classification of age

The main purpose of this standard is to be the base for different age groupings (variants).

Current version: (Valid from January 1960)

Age 1960

Codes
About the version
Changes
All versions
Correspondences
Variants

Filter by code or name
Filter
Reset

Close hierarchy
Download CSV

000 - 0 years
001 - 1 year
002 - 2 years

<a> 코드 탭

Codes
About the version
Changes
All versions
Correspondences
Variants

About Age 1960

Description: The main purpose of this standard is to be the base for different age groupings (variants).

Owner: Haug, Magnus, division 320 - Division for population statistic

Valid from: January 1960

Published in: Norwegian (bokmål), Norwegian (Nynorsk), English

Based on: Not relevant

Legal base: Not relevant

Publications: Not relevant

Levels:

No.	Name	Description	Structure	Number of units
1	Level 1			

Change log:

Date	Time	Comments
Ingen endringer		

 버전 정보 탭

* 탭별 제공정보

- [코드] 탭: code, name 표시 (csv파일에서는 parent code, level 정보까지 제공)
- [버전 정보] 탭: 설명, 소유자, 유효시작연월, 발간언어, 참조 법령, 레벨 정보, 변경 로그 등
- [변경 사항] 탭: 이전 버전과 현재 버전 간의 변경 사항을 표시
- [모든 버전] 탭: 문서화된 모든 버전에 대한 링크 및 각 버전의 유효기간 제공
- [대응 관계] 탭: 현재 버전과 다른 버전 간의 대응표 제공 (예. '16년과 '14년의 지자체 대응표)
- [변형] 탭: 이를 기준으로, 대체할 수 있는 그룹화(예. 환경계정 작성을 위해 표준산업분류 변형 분류 사용)

출처: Statistics Norway, Metadata [웹사이트].

2) 변수 정의 (Variable definitions)

‘변수’란 관찰대상 단위의 특성이며, 수치적 값 집합 혹은 범주형 값 집합을 가질 수 있다고 설명한다. 메타데이터 시스템에서 제공하는 변수 목록은 완전하지 않으며, 향후 다양한 변수에 대한 문서화 작업과 병행하며 업데이트 예정이다. 현재 영문 버전 메타데이터 시스템에서는 총 1,193개 변수를 조회할 수 있다.

<표 22> 주제 및 통계단위별 변수 개수(단위: 개)

통계단위	계	주제													
		인구	가계	노동	사회·보건	교육	문화	산업·활동	국민·계정·무역	금융	재정	물가·경제	환경	기타	
계	1,193	144	149	72	74	20	25	422	106	35	59	31	19	37	
Enterprise (기업체)	405			13	12		20	260	68	26	3	2	1		
Person (개인)	355	62	143	46	61	16	4	8	1	2				12	
Establishment (사업체)	142			4		2	1	59	27	3	29	13	4		
Dwelling (거처)	45	39						1				5			
Household (가구)	37	20	6			1			6		1	3			
Vehicle (차량)	32							32							
General government (정부)	29								3		26				
Municipality (지자체)	19	1		1				7	1				3	6	
Family (가족)	14	14													
Building (건물)	13	7						5				1			
Region (지역)	12	1											1	10	
Agricultural and forestry property (농림지)	11							11							
Ship (선박)	10							10							
Election (선거)	9													9	
Fishing vessel (어선)	9							9							
Animal (동물)	8							8							
Repr. goods and services (재화·서비스)	7											7			
Employment (일자리)	7			7											
기타*	29			1	1	1		12		4			10		

* 기타 통계단위: Waste treatment plant (폐기물 처리시설) / Wastewater treatment plant (폐수 처리시설) / Shareholder (주주) / Forest area (임야) / Accident (사고) / Air craft (항공기) / Ground Parcel (필지) / Case (소송) / Course (강의) / Registered transfer (등기 이전) / Internet subscription (인터넷 가입) / - (비어있음)

출처: Statistics Norway, Metadata [웹사이트].

개념에 대한 상세 조회화면에서는 변수를 ‘Concept variable’로 지칭한다. 개념변수에 대해 통계 프로그램 담당자가 작성해야 하는 메타데이터 항목은 필수 항목과 선택 항목을 통틀어서 총 34개(<부록 9>)이다. 메타데이터 포털은 이 중에서 이름, 정의, 유효기간, 관리부서, 통계단위, 주제 분야는 고정적으로 제공하고, 그 외에 연계 변수(Linked to Variable Definition), 연계 통계표 링크(Linkded to StatBank tables), 통계분류, 산출식 등은 해당 사항이 있는 경우만 표출된다. 한편, 본 서비스에서 연계 통계표 링크가 제공되는 것처럼, 역으로 통계표 조회 시스템(StatBank)에서는 통상의 통계기관이 제공하는 ‘주제’ 필터링 및 키워드 검색뿐 아니라 ‘변수’ 필터링 서비스를 제공한다(<부록 10> 참고). 즉, 변수에 대한 설명자료를 재사용성 강화 및 정보(설명자료) 제공에만 그치지 않고, 통계 조회 서비스까지 보다 적극적인 역할로 활용할 수 있다.

<그림 31> 노르웨이 Conceptual Variable 조회 화면

Concept variable	
Employed persons	
Norwegian (bokmål)	
Name	Employed persons
Definition	Persons performing at least one hour of income-producing work during the week or day referred to, as well as persons who have this sort of work, but who were temporarily absent due to sickness, vacation, paid leave etc. Persons in the civil service and conscripts are considered employed persons. Involuntary laid off persons, with a continuous duration of up to three months, are defined as employed and temporarily absent.
Valid from	1980-01-01
Valid to	
Owner	260 - Division for labour market statistics
Statistical unit	Person
Subject	06 - Labour market, wages
Linked to Variable Definition	Employees
Linked to StatBank tables	Hide links
	11188: Employed persons total, part time employees and underemployed persons, by sex and industry (SIC2007) (1 000 persons) 11154: Employed persons, by sex and major industry division (SIC2007) (1 000 persons) 11153: Employed persons, by sex and major industry division (SIC2007) (1 000 persons) 11144: Location of R&D cooperation partner in the business enterprise sector, by industry (SIC2007) , employment group 11142: R&D cooperation in the business enterprise sector, by industry (SIC2007) and employment group 11140: Product and process related R&D in the business enterprise sector. Current cost, by industry (SIC2007) employment group and product/process R&D (NOK million) 11138: R&D type in the business enterprise sector. Current cost, by industry (SIC2007) and employment group (NOK million) 11136: Thematic area of R&D in the business enterprise sector. Current cost, by industry (SIC2007) and employment group (NOK million) 11134: Technology area of R&D in the business enterprise sector. Current cost, by industry (SIC2007) and employment group (NOK million) 11131: Employed persons, by sex, major industry division (SIC2007) and status (1 000 persons) 10815: Employed persons, man-years and hours worked, by provider 10585: Compensation and employees, by industry, education and sex 10439: Employed persons, man-years and hours worked, by provider (closed series)

출처: Statistics Norway, Metadata [웹사이트].

<표 23>에서 제시된 개념변수와 그 정의를 살펴보면, ‘일하였음(working)’이라는 최상위 개념이 서로 다른 통계단위를 측정하면서 각기 다른 개념변수로 설계되는 구조를 확인할 수 있다. 다만, 노르웨이 통계청에서 제공하는 개념변수 정보에는 개념이 아닌 개념변수에 대한 정의만 제공하며, 이는 개념 측정이 가능하게끔 구체적인 수준으로 정의되고 있어서 이들 세 변수의 가장 추상적인 수준의 개념 정보는 확인하기 어렵다.

<표 23> 노르웨이 통계청 개념변수 정의 사례

주제	통계단위	개념변수	정의
노동	Person	Employed persons	주간(일간) 1시간 이상 소득 창출 활동을 한 사람 - 일시적 결근 및 3개월 미만 일시 해고자 포함 - 공무원과 병역자 포함
	Employment	Job	일자리(job)과 고용(employment)은 소득으로 보상받는 일(work)을 의미하는 동의어로 사용됨 - 한 사람이 여러 사업체에서 여러 일자리/고용을 가질 수 있으나, 같은 사업체에서는 1개의 일자리로 집계
산업 활동	Establishment	Employed persons - structural business statistics	해당 사업장에 근무하는 모든 소유주와 직원의 합계 - 복수 취업시, 여러 사업장에 고용된 것으로 집계 - 고용 수치는 연평균값을 나타냄 - 파견 근로자 제외(단, 인력공급업에서는 포함)

출처: Statistics Norway, Metadata [웹사이트].

3) 단위 유형 정의 (Definitions of unit types)

단위 유형(Unit type)은 통계가 작성되는 대상들의 그룹을 나타내는 것으로, 노르웨이 통계청의 통계 생산·발간물에 사용되는 단위 유형을 지정하고 있다. 모든 자연인은 ‘개인’ 단위 유형에 속한다고 설명하는 것처럼, 캐나다 통계청과 달리 GSIM의 Universe가 아닌 GSIM의 Unit type에 국한하고 있다. 허용가능한 단위 유형값은 <그림 32>과 같이 총 31개이다. 다만, 앞서 변수 상세 조회화면에서와 같이 시스템 메뉴에 따라 ‘Statistical unit’ 표기가 혼용되어 있다.

<그림 32> 노르웨이 통계청 단위 유형

• Person 개인 • Family 가족 • Household 가구 • Building 건물 • Utility unit 기능 단위 • Address 주소 • County(administrative) 카운티(행정적) • County(geographic) 카운티(지리적) • Municipality(administrative) 지자체(행정적) • Municipality(geographic) 지자체(지리적)	• Legal unit 법인 단위 • Enterprise 기업 • Establishment 사업체 • Governmental enterprise 정부기업 • Goods/services 재화/서비스 • Legal right 법적 권리 • Securities 증권 • Transaction 거래 • Property 재산 • Legal right 법적 권리	• Employment 고용 • Kind-of-activity unit 활동유형 단위 • Cattle 소 • Vehicle 차량 • Ship 선박 • Course 강좌 • Election 선거 • Port call 기항 • Offence 범죄 • Traffic accident 교통사고 • Workplace accident 직무상 사고
---	---	--

* ‘Codelist for unit types’에는 위의 31개 외에 ‘Region’, ‘Aggregated’, ‘Not a unit type’ 범주 포함

출처: Statistics Norway, Metadata [웹사이트].

4) 메타데이터 문서 (Metadata documents)

마지막으로 제공하는 서비스 메뉴에서는 <표 24>와 같이 메타데이터 문서 링크를 제시한다. ‘정의’ 관련해서는 Neuchâtel 용어 모델 및 GSIM을 참조하였음을 명시하고 있다.

<표 24> 노르웨이 통계청 메타데이터 문서

구분	내용
General	· DOC 2006/05. DOC2006/05 Statistical Metadata in Statistics Norway. Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS), Geneva, 3 - 5 April 2006 · DOC 2005/08. Metadata strategy in Statistics Norway. Eurostat Metadata Working Group Luxembourg, 6 - 7 June 2005 · DOC 2004/14. Development of a variables documentation system in Statistics Norway. International Statistical Conference 'Investment in the future', Prague, Czech Republic, 6 - 7 September 2004
Definitions	· Neuchâtel Terminology Model - Part II: Variables and related concepts objects types and their attributes (Version 1) · General Statistical Information Model (GSIM): Statistical Classifications Model
External links	· RAMON- Eurostat's Metadata Server · SDMX - Content-Oriented Guidelines (COG) · OECD - Glossary of Statistical Terms · ISI Multilingual Glossary of Statistical terms · The Common Metadata Framework · Data Documentation Initiative (DDI) · UNECE Case Studies of Metadata use with GSBPM and GSIM

출처: Statistics Norway, Metadata [웹사이트].

제3절 해외 사례 비교

1. 개념 관리체계 비교

캐나다 통계청과 노르웨이 통계청은 GSIM을 적용하여 국가통계 생산에 활용되는 개념, 통계단위, 변수, 분류 및 코드목록 등을 체계적으로 관리하고 있다는 공통점이 있으나 이를 적용하고 구현하는 방식에는 차이가 있다. 캐나다는 기관 차원의 표준 정책과 중앙 메타데이터 저장소를 기반으로 표준을 선별하고 이를 재사용하는 데 중점을 두는 반면, 노르웨이는 변수 정의, 통계분류, 코드목록 등을 기능별 시스템에서 관리하고 이들 간의 관계를 연계하는 방식이 두드러진다.

<표 25> 캐나다·노르웨이 통계 개념 정보 관리체계 비교

구분	캐나다 통계청	노르웨이 통계청
관계 체계	기관 차원의 표준 정책 및 중앙 메타데이터 저장소 중심	기능별 메타데이터 시스템을 연계하는 구조
시스템	IMDB(Integrated Metadata Base)	Datadok, Vardok, Klass 등
거버넌스	표준부서, 방법·표준 위원회, 프로그램 담당자의 역할 분담 명확	변수 정보는 통계 프로그램 담당자가 입력
주요 특징	기관 표준화와 재사용성	기능별 전문화와 메타데이터 간 연계

주요 정보 클래스별로도 공통점과 차이점이 상존한다.

Subject Field의 경우, 캐나다는 주제 영역 총 31개 중에서 표준이 포함된 주제는 25개로 일부 주제에서는 지정된 표준이 없음을 알 수 있다. 노르웨이는 통계 개념이 사용되는 지식 분야로서의 주제 외에 ‘통계청 내부용’ 주제에서 통계분류 및 코드목록을 관리하고 있다는 점에서 통계 프로세스 단계에서 필요한 분류상의 표준화까지 고려하고 있다는 점이 인상적이다.

Unit Type의 경우, 캐나다는 ‘개인’, ‘여성’, ‘15세이상 인구’ 등과 같이 GSIM의 Unit Type뿐 아니라 Universe 개념까지 포괄하여 Statistical Unit으로 관리한다. 표준 Statistical Unit이 정의된 주제는 인구, 가족·가구, 다양성, 노동, 기업, 주거, 건설의 7개로 한정적이지만, 그밖에 표준으로 지정되지 않은 Statistical Unit도 통계 프로그램 수준에서 사용

된다.³³⁾ 반면, 노르웨이는 GSIM의 Unit Type 개념을 충실하게 적용한다. 55페이지의 <표 22>를 보면 다양한 통계단위들이 여러 가지 주제에 해당하는 변수의 측정 대상으로 사용되고 있음을 알 수 있다. 다만, 기업체, 개인, 사업체, 거처, 가구, 차량, 정부 등이 일부 통계단위들의 활용도가 높은 반면, 일부 통계단위는 이를 대상으로 측정하는 변수가 5개 미만으로 과소하다.

<표 26> 캐나다·노르웨이 주요 정보 클래스 관리 비교

구분		캐나다 통계청	노르웨이 통계청
Subject Field		Subject 25개 (표준이 제공되는 주제에 한함)	Subject 24개 (‘통계청 내부용’ 주제 포함)
Unit Type		Statistical Unit 33개 (맥락이 부여된 Universe까지 포괄)	Unit Type 31개 (일부 메뉴에서는 Statistical Unit으로 지칭)
Variable	유형	변수 162개 (‘개념 of 통계단위’로 명명)	Conceptual Variable 1,193개 (통계표 조회 시스템과 연계)
	정보	정의, 용법, 도출방법/측정산식, 관련 분류, 표준 의무수준 등 (통계 프로그램별 변수는 정의, 관련 분류만 제공)	(고정적) 정의, 통계단위, 유효기간, 관리부서 등 (해당시) 관련 변수·통계표, 통계분류, 산출식 등
	표준	표준 변수와 통계 프로그램별 변수 관리 별도	표준 여부 미구분
	범위	Classification의 범주는 불포함하기도 함	Statistical Classification 및 Codelist의 범주 여부와 무관
Node Set	유형	Classification 226개 (통계분류와 코드목록 미구분)	Statistical Classification 151개, Codelist 596개 (둘을 엄격 구분)
	조회	범주 정의와 사용법 등을 제공하나, 전체 구조 조회에는 상대적으로 제약	전체 코드 구조, 버전, 변경이력, 대응표, 변형 등을 구조적으로 제공하나, 범주 정의는 미제공

Variable은 캐나다의 표준 변수와 노르웨이의 변수 모두 GSIM의 Conceptual Variable에 상응한다는 공통점이 있다. 변수 등록 범위는 차이가 있는데, 캐나다는 162개 변수가 표준으로 선별되어 있다. 또한, Classification의 범주는 표준 변수에 등재되지 않은 경우가 많다. 예컨대, ‘Classification of Labour Force Status’의 한 범주인 ‘Unemployed’와

33) 캐나다 환경 주제의 통계 프로그램 중에서 ‘Biennial Industrial Water Survey’의 주요 측정 대상인 ‘Location’은 ‘Business performance and ownership’ 주제의 표준 통계단위인 반면, ‘Biennial Agricultural Water Survey’의 주요 측정 대상인 ‘Farm’은 표준 통계단위로 지정되어 있지 않음.

관련된 변수는 표준으로 등록되어 있지 않다. 반면, 노르웨이는 범주 여부와 무관하게 통계 프로그램에서 작성되는 전체 변수들을 조회할 수 있다.

변수에 대한 제공 정보에 있어서는, 변수 이름, 변수 정의, 통계단위, 주제를 제공한다는 점에서는 공통점이 있으나, 캐나다는 용법(usage), 도출방법(derivation), 국제표준 등의 정보를 상세하게 제공하고, 노르웨이는 GSIM의 Identifiable Artefact 속성에 해당하는 버전 정보를 제공하며 연계된 통계표 링크를 통해 데이터와의 연계성을 높였다는 데에서 차이가 있다.

한편, GSIM에서 Conceptual Variable이 단 하나의 Conceptual Domain을 참조한다고 설명하고 있으나, 두 국가 사례 모두 변수에 대한 Conceptual Domain 정보는 제공하지 않는다는 점이 주목할 만하다. 특히, 캐나다는 44페이지의 <표 18>과 같이 하나의 표준 변수에 여러 분류를 연계하고 있어, GSIM에서 제시하는 정보 클래스 간 관계를 실제 통계작성 환경에 맞게 변형하여 적용한 사례로 볼 수 있다.

두 국가의 차이는 통계분류와 코드목록의 관리방식에서도 명확하게 나타난다. 캐나다는 Classification이라는 명칭 아래 GSIM의 Statistical Classification과 Code List에 해당하는 정보를 함께 제공하는 반면, 노르웨이는 GSIM 설계대로 이 둘을 엄격하게 구분한다. 노르웨이는 GSIM의 설계에 따라 통계분류는 공식적이고 상호 배타적이며 완전한 구조를 갖고, 코드목록은 특정 통계 또는 분석 목적에 따라 구성할 수 있는 것으로 관리한다. <그림 33>과 같이 변형분류를 나타내는 방식에서도 노르웨이는 이를 명시할 수 있는 ‘Variants’ 항목을 기본적으로 제공하며, 캐나다는 분류명에 명시를 하거나 분류 설명에서 ‘Usage’로서 부가 설명을 제공한다. 종합해 보면, 국내 적용 시 노르웨이의 구조적 조회 방식과 캐나다의 범주 정의 제공 방식을 상호 보완적으로 고려할 필요가 있다. 또한 노르웨이에서 제공하는 버전, 변경이력 및 대응표는 코드목록 및 통계분류를 관리하는데 있어 필수적으로 생각되며, 국내 적용시 코드목록 버전 간 대응표뿐 아니라, 표준 코드목록과 통계 프로그램별 코드목록의 대응표 또한 개념 조화를 이루기 위해 중요한 요건으로 생각된다.

<그림 33> 캐나다·노르웨이 변형분류 정보 제공 비교

Name	Classification of sex	Codelist for gender		
Statistical unit	Person	Person		
Subject	Population	Population		
Classification	1	Male	0	Both genders
	2	Female	1	Male
			2	Female
Description	Statistics Norway's official sex codes.	Codelist for genders, 'both genders' included		
Variants	Gender, included 'Both genders' - variant of sex	-		

<a> 노르웨이 'Classification of sex' 및 'Codelist for gender' 설명 일부

1 - Male

This category includes persons whose reported sex at birth is male.

2 - Female

This category includes persons whose reported sex at birth is female.

※ **Usage:** Information on sex at birth may be based on self-reported data or reported by proxy depending on the statistical program. Therefore, the expression 'reported sex at birth' is used in these definitions when referring to self-reported data and answers reported by proxy.

<b1> 캐나다 'Classification of sex at birth' 범주 정의

1 - Male

This category includes persons whose reported sex at birth is male.

2 - Female

This category includes persons whose reported sex at birth is female.

3 - Intersex

This category includes persons whose reported sex at birth is intersex.

※ **Usage:** Information on sex at birth may be based on self-reported data or reported by proxy depending on the statistical program. Therefore, the expression 'reported sex at birth' is used in these definitions when referring to self-reported data and answers reported by proxy.

According to the Government of Canada's LGBTQ2 Secretariat's glossary, the term 'intersex' is 'an umbrella term to capture various types of biological sex differentiation. Intersex people have variations in their sex characteristics, such as sex chromosomes, internal reproductive organs, genitalia, and/or secondary sex characteristics (e.g. muscle mass, breasts) that fall outside of what is typically categorized as male or female.' For reasons related to the small size of this population and the challenges in identifying intersex people, Statistics Canada does not currently collect specific information on intersex Canadians.

<b2> 캐나다 'Classification of sex at birth, variant' 범주 정의

* 노르웨이는 변형관계를 명시하도록 기본 틀로 관리하고 있으나, 'Both genders'가 추가된 사실과 변형관계만 제공할 뿐 더 상세한 내용까지 제공하지 않음. 캐나다는 '인구규모가 작고 식별이 어려워 실제로는 Intersex(3)의 특정 정보를 수집하지 않는다'는 것을 용법(Usage)에 명시하고 있지만 모든 변형분류에서 이처럼 작성되도록 필수 항목으로 관리하고 있지 않음.

2. 개념 및 변수 정의의 추상화 수준 비교

앞에서 개념 관리체계를 구조적으로 비교했다면, 이번에는 개념 및 변수의 정의를 실제로 어떻게 작성하고 있는지 내용 면에서 비교한다.

GSIM을 개념적으로 적용하면 Conceptual Variable은 Unit Type과의 결합을 통해 구체화되며, 그 외에는 Concept 수준의 추상성을 유지할 것으로 생각했었다. 하지만 <표 27>를 보면 캐나다와 노르웨이 모두 Conceptual Variable에 상응한다고 판단되는 변수에서 상당히 조작적 정의에 가까운 수준의 정보를 다루고 있음을 확인할 수 있다. 즉, ‘Employed (person)’을 정의하기 위해, 수입목적 근로, 무급가족 종사, 질병·휴가·노동쟁의 등에 따른 일시적 결근 등의 개념을 구체적으로 명시하고 있다. 다만, 캐나다가 변수 정보 중 Definition과 Derivation을 분리하여 세부 기준을 기술하는 반면, 노르웨이는 ‘주 1시간 이상 소득창출활동’, ‘3개월 이내 일시해고’와 같은 수치 기준을 정의문 안에 직접 포함한다는 점에서 차이가 있다. 즉, GSIM은 Variable Cascade 개념을 적용하여 재사용성을 강화하고자 하지만, 이번 사례에 따르면 양국 모두 변수의 판정기준과 다른 변수와의 식별가능성을 확보하는데 무게를 두어, 조작적 정의를 변수 수준에서 실무적으로 구현하고 있는 것으로 나타난다.

<표 27> 캐나다·노르웨이 '취업자' 정의 비교

캐나다		노르웨이
'Employed' 범주 정의 (Classification of Labour Force Status)	'Employed person' Derivation 일부 (Labour force status of person)	'Employed persons' 변수 정의
This category includes persons who, during the reference period: (a) Did any work at all at a job or business, that is, paid work in the context of an employer-employee relationship, or self-employment. It also includes persons who did unpaid family work, which is defined as unpaid work contributing directly to the operation of a farm, business or professional practice owned and operated by a related member of the same household; or (b) Had a job but were not at work due to factors such as their own illness or disability, personal or family responsibilities, vacation or a labour dispute. This category excludes persons not at work because they were on layoff or between casual jobs, and those who did not then have a job (even if they had a job to start at a future date).	Employed persons include those who during the reference period: 1. Had paid work either as an employee or were self-employed; 2. Did unpaid family work, unpaid family work is defined as work contributing directly to a farm, business or professional practice owned and operated by a related member of the household; 3. Had a job but were absent from work due to illness or disability, personal or family responsibilities, vacation or labour dispute; 4. Were not at work because of their work schedule (e.g. shift work); or 5. Were not working because they were self-employed and had no work available. Persons not at work because they are on layoff or are between casual jobs or because their job starts at a future date are not considered employed.	Persons performing at least one hour of income-producing work during the week or day referred to, as well as persons who have this sort of work, but who were temporarily absent due to sickness, vacation, paid leave etc. Persons in the civil service and conscripts are considered employed persons. Involuntary laid off persons, with a continuous duration of up to three months, are defined as employed and temporarily absent.

출처: Statistics Canada, Classifications, variables and statistical units; Statistics Norway, Metadata [웹사이트].

혼인상태(Marital status)는 캐나다와 노르웨이 모두 개인을 통계단위로 하는 인구학적 기본 변수이지만, 두 나라가 '무엇을 하나의 변수로 볼 것인가'를 정하는 기준 자체가 다르다. 캐나다는 사실혼 여부라는 하나의 개념적 축을 기준으로 변수를 아예 둘로 분리하고 각각을 표준으로 승인하고, 각각의 범주명에 '(not living common law)' 또는 '(including living common law)'를 병기하여 두 변수 간 대응 관계를 명시적으로 드러낸다. 반면 노르웨이는 단일 변수('Marital status')로 법률혼 여부만을 다룬다. 범주 안에 2009년 동성혼 법제화 이전에 성립했던 관계 유형(등록 동반자, 별거 동반자 등) 4개 범주가 그대로 남아 있는데, 개념을 새롭게 분리·결합한 결과라기 보다는 제도적 상황을 반영된 결과로 볼 수 있다. 즉, 사실혼이라는 개념 자체는 고려되지 않았다.

즉, 동일한 상위 개념(혼인상태)이라 하더라도 그 하위 개념(사실혼/법률혼)을 변수 수준에서 분리하여 포착할 것인지는 고정된 기준에 따라 결정되는 것이 아니라 통계 이용 목적과 제도적 맥락에 따라 달라질 수 있다. 따라서 국가통계 개념정보를 구축할

때, 개념과 변수의 관계를 확일적으로 설정하기보다는 상·하위 개념 간 관계와 변수의 분리·통합 필요성 및 제도적 맥락을 함께 고려할 필요가 있다.

<표 28> 캐나다·노르웨이 '혼인상태' 변수 비교

캐나다		노르웨이
Marital status of person	Legal marital status of person	Marital status
Marital status refers to whether or not a person is living in a common-law union as well as the legal marital status of those who are not living in a common-law union. All persons aged less than 15 are considered as never married and not living common law. Person refers to an individual and is the unit of analysis for most social statistics programs.	Legal Marital status refers to the marital status of the person under the law, not taking into account common-law unions. All persons aged less than 15 are considered as never married. Civil unions are included. Same-sex marriage has been legal in all provinces and territories of Canada since 2005. Person refers to an individual and is the unit of analysis for most social statistics programs.	This variable shows the person's marital status according to marital legislation.

출처: Statistics Canada, Classifications, variables and statistical units; Statistics Norway, Metadata [웹사이트].

<표 29> 캐나다·노르웨이 '혼인상태' 분류 비교

캐나다		노르웨이
Classification of marital status ¹⁾	Classification of legal marital status ²⁾	Classification of marital status ³⁾
1 Married	1 Married	1 Unmarried
2 Living common law	2 Never married (including living common law)	2 Married
3 Never married (not living common law)	3 Separated (including living common law)	3 Widow/widower
4 Separated (not living common law)	4 Divorced (including living common law)	4 Divorced
5 Divorced (not living common law)	5 Widowed (including living common law)	5 Separated
6 Widowed (not living common law)		6 Registered partner
		7 Separated partner
		8 Divorced partner
		9 Surviving partner

- 1) Usage: 분류 사용시, 사실혼 관계(living common law)는 제외된다는 점을 각주로 명시할 수 있습니다.
 2) Usage: 분류 사용시, 사실혼 관계(living common law)의 포함된다는 점을 각주로 명시할 수 있습니다.
 3) Description: 이러한 혼인상태 분류는 1993년부터 인구통계에 사용되어 왔습니다. 2009년 1월 1일, 동성 결혼이 허용되도록 결혼법(Marriage Act)이 개정되면서 동거관계 등록제도가 폐지됨에 따라, 이미 동거 관계로 등록된 사람들은 동거 관계를 유지하거나 동거 관계를 혼인 관계로 전환할 수 있습니다. 따라서 모든 혼인 상태 코드는 여전히 유효하며, 동거 관계도 포함됩니다.

출처: Statistics Canada, Classifications, variables and statistical units; Statistics Norway, Metadata [웹사이트].

제 4 장

국내 활용 방안 검토

제3장에서 확인한 바와 같이, 캐나다와 노르웨이 모두 GSIM을 활용하고 있으나 이를 각국의 통계 작성 환경에 맞게 선택·변형하여 운영하고 있다. 특히, 변수에 대해 조작적 정의에 가까운 정보를 제공한다거나, 제도적 맥락에 따라 서로 다른 변수와 범주로 구성하는 사례도 확인되었다. 이는 GSIM을 국내에 적용할 때 정보 클래스 구조를 그대로 도입하기보다 각 정보 클래스에 대한 관리 실익과 국내 통계환경을 함께 고려하여 구체화할 필요가 있음을 시사한다.

본 장에서는 이러한 시사점을 바탕으로 국가통계 개념 정보 관리체계 구축방안을 제시한다. 먼저 개념 정보를 표준으로 관리하기 위한 거버넌스 방향을 제시하고, GSIM의 주요 정보 클래스별 관리 방안과 그 밖에 실무적으로 결정할 사안이나 추가 검토가 필요한 사항에 대해서 다룬다.

제1절 개념 정보 관리 거버넌스

국내에서는 개념을 정의하기 위한 틀이나, 이를 관리·조정하는 절차가 부재하며, 정의가 개정되었을 때 이를 공식적으로 문서화하고 공지하는 체계도 마련되어 있지 않다.

캐나다는 기관 차원의 표준 정책을 통해 개념 정보의 표준화와 재사용성을 확보하고 있다. Policy on standards(2004)는 약 2페이지 분량의 간결한 지침이지만, 모집단·통계단위·개념·변수·분류 전반에 걸쳐 일관된 명칭과 정의를 제공한다는 목표를 명확히 제시하고, 표준의 검토·문서화·승인·모니터링이라는 관리 절차를 규정하고 있다는 점에서 시사하는 바가 크다.

이를 벤치마킹하여, 국내에서도 개념 정보에 대한 표준 관리 지침을 마련할 것을 제

안한다. 지침을 통해, 표준으로 관리할 정보의 범위(개념·변수·단위유형·분류 등)와 각 정보의 표준 명칭·정의를 규정하고, 캐나다의 부처 표준·권고 표준·프로그램별 표준과 같이 의무 수준을 구분하여 운영함으로써 표준화의 강제력과 유연성을 동시에 확보하는 방안을 고려할 수 있다. 또한 표준의 제정·개정·승인·폐기에 관한 관리 업무와 절차, 그리고 이를 수행할 관리 주체를 명확히 할 필요가 있다. 현행 통계법에서는 ‘통계 표준분류 등 통계의 작성·보급 및 이용의 기준에 관한 사항’을 국가통계위원회 심의·의결 사항으로 규정하고 있어서, 국가통계위원회 및 분과위원회에서 캐나다의 정책위원회, 방법·표준 위원회와 같은 역할을 수행할 수 있을 것으로 기대된다. 개념에 대한 표준을 문서화하고 모니터링하면서 통계 프로그램 담당자와의 소통 등 상시적인 관리를 위해서는 표준 관리 전담 부서 또한 필요하다.

제2절 정보 클래스별 관리 방안

1. Unit Type

현재 통계정책관리시스템에서는 통계조사별로 ‘조사(작성)단위’를 작성하고 있으나, 변수 수준에서는 이를 별도로 관리하고 있지 않다. 또한 현행 조사(작성)단위는 ‘개인’, ‘가구’, ‘사업체’, ‘기업체’, ‘기타’의 5개 범주 중 하나를 선택하도록 하고 있는데, 이 중에서 ‘기타’로 분류된 사례가 전체 1,379개 승인 통계 중 497개로 약 36%에 달한다. 특히, 환경, 국민계정, 보건, 농림, 정부·재정, 교통·물류에서 ‘기타’ 비중이 높게 나타나고 있어, 현행 5개의 범주만으로는 실제 통계에서 활용되는 다양한 조사(작성)단위를 충분히 포괄하기에 부족한 것으로 판단된다.

캐나다와 노르웨이는 이러한 통계단위 정보를 변수 수준에서 관리하고 있다. 다만, 캐나다와 같이 Universe까지 포괄하여 표준화할 경우, 통계 작성 맥락에 따라 ‘취업자’, ‘여성’ 등 다양한 개념들이 추가될 수 있어 이를 공통 표준값으로 지정·운영하기에는 한계가 있을 것으로 예상된다. 따라서 국내 적용을 위해서는 노르웨이와 같이 GSIM의 Unit Type 개념에 준하여 통계단위를 표준화하는 방안이 더욱 적합할 것으로 판단된다.

국내 다양한 통계분야 전반을 포괄할 수 있는 Unit Type에 대한 표준값을 정립하고, 이를 변수 수준에서 적용·관리하기 위해, 표준 관리부서는 <표 30>과 같이 ‘기타’ 비율이 높은 통계 분야를 중심으로 변수별 조사(작성)단위의 유형을 검토하여 국가통계 전반에서 활용할 수 있는 Unit Type 표준값 집합을 마련할 필요가 있다.

아울러, Unit Type을 지칭하는 통계용어 표준화가 필요하다. 국내 조사 수준에서는 ‘조사(작성)단위’(통계정책관리시스템), ‘조사단위’(통계설명자료서비스) 등이 서로 다른 용어가 사용되고 있고, 캐나다와 노르웨이는 이를 Statistical Unit으로 명명하고 있다. ‘명’, ‘원’과 같은 측정단위(Measurement Unit)와 혼동되지 않으면서, 조사 수준과 변수 수준을 아우를 수 있는 국내 표준 통계용어 마련이 필요하다.

<표 30> 통계분야별 단위 유형 현황

통계분야	단위 유형별 통계 개수(개)									‘기타’ 비율
	계	개인	가구	사업체	기업체	기타	보고통계	가공통계	조사통계	
전체	1,379	333	236	237	76	497	374	81	42	-
국민계정	23			1		22		22		96%
국토이용	12		1	1	1	9	5	1	3	75%
정부·재정	21	7				14	13		1	67%
주거	15	1	4			10	7	1	2	67%
환경	42	1	1	12	2	26	16	5	5	62%
교통·물류	28	7		6	2	13	11		2	46%
금융	17		3	2	5	7	7			41%
경제일반·경기	21	1	1	8	3	8		7	1	38%
무역·국제수지	19	2		8	2	7	3	4		37%
수산	20	1	3	8	1	7	3	1	3	35%
교육·훈련	18	10	2			6	4		2	33%
물가	9	1		4	1	3		2	1	33%
범죄·안전	35	12	4	6	2	11	11			31%
농림	49	3	18	12	1	15	5	2	8	31%
보건	64	27	5	12	1	19	12	6	1	30%
과학·기술	22	1		5	10	6	3		3	27%
문화·여가	32	12	4	8		8	5	1	2	25%
복지	31	16	8			7	4	2	1	23%
건설	18			10	4	4	4			22%
에너지	17		2	9	3	3	1	1	1	18%
광업·제조업	20			14	3	3		1	2	15%
인구	12	8	2	1		1		1		8%
정보통신	40	6	6	16	9	3	2	1		8%
사회일반	35	21	10	1	1	2	2			6%
도소매·서비스	18			15	2	1			1	6%
노동	36	13	3	15	4	1		1		3%
기업경영	30	1		13	16					0%
소득·소비·자산	8	2	6							0%
임금	10	1		7	2					0%
지역통계	657	179	153	43	1	281	256	22	3	43%

출처: 국가데이터처, 통계정책시스템 [웹사이트].

단위 유형이 ‘기타’인 통계 497개 중 지역통계(281개)를 제외한 216개 통계는 주로 보고통계(374개), 가공통계(81개)에 해당하며, 조사통계는 42건에 해당한다. 따라서, 작

성방법에 따른 개념 관리 대상 범위에 대한 고려와 함께 Unit Type 표준값 작성 검토가 이루어져야 할 것으로 사료된다.

<표 31> 조사(작성)단위가 ‘기타’이면서 ‘조사통계’인 사례

통계분야(주)	통계명	공표주기	조사(작성)단위 비고
국토이용	국가해안선조사	1년	
	상업용부동산임대동향조사	분기	빌딩
	전국지가변동률조사	월	토지
정부·재정	통계인력및예산조사	2년	정부승인통계 작성기관(인력 및 예산)
주거	전국주택가격동향조사	월	주택
환경	영농폐기물조사	1년	농가+재활용업체
	전국폐기물통계조사	5년	주택, 사업장, 상가, 학교, 재활용기능자원집하장 등
	수질오염실태보고	월	하천, 호수 등
교통·물류	도로교통량조사	1년	고속국도, 지방도, 일반국도 등
수산	연안어업실태조사	1년	연안어업 어선
	어촌계현황조사	1년	어촌계
물가	소비자물가조사	월	품목
농림	농업면적조사	1년	경지
	가축동향조사	분기	농장
	농작물생산조사	1년	농경지에 재배되고 있는 작물
보건	의료기관별급여적정성평가현황	1년	의료기관
복지	장애인편의시설설치현황조사	5년	공원, 공공건물 및 공공이용시설, 공동주택 등 대상시설
광업·제조업	전국산업단지현황통계	분기	산업단지
	광물자원매장량현황	2년	광산

출처: 국가데이터처, 통계정책시스템 [웹사이트].

2. Universe

현행 통계설명자료서비스에서는 ‘작성 대상 범위’에서 Universe 개념을 작성하고 있는 것으로 보인다. 다만, 동일한 조사안에서도 문항에 따라 응답대상이 다를 수 있으므로 마이크로데이터 이용자가 변수의 작성 대상을 명확히 파악하기 위해서는 개별 변수 수준에서의 조사대상 관리가 필요하다.

해외에서는 캐나다, 호주, 미국 등 다수 국가에서 마이크로데이터 변수별 Universe 제공 사례를 확인할 수 있었다.³⁴⁾ 캐나다, 호주는 ‘만 15세 이상 인구’와 같이 서술형 개념어로 작성하고, 미국은 ‘PRPERTYP=2 AND PRIOELG=1’와 같은 변수명을 사용한

34) 통계용어는 캐나다는 ‘Universe’, 호주는 ‘Population’, 미국은 ‘Edited Universe’로 상이함.

조건식으로 작성한다는 차이는 있다. 조건식은 마이크로데이터 분석을 위한 정확성 측면에서는 유리하지만, 변수를 직관적으로 이해하기에는 어려울 수 있다. 통계표 이용자 입장에서 변수를 이해하기 위해서는 개념어 작성방식이 보다 유용할 수 있다. 이때 개념어는 Unit Type을 기반으로 이를 한 단계 더 구체화한 수준에서 작성하는 것이 GSIM 설계 취지에 부합한다. 예컨대, Unit Type이 ‘개인’일 때, Universe는 ‘취업자’ 등으로 작성할 수 있다. 다만, 개념어를 어느 정도 구체적으로 작성해야 할지 표준 작성지침을 제시하는데 어려움이 있을 수 있는데³⁵⁾, 조사표의 질문 블록 수준에서의 작성 권고도 하나의 실무 지침으로 고려해 볼 수 있다.

Universe 작성 표준 지침은 필요하겠으나, Unit Type과 달리 변수 Universe는 각 조사의 맥락이 반영되는 것이므로 Universe 작성 표준값을 정립하고 적용하는데 실익이 크지 않을 것으로 생각된다. 데이터를 분석하고 통계결과를 이용하는 측면에서는 변수 Universe가 매우 유용한 정보이나, 통계 생산자 입장에서는 실무적인 부담이 따른다. 또한, 산출물이 마이크로데이터인지 KOSIS 통계표인지에 따라 표현방식의 적합성과 공표범위에 따른 작성 내용의 차이가 존재할 수 있다. 따라서, 표준 관리부서보다는 제공 영역에 따라 마이크로데이터 관리부서나 공표 통계표 관리부서의 검토·협업을 통해 통계 프로그램 담당자들이 작성해야 할 대상과 기준을 제시하는 것이 합리적일 것으로 생각된다.

<표 32> Universe 표현 방식 비교

개념어 방식		조건식 방식	
(캐나다 CPS 공공용 MD) Universe	변수 (개)	(미국 CPS 공공용 MD) Edited Universe	변수 (개)
전체	60	전체	190
All respondents	13	ALL HHLD's IN SAMPLE	12
Currently employed only	7	PRPERTYP = 1, 2, OR 3	12
Currently employed or worked within the past 12 months	4	PRPERTYP = 2	11
Currently employed, absent from work	2	ALL HHLDs IN SAMPLE	8
Currently employed, employees	5	PRIOELG = 1	6
Currently employed, part-time usual work hours at their main or only job was below 30 per week	1	PROELG = 1 AND PEMJOT = 1 AND HRMIS = 4 OR 8	5
Currently unemployed	2	PRERELG = 1	5
Currently unemployed, job searchers	7	ALL	5
⋮	⋮	⋮	⋮

출처: Statistics Canada(2026), Labour Force Survey (LFS) PUMF Codebook; U.S. Census Bureau(2025), CPS PUMF Data Dictionary.

35) 46-47페이지 <그림 25>와 <그림 26>으로 미루어 볼 때, 같은 개념어 서술 방식이라 하더라도 캐나다 통계 프로그램별 Universe 작성 방식은 소폭 차이가 있음.

3. Code List

통계법에서 지정 및 준수 의무를 규정하고 있는 한국표준산업분류, 한국표준직업분류 등의 표준분류는 GSIM의 Statistical Classification에 해당한다. 변수에 허용가능한 값 집합을 제공하고 개념 조화에 도달하기 위해서는 이보다 더 유연한 개념인 Code List까지 표준 관리 대상에 포함하여 코드목록 간의 관계와 그 의미를 이용자가 명확히 확인할 수 있도록 하는 것이 핵심 요건으로 생각된다.

국내 조사표에서의 ‘혼인상태’ 측정 사례를 보면, 범주의 개수도 다를뿐더러 같은 범주를 표현하는 지칭도 다소 차이가 있다. 예컨대, 2020 인구주택총조사의 ‘배우자있음’, 경제활동인구조사의 ‘유배우’, 2024 생활시간조사 ‘배우자있음(동거포함)’, 2024 전국다문화가족실태조사 ‘배우자있음(사실혼포함)’은 모두 같은 개념의 범주이지만, 세부 지침을 확인하지 않는다면 범주명만으로 이를 파악하기 어렵다. 따라서 혼인상태에 대한 표준 코드목록을 정의하고, 각 통계 프로그램에서는 표준과의 대응표를 작성함으로써 통계 간 비교가능성을 높일 수 있다. <표 33>에서 확인되는 세부 범주들은 ‘미혼’, ‘기혼’, ‘유배우’, ‘동거’, ‘사별’, ‘이혼’, ‘별거’ 등으로 모아진다. 이렇게 고려할 수 있는 세부 범주들을 검토하되, 표준 준수 의무, 비교가능성을 위한 구심점 역할을 고려하여 대표적인 범주로 구성된 ‘혼인상태’ 표준 코드목록 지정이 필요할 것이다.

<표 33> 국내 혼인상태 조사표 문항 범주 사례

구분	조사표 사례				
배우자 상태 (= 4개)	'20 인구주택총조사	경제활동인구조사	'24 생활시간조사	'24 전국다문화가족실태조사	
	1.미혼 2.배우자있음 3.사별 4.이혼	1.미혼 2.유배우 3.사별 4.이혼	1.미혼 2.배우자있음(동거포함) 3.사별 4.이혼(별거포함)	1.미혼 2.배우자있음(사실혼포함) 3.배우자사망 4.이혼·별거	
배우자 상태 (≠4개)	'15 인구주택총조사	'23 청소년종합실태조사	'25 장애인실태조사	'24 한부모가족실태조사	'24 노숙인등의실태조사
	1.미혼 2.배우자있음 3.동거 4.사별 5.이혼 6.별거	1.배우자있음 2.배우자없음 (미혼, 이혼, 사별 포함)	1.미혼 2.배우자있음 (사실혼, 동거포함) 3.이혼 4.별거 5.사별	1.배우자있음(배우자와함께거주(사실혼포함)) 2.미혼·비혼 3.이혼 4.사별 5.별거(가출또는생사불명포함, 단순분거는제외) 6.기타(사유기재)	1.비해당(만18세미만) 2.미혼(18세이상) 3.유배우 4.동거 5.별거 6.사별 7.이혼
결혼 경험	'23 인적자본기업패널조사	제7차('23년) 근로환경조사			
	1.미혼 2.기혼(재혼포함) 3.이혼또는사별	1.결혼함 2.결혼안함 3.모름/무응답 9.거절			

출처: 국가데이터처, 통계정책시스템 [웹사이트].

해외사례를 통해서 두 가지 시사점을 얻을 수 있었다. 하나는 코드목록의 지위이고, 다른 하나는 코드목록을 서비스하는 방식이다. 노르웨이는 Statistical Classification과 Code List를 엄격히 구분하는 반면, 캐나다는 이를 Classification으로 통칭하되 의무 수준에 따라 차별화한다. 우선, 노르웨이처럼 통계분류와 코드목록을 구별하여 통계 프로그램 담당자 및 이용자 입장에서 현행 통계분류(표준분류, 특수분류, 일반분류)와 혼동되지 않도록 관리 및 안내가 필요하다. 더불어 표준 지정 및 준수에 대한 안정성과 실무적용 가능성을 고려하여, 캐나다와 같이 의무 준수 수준을 차별화할 것을 제안한다.

코드목록에 대한 구체적인 서비스 방식은 정보시스템 구현에서 검토되겠지만, 어떤 정보를 어떻게 제공해야 할지에 따라 통계 프로그램 및 표준별로 필요한 정보가 달라 지므로 방향성 수준에서의 서비스 방안에 대한 검토가 필요하다. 해외사례의 이점만을 종합하여, 노르웨이처럼 개념에 대한 범주 위계구조를 한눈에 볼 수 있도록 구성하고, 캐나다처럼 범주명만으로 설명이 되는 경우를 제외하고는 개별 범주에 대한 정의를 제공할 것을 제안한다. 범주 정의를 통해 범주가 의미하는 바를 제대로 이해할 수 있을 뿐 아니라³⁶⁾, 개별 범주가 변수로서 측정되는 경우에는 다소 나열적으로 제공될 수밖에 없는 변수의 개념을 코드목록 하에서 보다 체계적으로 이해할 수 있는 이점까지 기대할 수 있다. 예컨대, 변수-범주 간 연계를 통해 코드목록 하에서 ‘취업자’, ‘임금근로자’라는 개별 변수가 상하 관계에 해당한다는 구조적 맥락을 이해할 수 있다.

통계 프로그램에서 작성해야 하는 표준 코드목록과의 대응표와는 별개로, 표준 코드목록 간 대응표 및 버전·변경이력이 기본적인 틀로 제공될 필요가 있다. 즉, 서로 독립적으로 발전해 온 코드목록 간의 관계를 명시적으로 확인할 수 있는 대응표가 필요하다. <표 34>의 ‘Classification of Employed Status’에서 주석을 통해 ‘Not Employed’가 비경제활동인구를 포함한다는 정보를 제공하고 있지만, 코드목록간 대응표를 통해 ‘(Collapsed) Classification of Labour Force Status’와의 범주 간 관계를 보다 명시적으로 이해할 수 있을 것이다. 다만, 하나의 표준 코드목록에서 범주 통합·추가와 같은 경미한 변형이 있는 경우에는, 캐나다와 노르웨이 사례처럼 원분류의 변형(Variant)으로 관리할 수 있다.

36) 노르웨이 ‘Codelist for gender’는 {Both genders(0), Male(1), Female(2)}로 구성되는데, 범주별 정의가 제공되지 않아서 ‘Both genders(0)’가 자기 결정에 따른 성별 정체성을 의미하는 것인지, 집계용 총계 의미인지 모호함.

<표 34> 캐나다 '노동력상태' 관련 분류 간 대응관계

Classification of Labour Force Status		Collapsed Classification of Labour Force Status		Classification of Employed Status*	
코드	범주	코드	범주	코드	범주
1	In labour force				
11	Employed	1	Employed	1	Employed
111	Employed and at work				
112	Employed and not at work				
12	Unemployed	2	Unemployed	2	Not Employed
2	Not in labour force	3	Not in labour force		

* (Usage) When using the class term Not employed in a table or text, a **footnote should be included** to state that Not employed includes persons who were unemployed as well as those who were not in the labour force.

출처: Statistics Canada, Classifications, variables and statistical units [웹사이트].

4. Variable

서론에서 언급하였듯이 현행 각종 통계설명자료는 개념 및 변수 정보 작성의 구조화가 다소 미진하다. PDF 파일로 제공되는 이용자용 통계정보보고서는 '1-1. 주요 용어 및 항목별 정의' 및 '1-2. 주요 용어에 대한 국내·국제기준 비교' 챕터를 작성하여야 하나 상세한 작성 틀이 부재하여 통계 프로그램마다 편차가 발생한다. 통계설명자료서비스의 용어 조회 메뉴에서 활용되는 조사 용어도 통계 프로그램별 일관성이 매우 부족하다. 통계용어사전은 선별된 통계용어에 대해 정의, 해설, 관련용어 등을 정제하여 제공하고 있으나, 관련통계명과 국가통계포털 통계표와의 연계가 아쉽고, 단위유형, 변경이력, 개념체계에 대한 정보가 구조적으로 제공되지 않는다.

제2장에서 설계와 구현을 분리하여 개념의 재사용성과 비교가능성을 확보하고자 하는 GSIM의 핵심 원칙을 확인할 수 있었다. 제3장 해외사례에서는 변수에 대한 구조화된 정보를 제공하되 각국의 기조와 상황에 맞게 현지화하여 적용하고 있음을 확인하였다.

<표 35> 캐나다·노르웨이 변수에 대한 제공 정보

노르웨이 변수	캐나다 표준 변수	캐나다 통계 프로그램 변수
Variable Name	Variable Title (Statistical unit 포함)	Variable Title (Statistical unit 포함)
Statistical unit		
Definition	Definition	Definition
	Usage	
	Derivation	
	Conformity to relevant internationally recognized standards	
Subject	Subject (목록에서만 조회)	Subject (목록에서만 조회)
-	Classifications (Link)	using classifications (Link)
Linked to Variable Definition (Link)	-	-
Linked to StatBank tables (Link)	-	-
Valid From	Relation to previous version (Link)	-
Valid to		-
Owner	-	-
-	Status	-

출처: Statistics Canada, Classifications, variables and statistical units; Statistics Norway, Metadata [웹사이트].

<표 36>은 GSIM 정보 클래스를 적용하여 시범적으로 작성해 본 결과이다(정의 등 정보 클래스별 내용은 간략하게 기술하였음). 우선 노동 영역의 대표적 추상적 개념인 Employment에서 시작하여, 어떤 단위유형을 측정하는지에 따라 ‘개인의 취업여부’와 ‘사업체의 종사자 수’라는 서로 다른 개념변수를 정의하였다. 변수 간 실질적인 비교가능성과 재사용성을 동시에 확보하기 위해서 조작적 정의를 작성하였다. GSIM에서는 개념변수 및 표현변수의 속성에 ‘정의’를 포함하고 있지 않지만, 최상위 정보 클래스인 Identifiable Artefact의 속성을 상속받을 수 있으므로 Description 속성을 활용하여 GSIM의 설계안에서 조작적 정의를 나타낼 수 있을 것으로 생각된다.³⁷⁾

37) 기존 변수를 활용한 파생변수는 비즈니스 그룹의 Rule 정보 클래스로 설명할 수도 있으나, 보다 범용적으로 적용하기 위해서는 Identifiable Artefact의 Description 속성이 합당할 것으로 보임

<표 36> GSIM 정보 클래스 시범 적용

주제영역		노동			
개념	명칭	Employment			직장 규모
	정의	임금 또는 수익을 위해 재화를 생산하거나 서비스를 제공하는 활동에 참여			직장에 종사하는 인력 규모
개념변수	단위 유형	개인	사업체		개인
	명칭	개인의 취업여부	사업체의 종사자 수		개인의 직장규모
	개념도메인	{취업, 미취업}	인원수	인원수	{1~4명, 5~9명, ..., 300~499명, 500명이상}
	설명 (조작적 정의)	수입목적1시간이상 근로 or 18h이상 무급가족일 or 일시휴직	소속외근로자(파견, 용역), 3개월 이상 장기 결근자 제외	자영업자, 무급가족종사자, 소속외근로자(파견, 용역) 제외	해당 본/지점만 파악, 고용주·자영자·무급가족종사자 포함
표현변수	값도메인	{1 취업, 2 미취업}	{0 이상의 정수}	{0 이상의 정수}	{1. 1~4명, 2. 5~9명, ..., 6. 300~499명, 7. 500명이상}
	측정단위	-	명	명	명
유니버스		만 15세이상 인구	모든 사업체	종사자 1인 이상을 고용하고 있는 모든 사업체	만 15세이상 취업자
인스턴스	조사모집단	특수시설조사구 제외	표준산업분류 T, U 제외	표준산업분류 A, T, U 제외 (공무원 포함)	특수시설조사구 제외
	통계프로그램	경제활동인구조사, 지역별고용조사	전국사업체조사	사업체노동력조사	경제활동인구조사, 지역별고용조사
	관련통계표	Link	Link	Link	Link

한편, 개념 도메인 작성을 시도해 보았으나, 앞서 혼인상태라는 개념에 대해 다양한 범주 구성이 가능하다는 점에서도 알 수 있듯이 개념변수가 하나의 개념 도메인만을 참조해야 한다는 GSIM의 관계 설정은 개념 재사용성에 제약 요소로 작용할 것으로 보이며 캐나다와 노르웨이 사례에서도 개념 도메인을 별도로 분리하여 정의하고 있지 않다. 따라서 본 연구에서는 개념 도메인은 관리 대상에 포함하지 않고자 한다. 한편, 값도메인은 통계 프로그램 단위에서는 비교가능성 확보를 위해 작성하되 표준에서는 코드목록과의 연계로만 표시하는 방안을 제시한다. 이를 반영하여 표준 변수와 통계 프로그램 변수 각각에 필요한 정보를 <표 37>, <표 38>과 같이 제안한다.

<표 37> 표준 변수에 대한 설명자료 시범 작성

주제영역	노동	
개념	Employment	
개념 정의	임금 또는 수익을 위해 재화를 생산하거나 서비스를 제공하는 활동에 참여	
단위유형	개인	사업체
개념변수	개인의 취업여부	사업체의 종사자 수
개념변수 설명	수입목적1시간이상근로 or 18h이상 무급가족일 or 일시휴직	소속외근로자(파견, 용역) 제외
관련 코드목록 ¹⁾	Link	Link
관련 통계 프로그램 변수 ²⁾	Link	Link
관련 통계표	Link	Link

1) 제4장 2절 3에서 구축 제안한 표준 코드목록에 대한 Link(s)

2) <표 38>과 같이 구축된 통계 프로그램 변수 설명자료에 대한 Link(s)

3) 표준 변수가 사용된 국가통계포털 통계표에 대한 Link(s)

<표 38> 통계 프로그램 변수에 대한 설명자료 시범 작성

주제영역	노동																	
개념변수	개인의 취업여부	사업체의 종사자 수		개인의 직장규모														
단위 유형	개인	사업체		개인														
변수 설명	수입목적1시간이상 근로 or 18h이상 무급가족일 or 일시휴직	소속외근로자(파견, 용역), 3개월 이상 장기 결근자 제외	자영업자, 무급가족종사자, 소속외근로자(파견, 용역) 제외	해당 본/지점만 파악, 고용주·자영자· 무급가족종사자 포함														
값 도메인	{1 취업, 2 미취업}	{0 이상의 정수}	{0 이상의 정수}	LINK <table><tr><td>1</td><td>1~4명</td></tr><tr><td>2</td><td>5~9명</td></tr><tr><td>3</td><td>10~29명</td></tr><tr><td>4</td><td>30~99명</td></tr><tr><td>5</td><td>100~299명</td></tr><tr><td>6</td><td>300~499명</td></tr><tr><td>7</td><td>500명 이상</td></tr></table>	1	1~4명	2	5~9명	3	10~29명	4	30~99명	5	100~299명	6	300~499명	7	500명 이상
1	1~4명																	
2	5~9명																	
3	10~29명																	
4	30~99명																	
5	100~299명																	
6	300~499명																	
7	500명 이상																	
측정단위	-	명	명	명														
통계 프로그램	경제활동인구조사, 지역별고용조사	전국사업체조사	사업체노동력조사	경제활동인구조사, 지역별고용조사														
관련 통계표1)	Link	Link	Link	Link														

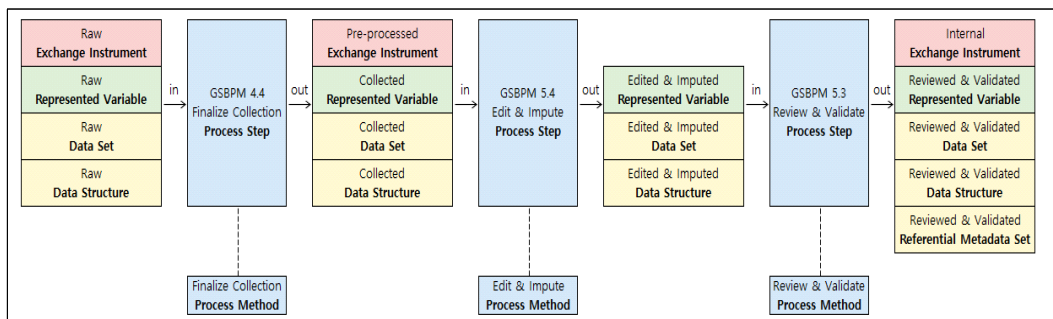
1) 통계 프로그램 변수가 사용된 국가통계포털 통계표에 대한 Link(s)

제3절 실무 적용 고려사항

1. 업무 프로세스에 따른 개념 진화

데이터는 업무 프로세스(Business Process) 전반을 거치며 진화한다. 원시 데이터가 점차 정제된 데이터로 이동함에 따라, 프로세스 단계(Process Step)별 프로세스 방법(Process Method)을 통해 새로운 데이터세트와 변수가 생성된다. 표현변수의 경우, 설계, 수집, 내용검토(editing), 자료처리 단계 등을 거치면서 여러 버전의 형태로 존재³⁸⁾하며, 단위 유형이 파생되거나 신규 변수가 파생되기도 한다.

<그림 34> 업무 프로세스에 따른 정보 클래스 입·출력 흐름 예시



출처: UNECE Statistics, GSIM 2.0 Webinar 2025 [YouTube] (2025.9. 게시) 재구성.

1) 정보 작성이 필요한 변수 범위 - 조사표, 마이크로데이터, 공표통계표

조사표 질문(Question)은 개념변수에 기반³⁹⁾하여 작성되는 정보 클래스로서, 동일한 개념을 측정하더라도 질문 설계는 조사별로 상이할 수 있다. <그림 35> 사례와 같이, 조사에 따라 ‘취업여부’ 개념을 측정하는 질문이 1개 또는 4개이거나, 하나의 질문에서 ‘취업’으로 판정될 수 있는 질문 값 도메인이 1~2개로 상이하다. 개념(Concept)은 정의(특성에 따라 구분되는 사고 단위)에 따르면 조사표 질문도 각각 하나의 개념으로 대응될 수 있으나, 어느 정도까지의 개념을 설명자료 작성범위에 포함할 것인지에 대한 지침이 필요하다.

38) <부록 11-1>, <부록 11-2>의 미국 CPS 공공용 마이크로데이터와 같이 변수명과 플래그 변수값을 통해 어떤 자료처리가 수행되었는지 정보를 제공하기 위해서는 통계 생산 시스템의 재편이 필요한 사항이므로 조사시스템 재편 가능성 및 효과성 등에 대한 장기적인 검토가 필요함

39) GSIM 다이어그램에 ‘Question is based on Conceptual Variable’ 관계가 나타남

<그림 35> 취업여부를 측정하는 조사표 사례

6 지난주에 수입을 목적으로 1시간 이상 일하였습니까?

1. 예 → (**13** 번으로) 2. 아니오

7 지난주에, 함께 사는 가족의 사업체 또는 농장에서 보수를 받지 않고 1시간 이상 일하였습니까?

1. 예 → (**13** 번으로) 2. 아니오

8 지난주에 일을 하지는 않았지만 직장(일)을 가지고 있었습니까?

1. 있었음 2. 없었음 → (**29** 번으로)

8-1 지난주에 일을 하지 않는 이유는 무엇입니까?

① 일시적 병, 사고 → (**10** 번으로)

② 휴가, 연가

③ 교육 훈련

④ 육아

⑤ 가족적 이유

⑥ 노사분규

⑦ 사업부진, 조업중단 → (**29** 번으로)

⑧ 기타()

12 지난주에 실제 몇시간 일하였습니까?

1. 주된 일 : 시간

2. 다른 일 : 시간

3. 총 계 : 시간

① 1 ~ 17시간 무답가측종사자 → (**29** 번으로)

② 0 ~ 35시간 → (**13** 번으로)

③ 36시간 이상 → (**17** 번으로)

1-1 귀하는 지난 1주간(8월 1일~ 8월 7일) 주로 무엇을 하였습니까?

① 일하였음(일시휴직 포함) → **2-1** 번으로

'일하였음'과 '일시휴직'은 아래와 같은 경우를 말합니다.

- 일하였음: ① 임금이나 수입(소득)을 목적으로 1주일에 1시간 이상 일했거나
- ② 가족이 경영하는 사업체, 농장 등에서 무보수로 1주일에 18시간 이상 일한 경우 (가족: 혈연, 혼인, 입양 등으로 이루어진 사람들의 집단, 또는 그 구성원)
- 일시휴직: 자신이나 가족의 병, 사고, 휴가(연가), 교육, 노사문제 등으로 일시적으로 일을 하지 않으나 복귀가 가능한 경우

② 구직활동(이력서 제출, 면접 등 실제활동)

③ 육아

④ 가사

⑤ 정규교육기관 통학

⑥ 임시학원 통학

⑦ 취업을 위한 학원·기관 통학 (고시학원, 직업훈련기관 등)

⑧ 취업준비(자격증 시험, 공무원 시험 준비 등)

⑨ 진학준비

⑩ 일시적 병, 사고

⑪ 쉬었음

⑫ 기타 (직접 작성해 주십시오)

→ **2-2** 번으로

<a> 경제활동인구조사(2025년)

 여성의 경제활동 및 경력단절 실태조사(2025년)

8 ~ 14 는 만 15세 이상(2009년 2월 28일 이전 출생)만 작성합니다.
지난 1주일 동안(2024. 3. 8. ~ 3. 14.)의 경제 활동에 관한 내용입니다. ※ 는 통계청에서 작성합니다.

8 지난 1주일 동안 수입을 목적으로 1시간 이상 일을 하였습니까?

• 돈을 받지 않고 동일 가구 내 가족이 경영하는 사업장 등에서 18시간 이상 일 한 경우도 「1」 일을 하였음」에 표시합니다.

1 일을 하였음

2 휴가 또는 일시 휴직

3 일을 하지 않았음 →

8-1 일을 하지 않은 가장 주된 이유는 무엇입니까?

① (손)자녀 양육 ② 가사((손)자녀 양육 제외)

③ 학업(학교, 학원 등) ④ 취업 준비

⑤ 정년퇴직, 연로 ⑥ 심신 장애

⑦ 직장의 휴폐업 ⑧ 기타()

⇨ 응답 후 **14** 번으로 이동

<c> 생활시간조사(2024년)

* 붉은 음영은 취업에 해당하는 조건을 표시

출처: 국가데이터처, 통계설명자료서비스 [웹사이트].

국가통계 개념의 비교가능성 및 재사용성 제고라는 목적과 실무적인 관리 가능성을 고려했을 때, 조사표 문항 각각보다는 상위 수준에서 개념을 관리할 것을 고려할 수 있다.⁴⁰⁾ 즉, 질문에 대응하는 입력자료보다는 공표 통계표에 사용되는 개념 및 변수를 우선적으로 포함하는 것이다.

40) Question(질문)은 교환 그룹에 해당하는 정보 클래스로서, 이에 대한 재사용성에 대해서는 별도의 관리 방안 연구가 필요한 영역임

다만, 조사표와 공표 통계표의 대응관계가 1:1에 가까운 통계도 있을 수 있어서, 정보 작성이 필요한 변수 선정에 있어서 통계 프로그램 담당자에게 어느 정도 자율 부여는 필요할 것으로 생각된다.

<그림 36> 조사표 문항이 공표 통계표로 대부분 대응되는 사례

통계목록 ▼ 단어검색

맨위로

오름차순

검색

청년의삶실태조사

- ☞ 일반사항
- ☞ 주거
- ☞ 건강
- ☞ 교육·훈련
- ☞ 노동
 - 수입을 목적으로 1시간 이상 일한 경험 (2년 2022~2024)
 - 무급가족종사자 여부 (2년 2022~2024)
 - 일시휴직자 여부 (2년 2022~2024)
 - 일시휴직 이유 (2년 2022~2024)
 - 복수일자리 해당 여부 (2년 2022~2024)
 - 복수일자리 수 (2년 2022~2024)
 - (복수일자리) 일자리 경험 (2년 2022~2024)
 - 현재 일자리 종사상 지위 (2년 2024~2024)
 - 사업 소유 유무 (2년 2022~2022)
 - 고용 계약기간 유무 (2년 2024~2024)
 - 고용 계약기간 (2년 2022~2022)

<a> 청년삶실태조사 공표 통계표 목록

← E1

← E1-1

← E1-1-1, E1-1-2

← E1-1-2

← E2

← E2-1

← E2-1-1

← E3

← E3

← E3-1

← E3-1-1

E1 지난해 수입을 목적으로 1시간 이상 일하셨습니까?

① 예 ☒ E2 **회으로 이동** ② 아니요 ☐

E1-1 지난해에 함께 사는 가족의 사업체 또는 농장에서 보수를 받지 않고 1시간 이상 일하셨습니까?

① 예 ☒ E2 **회으로 이동** ② 아니요 ☐

E1-1-1 지난해에 일하지는 않았지만 직장(일)을 가지고 있었습니까?

① 예 ☐ ② 아니요 ☒ E15 **회으로 이동**

E1-1-2 지난해에 일하지 않은 이유는 무엇입니까?

① 일시적 병, 사고
② 휴가, 연가
③ 교육, 훈련
④ 육아
⑤ 가족적 이유
⑥ 노사분규
⑦ 사업 부진, 조업 중단
⑧ 기타()

← 응답 후 E2 회으로 이동

E2 평소 수입을 목적으로 일하는 직장(일)이 두 개 이상입니까?

• 프리랜서로 같은 일을 여러 곳(정소, 업체, 고객)에서 하고 있는 경우 하나의 일로 보고 응답합니다. 다만, 일의 내용이 다를 경우에는 다른 일로 보고 응답합니다.

① 예 ☐ ② 아니요 ☒ E2-1-1 **회으로 이동**

E2-1 총 몇 개의 일자리를 갖고 있습니까?

총 개 **← 응답 후 E2-1-1 회으로 이동**

E2-1-1 모든 일자리에 대하여, 가장 많은 시간 일하는 일자리부터 순서대로 응답해주세요.

• 무급가족종사자인 경우에도 두 개 이상의 일을 할 경우 응답해주세요.

구분	첫 번째 일자리	두 번째 일자리	세 번째 일자리
1) 일주일에 며칠을 일하십니까?	주당 일	주당 일	주당 일
2) 일주일에 몇 시간을 일하십니까?	주당 시간	주당 시간	주당 시간
3) 이 일자리에서 "세금 공제 전" 소득은 얼마입니까?	월 만원	월 만원	월 만원
4) 어떤 일을 합니까?	① 관리자 ② 전문가 및 관련 종사자 ③ 사무 종사자 ④ 서비스 종사자 ⑤ 판매 종사자	⑥ 농림·어업·숙련 종사자 ⑦ 기능원 및 관련 기능직 ⑧ 장차 기계 조작 및 조립 종사자 ⑨ 단순노동종사자 ⑩ 군인	

 2024년 청년삶실태조사 조사표

출처: 국가데이터터, 국가통계포털; 통계설명자료서비스 [웹사이트].

역으로, <표 39>에서와 같이 교육정도(컨버전코드)를 작성하기 위해 측정해야 하는 학력구분코드, 수학기분코드, 계열코드는 다양한 통계에서 활용되는 범주형 변수이므로 표준화가 필요하지만, 그 측정 결과가 공표 통계표에 직접 드러나지 않는다. 따라서, 공표되는 변수와 그밖에 인구통계학적 변수를 포함하는 방안, 혹은 마이크로데이터에 제공되는 변수는 포함하는 방안 등 정보 작성이 필요한 변수 범위에 대한 지침 마련이 필요하다.

<표 39> 경제활동인구조사 항목별 자료구분

자료구분 ¹⁾			항목명	값 도메인 ²⁾
조사표	MD	KOSIS		
37개	33개	26개		
명부	1	-	동부읍면부코드	1. 동부 / 2. 읍면부
명부	51	공표	조사연월	
명부	52	공표	농가비농가가구구분코드	1. 농가 / 2. 비농가
1	2	-	가구주관계코드	1. 가구주 / 2. 배우자 / 3. 미혼자녀 / 4. 기혼자녀 / 5. 손자녀 / 6. 부모(장인,장모) / 7. 조부모 / 8. 미혼형제자매 / 9. 기타
2	3	공표	성별코드	1. 남자 / 2. 여자
-	54	공표	만연령	
-	55	공표	연령5세단위코드	01. 15~19세 / 02. 20~24세 / (...중략...) / 12. 70~74세 / 13. 75세이상
3	4	-	출생연도	
3	-	-	출생월	
3	-	-	출생일	
-	53	공표	교육정도컨버전코드	1. 초졸이하 / 2. 중졸 / 3. 고졸 / 4. 초대졸 / 5. 대졸 / 6. 대학원졸
4	5	-	교육정도_학력구분코드	0. 무학 / 1. 초등학교 / 2. 중학교 / 3. 고등학교 / 4. 전문대(초급대, 2·3년제 대학포함) / 5. 대학교(4년제 대학포함) / 6. 대학원(석사) / 7. 대학원(박사)
4	7	-	교육정도_수학기분코드	1. 졸업 / 2. 재학 / 3. 중퇴 / 4. 휴학 / 5. 수료
4	6	-	교육정도_계열코드	001. 인문계열 / 002. 예술,체육계열 / (...후략...)
4	8	-	졸업연도	
5	9	공표	혼인상태코드	1. 미혼 / 2. 유배우 / 3. 사별 / 4. 이혼
6	10	-	현재일_지난주수입목적근로여부	1. 예 / 2. 아니오
7	11	-	현재일_지난주무급가족일근로여부	1. 예 / 2. 아니오
8	12	-	현재일_지난주일시휴직여부	1. 있었음 / 2. 없었음
8-1	13	-	현재일_지난주일시휴직사유코드	1. 일시적 병, 사고 / 2. 연가, 휴가 / 3. 교육, 훈련 / 4. 육아 / 5. 가족적이유 / 6. 노사분규 / 7. 사업부진, 조업중단 / 8. 기타
9	14	-	현재일_지난주부업여부	1. 있었음 / 2. 없었음
10	-	-	현재일_평소1주간근무시간	
10	-	-	현재일_평소1주간근무시간구분코드	1. 0~35시간 / 2. 36시간 이상
11	15	-	현재일_평소1주간36시간미만근무사유코드	1. 정규 근무시간이 36시간 미만 / 2. 건강 / 3. 육아 / 4. 가사 / 5. 통학 / 6. 본인이 원해서 / 7. 평소 일거리가 없어서 / 8. 기타()
12	16	-	현재일_주업실제취업시간수	
12	17	-	현재일_지난주부업근무시간수	
12	18	공표	현재일_총실제취업시간수	※ KOSIS 재코딩: 1-17/18-35/36-44/45-53/54+/-35/36+/1-14/15-35/36-52/53+/ 평균
12	19	-	현재일_지난주총근로시간구분코드	1. 1~17시간 무급가족종사자 / 2. 0~35시간 / 3. 36시간 이상
13	20	-	현재일_지난주36시간미만근무사유코드	01. 평소 1주일에 36시간미만 일함 / 02. 일시적 병, 사고 / 03. 알기불순 / 04. 연휴가 공휴일 / 05. 교육, 훈련 / 06. 육아 / 07. 가족적 이유 / 08. 노사분규 / 09. 일시적으로 일거리가 없어서 / 10. 사업부진, 조업중단 / 11. 기타()
14	21	-	현재일_추가취업전직희망코드	1. 현재 보다 시간을 늘리고 싶음 / 2. 현재 보다 다른일도 하고 싶음 / 3. 더 많이 일할수 있는 일로 바꾸고 싶음 / 4. 계속 그대로 일하고 싶음
15	22	-	현재일_추가취업가능성여부	1. 예 / 2. 아니오

자료구분 ¹⁾			항목명	값 도메인 ²⁾
조사표	MD	KOSIS		
37개	33개	26개		
16	23	-	현재일_지난4주내추가취업구직여부	1. 있었음 / 2. 없었음
17	-	-	현재일_사업체명	
17	-	-	현재일_사업체가주로하는일	
17	-	-	현재일_11차산업중분류코드	
-	24	공표	현재일_11차산업대분류코드	A. 농업, 임업 및 어업(01~03) / (...후략...)
18	25	-	현재일_종사자규모코드	1. 1-4명 / 2. 5-9명 / 3. 10-29명 / 4. 30-99명 / 5. 100-299명 / 6. 300명이상
20	27	공표	현재일_종사상지위코드	1. 상용근로자 / 2. 임시근로자 / 3. 일용근로자 / 4. 고용원이 있는 자영업자 / 5. 고용원이 없는 자영업자 / 6. 무급가족종사자
29	31	-	구직사항_지난4주내구직여부	1. 있었음 / 2. 없었음
30	32	-	구직사항_취업가능성유무	1. 있었음 / 2. 없었음
44	-	-	추가무급_추가무급가족일근로여부	1. 예 / 2. 아니오
-	-	공표	경제활동인구	※ 취업자 + 실업자
-	-	공표	취업자	※ 수입목적1시간이상근로 or (무급가족일 and 18h이상) or 일시휴직
-	-	공표	실업자	※ 취업자아님 and 지난4주간구직활동 and 지난주취업가능
-	-	공표	비경제활동인구	※ 취업자아님 and 실업자아님
-	-	공표	시간관련_추가취업가능자	※ 취업자 and 실제취업시간<36h and 추가취업희망 and 추가취업가능
-	-	공표	잠재경제활동인구	※ 경제활동인구 + 잠재경제활동인구
-	-	공표	잠재취업가능자	※ 잠재취업자 + 잠재구직자
-	-	공표	잠재구직자	※ 취업자아님 and 지난4주간구직활동 and (not_지난주취업가능)
-	-	공표	확장경제활동인구	※ 취업자아님 and (not_지난4주간구직활동) and 지난주취업가능
-	-	공표	18시간미만 취업자 중 추가취업희망자	※ 취업자 and 실제취업시간<18h and 추가취업희망
-	-	공표	36시간미만 취업자 중 추가취업희망자	※ 취업자 and 실제취업시간<36h and 추가취업희망
-	-	공표	경제활동참가율	※ (경제활동인구) / (15세이상인구) × 100
-	-	공표	실업률	※ (실업자) / (경제활동인구) × 100
-	-	공표	고용률	※ (취업자) / (15세이상인구) × 100
-	-	공표	고용보조지표1	※ (실업자 + 시간관련추가취업가능자) / 경제활동인구 × 100
-	-	공표	고용보조지표2	※ (실업자 + 잠재경제활동인구) / 확장경제활동인구 × 100
-	-	공표	고용보조지표3	※ (실업자 + 잠재경제활동인구 + 시간관련추가취업가능자) / 확장경제활동인구 × 100

1) [조사표]는 조사표 문항번호, [MD]는 마이크로데이터 파일설계서 연번을 표시함.

2) KOSIS 공표항목이 파생변수인 경우 파생규칙을 기재함.

출처: 국가데이터처(2025), 경제활동인구조사 조사표; 국가데이터처(2026), 경제활동인구조사 지침서 외부용; 국가데이터처(2026), 경제활동인구조사 MD 파일설계서; 국가데이터처, 국가통계포털 [웹사이트].

2) 차원 데이터 구조에만 존재하는 변수의 Unit Type 작성

실업률, 고용률 등 차원 데이터 구조(Dimensional Data Structure), 즉, 집계표 형태에서만 작성될 수 있는 파생 변수에 대한 Unit Type 값을 어떻게 작성해야 할지에 대한 검토가 필요하다. 차원 데이터에서 Identifier Component(식별자 구성요소)는 구조 그룹(Structure Group)에 해당하는 역할일 뿐이므로 파생 변수의 Unit Type으로 생각하기 어렵다. 해외 일부 변수를 조회해 본 결과, 캐나다는 ‘economy’, ‘location’, ‘geographic area’ 등 주제, 지역으로 혼재되어 있고, 노르웨이는 ‘재화·서비스’, ‘기업체’, ‘사업체’ 등 단위 데이터에서의 Unit Type으로 명시하는 사례를 확인할 수 있었다. 국내에서도 이에 대한 작성 지침 마련이 필요하다.

<표 40> 캐나다·노르웨이 차원 데이터 구조 변수에 대한 Unit Type 작성 사례

국가	변수	Unit Type	통계 프로그램
캐나다	Consumer Price Index	Economy	Consumer Price Index
	Unemployment rate	Geographic area	Labour Force Survey
	Job vacancy	Location	Job Vacancy and Wage Survey
노르웨이	Consumer Price	Repr. Goods and services	-
	Consumer Price Index - adjusted	Enterprise	-
	Job vacancy rate	Establishment	-

출처: Statistics Canada, Classifications, variables and statistical units; Statistics Norway, Metadata [웹사이트].

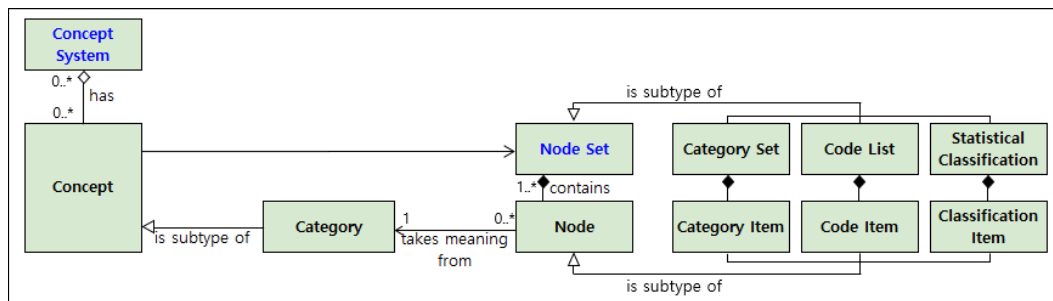
2. 개념의 체계적 이해를 위한 보조적 수단

1) Concept System

본 연구의 목적은 국가통계 개념 정보를 일관된 구조를 통해 체계적으로 관리하는데 있으며, 개념 자체에 대한 의미론적 체계화 작업은 연구 범위를 벗어난다. 다만, 개별 개념들이 국가통계의 전체 개념 구조에서 어떠한 위치에 있는지 파악하고, 동일 개념의 반복적인 생성이나 유사 개념의 중복 관리 여부를 식별하는 데 활용할 수 있는 보조적 수단으로서의 Concept System(개념체계) 활용 가능성을 간략히 검토해 보았다.

GSIM에서 Concept System은 Concept의 상위에 있는 정보 클래스로 ‘관계에 의해 구조화된 개념들의 집합’으로 정의되는데, 개념 간 상·하위 관계, 연관 관계, 부분·전체 관계 등 다양한 의미적 관계를 구조화하여 표현하는 수단으로 활용할 수 있을 것으로 보인다. 개념을 외연적으로 정의하는 범주(개념)는 앞서 표준 관리대상으로 선정한 Code List로 그룹화하여 관리 가능할 수 있으나, 변수 등 그 외 다른 개념 유형에 대해서는 Concept System으로 그룹화하는 접근을 고려할 수 있다.

<그림 37> 개념체계와 노드세트 관계



구성요소 및 관계, 표현방식 등에 대한 제약이 없어서 동일 주제 내에서 개념 간의 관련성을 유연하게 나타낼 수 있는 반면, 관리 주체, 표준 지정 여부 등에 대한 관리지침은 필요할 것으로 보인다.

앞서 48페이지 <그림 27>에서 살펴본 캐나다 2011년 국민가구조사 공공용 마이크로 데이터에서 133개 변수를 인구·가구·노동·소득 등 17개의 ‘Variable Groups’로 묶어 제공했는데, 이는 모든 통계에 공통 적용되는 표준은 아니지만 해당 통계 프로그램 내에서 관련 개념들을 하나의 체계로 조직화한다는 점에서 Concept System과 유사한 역할을 수행한다고 볼 수 있다. 그밖에 본 연구에서 사례로 다루었던 Employment, 혼인상태, 성별과 같이, 서로 관련되면서도 세분화 방식이 다른 개념들 역시 하나의 상위 체계로 묶어 관리할 수 있는 후보군이 될 수 있다.

다만 이러한 개념 간 의미 체계화 작업은 방대한 개념 전수조사와 합의 절차를 요구하므로, 본 연구에서는 그 필요성과 개략적인 적용 가능성을 제시하는 데 그치고, 구체적인 후속 연구는 향후 과제로 남겨둔다.

<표 41> 개념체계와 노드세트 비교

구분	Concept System (개념체계)	Node Set (노드세트)
정의	관계에 의해 구조화된 개념 들의 집합 set of Concepts structured by the relations among them	노드 들의 집합 set of Nodes
설명문	Here are 2 examples 1) Concept of Sex: Male, Female, Other 2) ISIC (the list is too long to write down)	Node Set is a kind of Concept System. Here are 2 examples: 1) Sex Categories: Male, Female, Other 2) Sex Codes: <m, male>, <f, female>, <o, other>
요소	Concept (Unit Type, Variable, Category 등)	Category
관계	제약 없음 (위계, 인과, 연관, 순서 등)	고정됨 (단일 특성을 외연적 정의)
표현	코드 부여 여부와 무관함	코드 유무에 따라, Code List 및 Category Set

출처: UNECE, GSIM Version 2.0 - 클래스 정의 [웹사이트]; 저자 해석.

2) Subject Field

통계정책시스템의 통계분야 및 국가통계포털 주제는 30개 주제 영역으로 제공되고 있다. GSIM에서는 Subject Field(주제 영역)를 정보 클래스에 상-하위 관계로 설정할 수 있다는 명시적인 표현은 없다(parent-child 관계선 부재). 다만, 관련 표준인 ISO 11179-1

에서 ‘주제 영역이 세분될 경우, 세분된 결과 역시 각각 하나의 주제 영역이 된다’는 설명에 근거하여 하위 주제 영역 작성을 통해 개념 탐색을 위한 보완적 효과 도모해 볼 수 있다.

<표 42> 한국 국가통계 주제 영역

주제 영역	건설	교통·물류	노동	물가	소득·소비·자산	정보통신
	경제일반·경기	국민계정	농림	범죄·안전	수산	정부·재정
	과학·기술	국토이용	도소매·서비스	보건	에너지	주거
	광업·제조업	금융	무역·국제수지	복지	인구	지역통계
	교육·훈련	기업경영	문화·여가	사회일반	임금	환경

출처: 국가데이터처, 통계정책시스템; 국가통계포털 [웹사이트].

실제로 캐나다와 노르웨이는 통계조회시스템에서는 <표 43>와 같이 주제 영역을 계층적으로 조회할 수 있도록 제공하고 있다.

<표 43> 캐나다·노르웨이 통계조회시스템에서 제공하는 ‘노동’ 하위 주제 영역

국가	주제 영역 (자료 수)
캐나다	Labour (1897)
	Earnings, wages and non-wage benefits (478)
	Employment and unemployment (1015)
	Employment insurance, social assistance and other transfers (67)
	Hours of work and work arrangements (236)
	Job training and apprenticeship programs (60)
	Job vacancies, labour mobility and layoffs (60)
	Commuting to work (126)
	Globalization and the labour market (23)
	Unionization and industrial relations (39)
	Unpaid work (78)
	Work transitions and life stages (2)
	Workplace organization, innovation and performance (15)
	Workplace health and work absences (21)
	Other content related to Labour (74)
노르웨이	Labour market and earnings (1024)
	Earnings and labour costs (346)
	Employment (522)
	Unemployment (248)
	Working environment, sickness absence, strikes and lockouts (165)

출처: Statistics Canada, Data; Statistics Norway, Statbank Norway [웹사이트].

3) 자기 참조 관계

제1장의 10페이지 <표 3>에서 살펴본 GSIM 다이어그램은 보통 정보 클래스 간의 관계를 나타내지만, 정보 클래스 자기 자신을 참조하는 관계 설정도 나타난다. 예컨대, Universe는 채워진 마름모로 표기되는 (0..*) 차수의 parent-child 관계를 가질 수 있다. 이러한 자기 참조 관계를 활용하여 상·하위 관계, 연관 관계, 부분·전체 등 관련되는 개념들을 연계하여 설명할 수 있다.

<표 44> 자기 참조 관계를 갖는 정보 클래스

정보 클래스	관계	표기
Unit Type	is an aggregation of	◇— (포함)
Universe	parent-child	◆— (구성요소)
Population	parent-child	◆— (구성요소)
Node	parent-child / part-whole	◇— (전체-부분)
Node Set	relates	→ (연관)
Statistical Classification	is variant of	— (대응)
Process Step	has	→ (전체-부분)
Question	references	→ (전체-부분)
Referential Metadata Attribute	parent-child	◆— (구성요소)

출처: UNECE, GSIM Version 2.0 - 그룹 다이어그램 [웹사이트] (2026.1. 조회).

3. 개념의 역할 간 연계

제2장에서 개념(concept)은 크게 1) 특성, 2) 대상, 3) 범주 3가지 역할 수행한다고 정리했었다. 이때, 개념의 역할은 그 개념이 무엇인지에 따라 고정되는 것이 아니라, 그 개념이 어디에서 사용되는지에 따라 달라지는 것이다. 예컨대, ‘취업’이라는 개념은 1) 개인의 취업여부를 측정하는 Conceptual Variable이 될 수도 있고, 2) 취업자의 산업, 종사상지위 등을 측정할 관심대상이 되는 Universe가 될 수도 있으며, 3) 경제활동상태를 외연적으로 정의하는 Category가 될 수도 있다. 실제로 캐나다의 경우 ‘Labour Force Status’는 표준 분류로, ‘Labour force status of person’은 표준 변수로 지정하고 있다. 따라서, 앞 절에서 제안한 변수, 코드목록 등 개념 저장소를 설계할 때 이들 개념의 역할에 대한 범용 활용성에 고려가 필요하다.

제 5 장

결론 및 시사점

제1절 연구 결과

본 연구는 국가통계 개념 정보의 체계적 관리 필요성을 확인하고, 국제표준 통계정보 모델인 GSIM을 기반으로 국내 통계환경에 적용할 수 있는 방안을 검토하였다. 개념 그룹을 중심으로 주요 정보 클래스의 역할과 클래스 간 관계를 정리하였으며, 설계와 구현의 분리, 정보의 재사용성, 개념의 계층적 구체화 등 GSIM의 핵심 설계 원칙을 다각도로 분석하였다.

캐나다와 노르웨이의 사례를 비교한 결과, 양국 모두 GSIM을 활용하여 개념, 통계 단위, 변수, 통계분류 및 코드목록 등을 체계적으로 관리하고 있으나, 캐나다는 기관 차원의 표준 정책과 중앙 메타데이터 저장소를 기반으로 표준의 제정과 재사용을 강조하는 반면, 노르웨이는 변수, 코드목록 등을 기능별 시스템에서 관리하면서 최종 통계표 조회까지 연계하는 방식이 특징적이었다. 또한 Conceptual Variable에 대응하는 변수 정보가 추상적인 개념 수준에만 머무르는 것은 아니며, 판정기준이나 도출방법 등 구체적인 내용까지 포함하고 있음을 확인하였다. 다만 캐나다는 정의(Definition)와 도출방법(Derivation)을 구분하는 반면 노르웨이는 구체적인 판정조건을 정의문 자체에 포함하는 차이를 보였다. ‘혼인상태’ 사례에서는 보편적 상위 개념이라도 국가별 제도적·통계적 맥락에 따라 변수 및 범주 구성이 달라질 수 있음을 확인하였다.

이러한 검토 결과를 바탕으로 이용자 활용성과 통계작성기관의 관리 부담 등 국내 환경을 종합적으로 고려하여 관리가 필요한 핵심 정보 클래스를 선별하고, 해외사례의 장점을 결합하여 국내에 적용하기 위한 국가통계 개념 정보 관리 체계 구축 방향을 도출하였다.

제2절 국가통계 개념 정보 관리체계 구축 방향

국내 국가통계 개념 정보의 체계적인 관리를 위해서는 우선 개념 정보에 대한 표준 관리 원칙과 거버넌스를 정립할 필요가 있다. 표준으로 관리할 개념 정보의 범위와 작성 기준을 정립하고, 표준의 제정·개정·승인·폐기 및 모니터링 절차와 관리 주체를 명확히 해야 하며, 특히 공통 개념과 표준값을 관리하는 ‘표준 관리부서’와 개별 통계에 포함된 개념 정보를 관리하는 ‘통계 프로그램 담당자’ 간의 역할을 명확히 구분할 필요가 있다.

정보 클래스별 측면에서는 관리 실익과 재사용성을 고려하여 Unit Type, Code List 및 Variable을 주요 표준 관리대상으로 선정하고, Universe는 작성 지침 중심의 검토를 제안하였다. Unit Type은 현행 조사 수준에서 변수 수준으로 확장하여 국가통계 전반을 포괄할 수 있는 표준값 집합을 마련할 필요가 있다. 반면 Universe는 통계작성 맥락에 따라 달라지므로 공통 표준값을 직접 설정하기보다 마이크로데이터 관리부서와의 협의를 통해 작성원칙과 지침을 제공하는 방식이 적절할 것으로 판단된다.

Statistical Classification은 통계법에 기반하여 체계적으로 제정·관리되고 있으므로 현행 방식을 유지하는 것이 바람직하다. 이와 별도로, 동일 개념에 대한 다양한 범주 구성을 조화시키고 통계 프로그램별 변형을 관리하기 위해 Code List를 표준 관리대상으로 설정하였다. 표준 코드목록과 통계 프로그램별 코드목록 간 대응관계를 관리함으로써 서로 다른 범주 구성과 명칭을 사용하는 통계 간 개념적 차이와 공통점을 파악할 수 있는 기반을 마련할 필요가 있다. 아울러 작성 양식에 버전, 변경이력, 대응표 및 변형관계를 포함함으로써 코드목록의 지속적인 변화와 재사용을 체계적으로 지원할 것을 제안한다.

Variable은 표준 변수와 통계 프로그램별 변수로 이원화하여 관리하는 방안을 제안하였다. 표준 변수는 공통적으로 활용할 수 있는 개념과 정의를 관리하고, 통계 프로그램별 변수는 개별 통계의 작성 목적과 맥락에 따른 구체적인 정의와 용법을 관리하도록 하였다. 한편, Conceptual Domain은 개념변수가 하나의 개념 도메인만을 참조해야 한다는 GSIM의 제약으로 인해 표준으로서의 재사용성을 제한할 소지가 있으므로 별도의 관리대상에서 제외하였다. 대신 최상위 정보 클래스인 Identifiable Artefact의 Description 속성을 활용하여 변수의 조작적 정의를 명시하도록 하였다.

제3절 향후 과제 및 기대효과

본 연구는 캐나다와 노르웨이 등 일부 국가의 사례를 검토하고 노동 분야를 중심으로 시범 적용하였다는 점에서 한계가 있다. 향후 다양한 통계 분야로 실증 범위를 확대하고, 통계작성기관과의 긴밀한 협의를 통해 관리체계의 적용 가능성과 실효성을 다각도로 검증할 필요가 있다. 이와 함께 통계작성기관 및 이용자 간의 원활한 소통을 위해 GSIM 정보 클래스에 대응하는 명확한 국문 통계용어를 정립하고, 통계작성기관 담당자가 이해할 수 있도록 정보 클래스별 상세 작성 기준을 명확화하는 작업도 병행되어야 한다. 아울러 변수·코드목록 등에 대한 정보가 국가통계포털과 연계되어 하나의 개념 정보를 여러 시스템에서 일관되게 조회·활용할 수 있는 구조를 마련하는 것도 중요한 과제가 될 것이다.

표준 변수, 표준 코드목록 등 새로운 표준 개념 관리체계의 안정적인 정착률을 위해서는, 통계 프로그램별 변수 및 코드목록 정보를 우선 구조화하여 확보한 후 이를 비교·검토하여 공통성과 활용성이 높은 정보부터 단계적으로 표준을 지정·확대해 나가는 접근을 고려할 수 있다.⁴¹⁾ 이는 캐나다 통계청이 표준의 적용 의무 수준을 차등화하여 운영하는 방식과도 맥락을 같이한다.

이처럼 GSIM 기반의 국가통계 개념 정보 관리체계가 성공적으로 정착하여 개별 통계에 분산된 개념 정보를 공통 구조로 통합 관리하고 표준 개념과 연계하게 된다면, 통계 간 비교가능성과 재사용성을 획기적으로 향상시킬 수 있을 것이다. 나아가 통계·데이터의 연계·융합을 촉진하고 국가통계 데이터의 활용 범위를 대폭 확장하는 핵심 기반이 되기를 기대한다.

41) 국가데이터처 작성 통계 → 중앙부처 작성 통계 → 지역통계 순의 단계적 접근 고려 가능

참고문헌

- 국가데이터처(2025), 「경제활동인구조사 조사표(2025년)」.
- 국가데이터처(2026), 「2026년 경제활동인구조사 지침서 외부용(2026년)」.
- 국가데이터처(2026), 「경제활동인구조사 연간자료 공공용 마이크로데이터 파일설계서(2025년)」.
- 국가데이터처(2026), 「2026 UNECE 표준지원 워크숍 참가 결과보고서」, 내부자료.
- Cloutier, M. and Born, A.(2010), “Metadata at Statistics Canada: implementation opportunities and challenges”, IFC Bulletin, No. 33, pp. 65-73.
- DDI Alliance(2025), DDI-Cross Domain Integration Specification Overview.
- ISO/IEC(2023), ISO/IEC 11179-1:2023, Information technology — Metadata registries (MDR) — Part 1: Framework, Fourth edition.
- Object Management Group(2017), OMG Unified Modeling Language, Version 2.5.1.
- OECD(2008), OECD Glossary of Statistical Terms.
- SDMX(2020), SDMX Guidelines — SDMX Glossary, Version 2.1.
- Statistics Canada(2015), National Household Survey, 2011 [Canada] Public Use Microdata File (PUMF): Individuals File.
- Statistics Canada(2026), Labour Force Survey (LFS) Public Use Microdata File (PUMF) Codebook.
- U.S. Census Bureau and U.S. Bureau of Labor Statistics(2026), Frequently Used Variables: American Time Use Survey (ATUS).
- U.S. Census Bureau(2025), Monthly Current Population Survey (CPS) Public Use Microdata File; Data Dictionary.
- U.S. Census Bureau, Monthly Current Population Survey (CPS) Public Use Microdata File; Documentation.
- UN(2025), Handbook on Management and Organization of National Statistical Systems.
- UNECE HLG-MOS Supporting Standards Group(2021), ModernStats Models Governance Guidance.
- UNECE HLG-MOS Supporting Standards Group(2026), Business Case for Sentinel Values.
- UNECE HLG-MOS(2013), Report on Activities and Results for 2012.
- UNECE HLG-MOS(2026), Report on Activities and Results for 2025.
- UNECE(2012), Generic Statistical Information Model (GSIM): Specification, Version 1.0, December 2012.
- UNECE(2012), “Metadata Standards and Models — Timeline”, background documentation for the GSIM Sprint 1 .
- UNECE(2013), Generic Statistical Information Model (GSIM): Specification, Version 1.1, December 2013.
- UNECE(2016), CES Members’ Guidebook.
- UNECE(2019), Generic Statistical Information Model (GSIM): Glossary, Version 1.2, August 2019.

- UNECE(2019), GSIM Concept Group, presented by Catrin Karling (Statistics Sweden) and Mikko Saloila (Statistics Finland), GSIM e-training, 5 November 2019.
- UNECE(2022), *Linking GSBPM and GSIM*, Version 1.0, January 2022.
- UNECE(2024), *Generic Statistical Information Model (GSIM): Communication Paper for a General Statistical Audience*, Version 2.0, June 2024.
- UNECE(2024), *Generic Statistical Information Model (GSIM): Version 2.0 User Guide*.
- UNECE Statswiki(2009), *Statistical Unit Definitions in the Metadata Common Vocabulary(2009 version)*, by Steven Vale

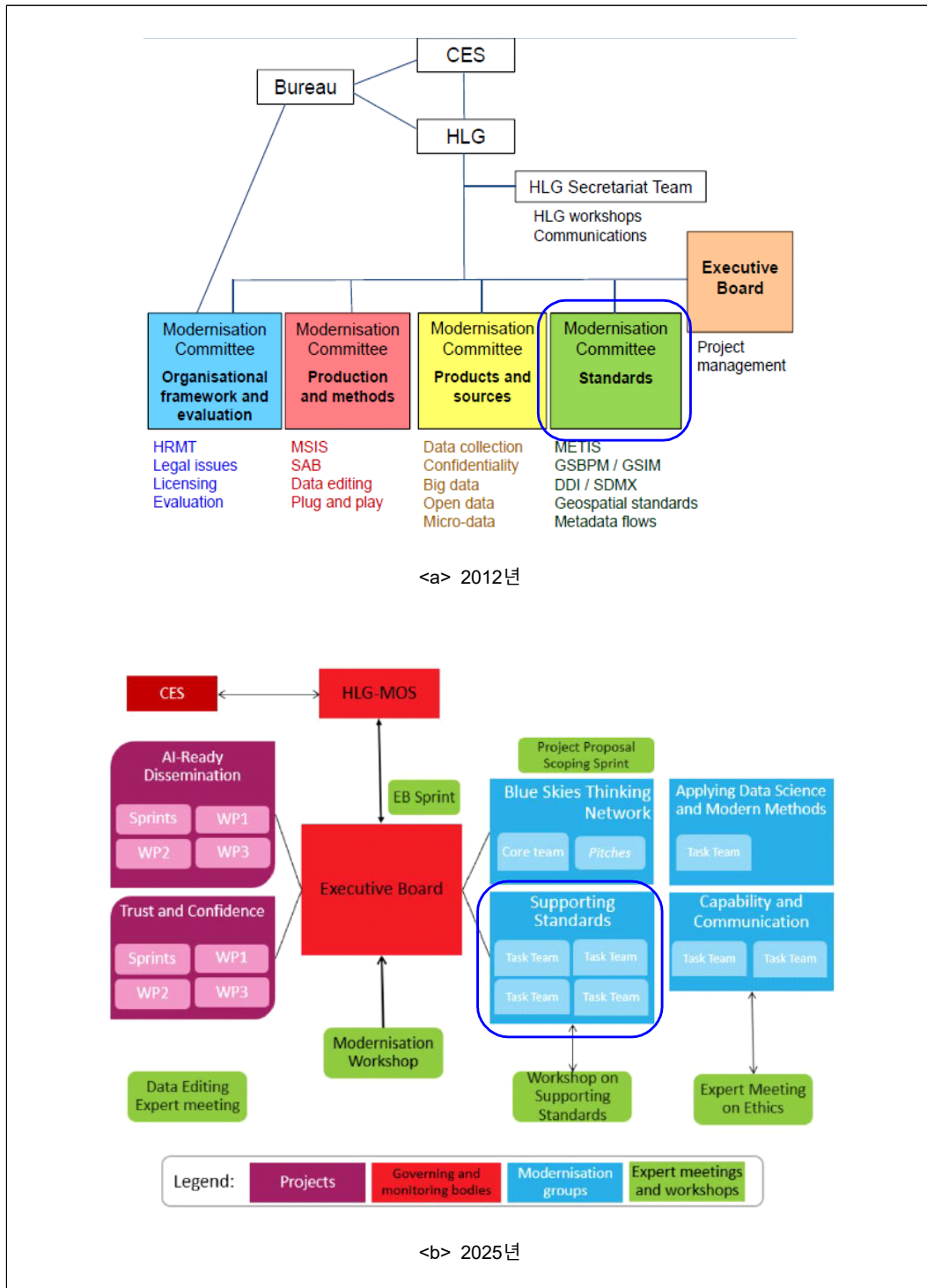
< 웹사이트 >

- 국가데이터처, 국가통계포털, <https://kosis.kr>.
- 국가데이터처, 통계설명자료서비스, <https://www.k-stat.go.kr/metascv>.
- 국가데이터처, 통계설명자료서비스 용어조회, <https://www.k-stat.go.kr/metascv/msda100/worddc>.
- 국가데이터처, 통계용어사전, <https://mods.go.kr/statisticalTermList.es?mid=a10506000000>.
- 국가데이터처, 통계정책시스템, <https://www.narastat.kr/pms/index.do>.
- Statistics Canada, Classifications, variables and statistical units, <https://www.statcan.gc.ca/en/concepts/search>.
- Statistics Canada, Data, <https://www150.statcan.gc.ca/n1/en/type/data>
- Statistics Canada, Policy on standards (revised July 14, 2004), <https://www.statcan.gc.ca/en/about/transparency/policy/standards>.
- Statistics Canada, Surveys and statistical programs, <https://www23.statcan.gc.ca/imdb-bmdi/pub/indexa-eng.htm>.
- Statistics Norway, About SSBs variable system (Published: 3 December 2021), <https://www.ssb.no/en/metadata/om-ssbs-variabelsystem>.
- Statistics Norway, Metadata, <https://www.ssb.no/en/metadata>.
- Statistics Norway, Metadata — Dapla-manualen (2025.1.20.), <https://manual.dapla.ssb.no/statistikkere/metadata.html>.
- Statistics Norway, Statbank Norway, <https://www.ssb.no/en/statbank>.
- Statistics Norway, How to use StatBank Norway, <https://www.ssb.no/en/statbank/hvordan-bruke-statistikkbanken>.
- UNECE, GSIM Version v2.0 (December 2023), <https://unece.github.io/GSIM-2.0>.
- UNECE Statistics, GSIM 2.0 Webinar 2025, YouTube (@UNECEStats), 2025.9.16., https://www.youtube.com/playlist?list=PL6uBaBVebZDQXgqM-G__clN-MAVhXQ08.
- UNECE Statswiki, Generic Statistical Information Model, <https://statswiki.unece.org/spaces/gsim/pages/59703371/Generic+Statistical+Information+Model>.

부 록

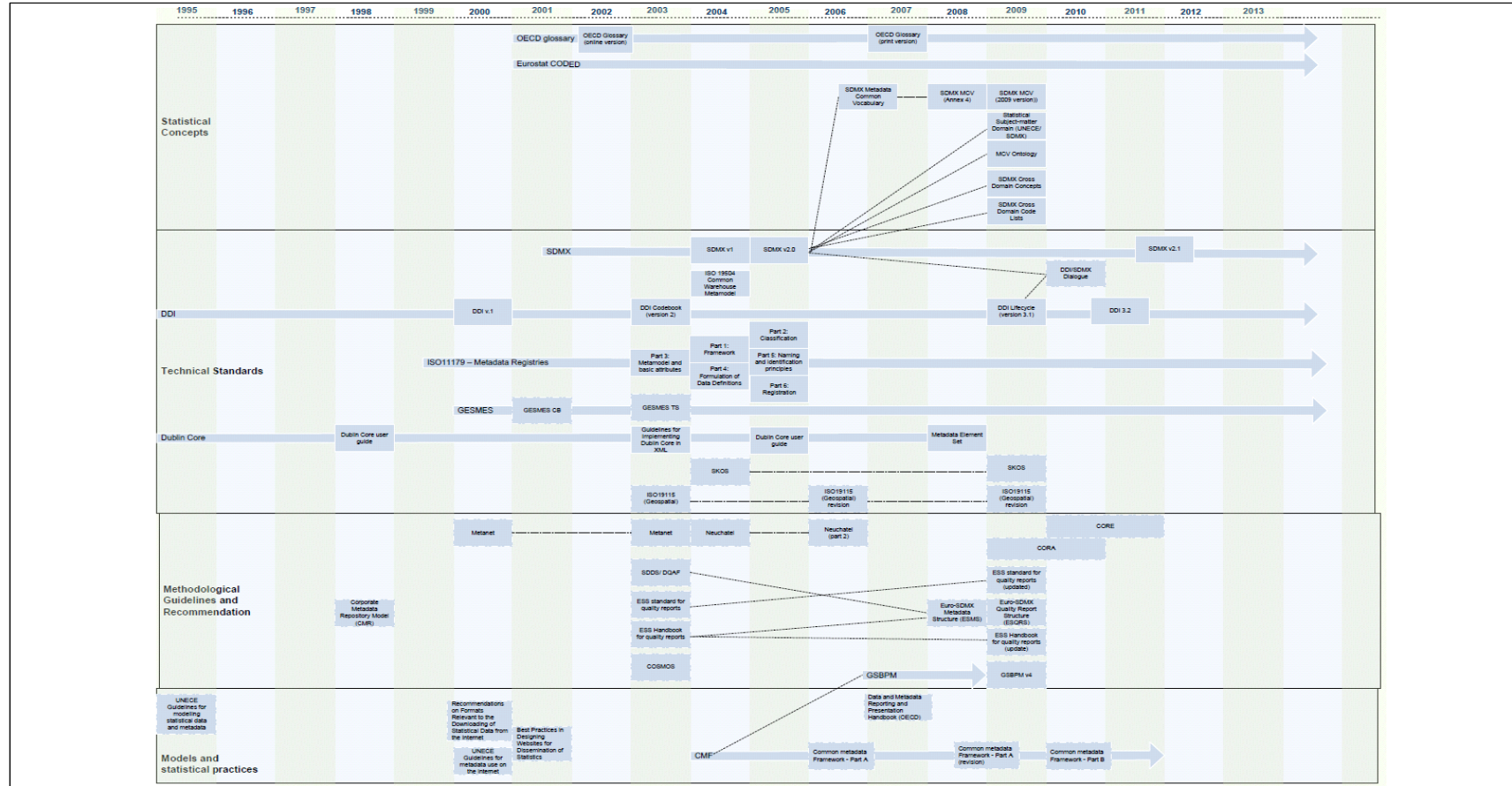
1. HLG-MOS 조직 체계 및 표준지원그룹	92
2. 메타데이터 표준/모델 타임라인	93
3. 개념 조화 시스템 구현 발표 사례(포르투갈 은행)	94
4. 조건식으로 변수 Universe 작성 사례(미국 CPS)	95
5. 다이어그램 표기 예시 및 해석	96
6. 비즈니스 그룹 정보 클래스(재배치)	97
7. GSIM Information Flow in GSBPM	
7-1. Sub-process 2.2 변수 설명 설계	98
7-2. Sub-process 5.5 새로운 변수·단위 도출	99
8. 캐나다 'Policy on standards' 전문	100
9. 노르웨이 통계청 변수 문서화 시스템 입력 항목	102
10. 노르웨이 통계청 StatBank의 Variable 필터 기능	103
11. 미국 빈출 변수 사례(CPS 공공용 MD)	104
12. 미국 빈출 변수 사례(ATUS 공공용 MD)	105
13. 미국 업무 프로세스 단계를 반영한 변수명 부여 사례(CPS)	106
14. GSIM 주요 정보 클래스별 정의 및 설명자료	107

<부록 1> HLG-MOS 조직 체계 및 표준지원그룹



출처: UNECE HLG-MOS(2013), Report on Activities and Results for 2012; UNECE HLG-MOS(2026), Report on Activities and Results for 2025.

<부록 2> 메타데이터 표준/모델 타임라인



출처: UNECE(2012), Metadata Standards and Models – Timeline, Background Documentation for the GSIM Sprint 1.

<부록 3> 개념 조화 시스템 구현 발표 사례(포르투갈 은행)

- (발표주제) 개념 조화(Concept Harmonisation)의 역할: 표준화를 통한 통계시스템 발전
- (개요) 공통 통계데이터 웨어하우스 구축·개선을 목표로, 단일 데이터 사전(single dictionary)을 통해 데이터의 조화 및 표준화 달성('22년 시작되어 5개월간 집중 투입)
- (주요 작업 내용) ❶ 400개 이상의 변수와 7,000개 이상의 코드를 통합해 조화된 코드 목록을 생성하는 데이터 사전 구축, ❷ 개별 개념을 표준 개념으로 연결하는 개념 매핑, ❸ 표준화된 코드 기반의 집계 테이블 생성 등을 수행, Python 기반 프로세스와 자연어 처리 기술을 활용, ❹ 개념 유사성 식별 및 조정 작업 효율화
- (준수 원칙) 영숫자 코드 사용 및 SDMX 코드 우선 적용, 명확·유의미한 명칭·코드 사용, 각 변수의 근본적 의미에 대한 문서화, 코드 길이 제한 등 원칙을 적용하여 데이터 관리 일관성 확보

IMPLEMENTING A HARMONISED SET OF CONCEPTS

TASK FORCE
5 EXPERTS FROM
DIFFERENT DOMAINS

Gather all statistical outputs
All variables (codelists), and their concepts (codes), that characterised data

Harmonisation cycle

1. Group all variables and codes that had the same underlying concept (e.g. CCE, Instrument, IVF ITEM, ...)
2. Define a robust governance
3. Study all concepts and check for consolidation/harmonisation opportunities
4. Proof of concept with use cases

Deliverables

- Harmonised codelists
- Mapping tables, from legacy to harmonised concepts
- Aggregation tables based on harmonised concepts

THE ROLE OF CONCEPT HARMONISATION | ADVANCING STATISTICAL SYSTEMS THROUGH STANDARDISATION

IMPLEMENTING A HARMONISED SET OF CONCEPTS

Stepwise approach to harmonise concepts across all statistical outputs

WITHIN COMPILATION PROCESSES

When the compilation system is updated, it should adopt the harmonised set of concepts

WHEN STORING STATISTICAL OUTPUTS IN THE DATAWAREHOUSE

42 statistical outputs translated into the new set of concepts, since 2023

WORK IN PROGRESS

All statistical outputs translated into the new set of concepts by the end of 2026

THE ROLE OF CONCEPT HARMONISATION | ADVANCING STATISTICAL SYSTEMS THROUGH STANDARDISATION
19

출처: 국가데이터처(2026), UNECE 표준지원 워크숍 참가 결과보고서, 내부자료.

<부록 4> 조건식으로 변수 Universe 작성 사례 (미국 CPS 공공용 MD)

Data dictionaries for monthly CPS public use files

Along with the monthly CPS public use files, data dictionaries can be downloaded from the Census FTP DataWeb page (<https://www.census.gov/data/datasets/time-series/demo/cps/cps-basic.html>). These will provide extensive information about each variable included in the data files.

Below is a sample entry from the monthly CPS public use data dictionary:

Name of variable

PEAFEVER

Size of the variable - the number of digits the variable uses

2

Description of variable or question wording used to collect data

DID YOU EVER SERVE ON ACTIVE DUTY IN THE U.S. ARMED FORCES?

Identifies the position on the file where the variable is located

131 - 132

EDITED UNIVERSE: PRTAG ≥ 17

VALID ENTRIES

1 YES

2 NO

The valid entries may be either a list of valid values or a maximum and a minimum value

Identifies when the variable is defined

<a> Data Dictionary

Edited Universe	변수(개)
ALL HHLD's IN SAMPLE	12
PRPERTYP = 1, 2, OR 3	12
PRPERTYP = 2	11
ALL HHLDs IN SAMPLE	8
PRIOELG = 1	6
ALL	5
PRERELG = 1	5
PRIOELG = 1 AND PEMJOT = 1 AND HRMIS = 4 OR 8	5
PEAFEVER = 1	4
PEMLR = 1-7	4
PRPERTYP = 1-3	2
PRPERTYP = 2 OR 3	2
ALL HOUSEHOLD MEMBERS	1
⋮	

 Edited Universe 집계 (저자 작성)

출처: U.S. Census Bureau(2025), CPS PUMF Data Dictionary.

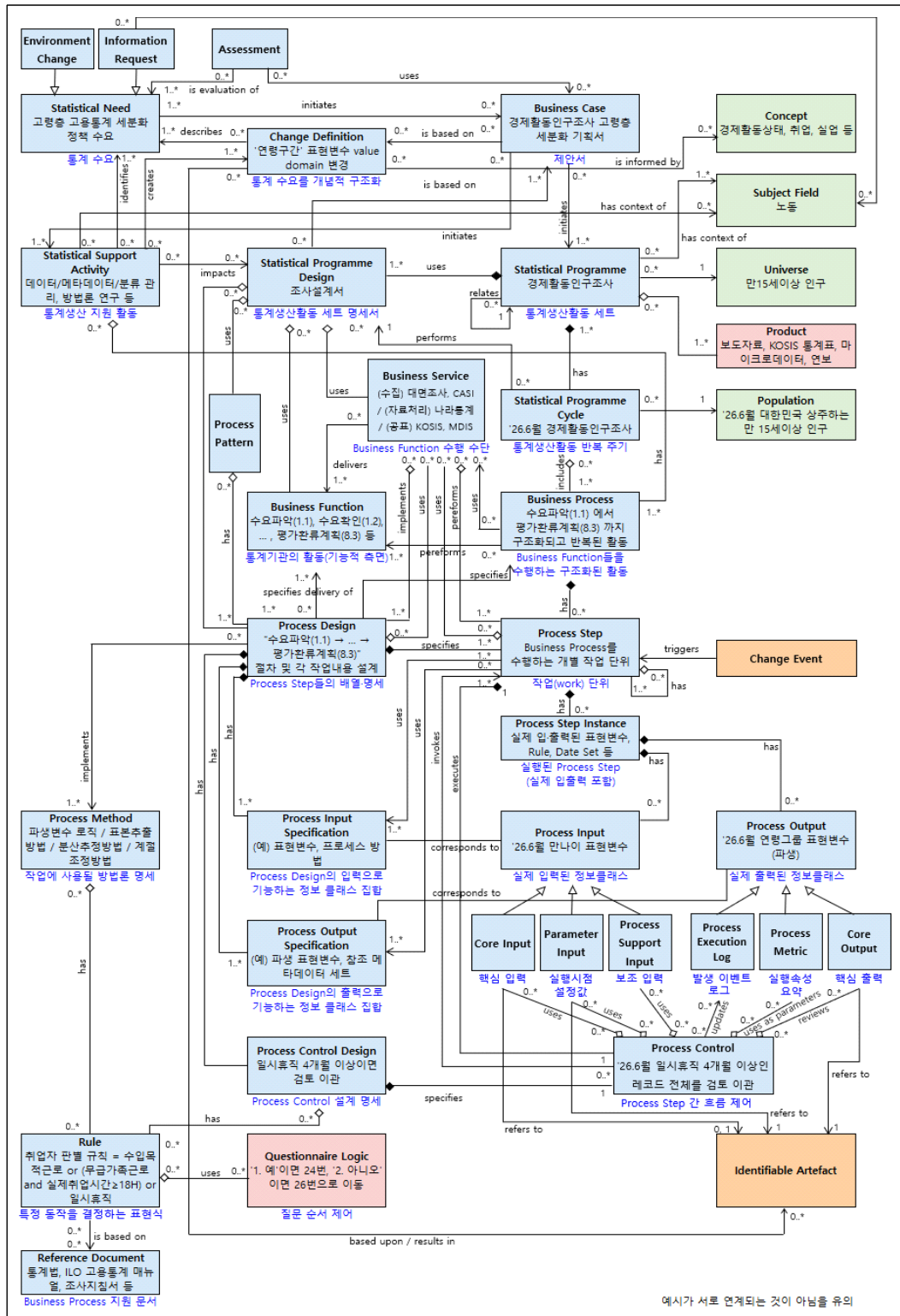
* 주소: <https://www.census.gov/data/datasets/2025/demo/cps/cps-basic-2025.html>
 경로: Census.gov / Data / Datasets / 2025 Basic Monthly CPS / Data Dictionary
 원파일명: 2025_Basic_CPS_Public_Use_Record_Layout_plus_IO_Code_list.txt

<부록 5> 다이어그램 표기 예시 및 해석

표기	예시 및 해석
 열린 화살표	개념변수 (0..*) → uses 개념 도메인 (1) ○ 개념변수는 개념 도메인을 이용함 ○ 개념변수는 하나의 개념 도메인만을 이용함 - 따라서 개념 도메인이 달라지면 다른 개념변수가 됨 - 즉, {남자, 여자}인 개념 도메인을 이용하는 개념변수와 {남자, 여자, 성소수자}인 개념 도메인을 이용하는 개념변수는 서로 다름 ○ 동일한 개념 도메인을 이용하는 개념변수는 여러 개일 수 있음 - {남자, 여자}인 개념 도메인을 이용하더라도, 단위유형에 따라 가구원 성별, 가구주 성별, 기업체대표 성별 등의 다양한 개념변수가 정의될 수 있음 ○ 개념 도메인을 이용하는 개념변수가 없을 수 있음 - {남자, 여자, 성소수자}라는 개념 도메인을 정의해놓더라도, 이를 측정하는 통계가 없다면 개념변수는 0개임
 실선	개념 도메인 (1) — corresponds to 값 도메인 (0..*) ○ 개념 도메인은 값 도메인에 대응됨 - 개념 도메인이 가지는 유효한 개념에 대응하여, 값 도메인이 정의됨 - 예를 들어, 개념 도메인에 유효한 개념이 3개가 있다면, 값 도메인도 이에 대응하는 3개의 값을 가짐 ○ 개념 도메인은 여러 개의 값 도메인에 대응될 수 있음 - {남성, 여성}이라는 개념 도메인은 {1, 2} 혹은 {F, M} 등 여러 개의 값 도메인에 대응될 수 있음 ○ 개념 도메인이 대응되는 값 도메인이 없을 수 있음
 빈삼각형	표현변수 → 개념변수 ○ 표현변수는 개념변수의 구체화임 - 즉, 개념변수의 속성을 그대로 상속받으면서, 값 표현방법 정보가 추가됨
 빈 마름모	모집단 (0..*) ◆ contains 단위 (1..*) ○ 모집단은 단위를 포함함, 즉, 모집단은 단위들의 집합으로 이루어짐 ○ 모집단은 최소 1개 이상의 단위를 포함해야 함 - 즉, 데이터 수집여부(응답률)와는 별개로 대상 집단의 실재해야 함을 의미함 ○ 단위는 어느 모집단에도 포함되지 않을 수 있음 - 예를 들어, 15일 주간에 국외 체류 중이었던 사람은 22일 주간에 국내에 있더라도 모집단(국내 상주인구)에서 제외됨
 채워진 마름모	연계표 ◆ contains 맵 (1..*) ○ 연계표는 하나 이상의 맵을 반드시 포함해야 함 - 예를 들어, 9차 산업분류에서 10차 산업분류와의 연계관계(맵)가 없는 연계표는 존재할 수 없음

42) (p.112) ‘composite’ indicates that the Property is aggregated compositely, i.e., the composite object has responsibility for the existence and storage of the composed objects. If a composite object is deleted, all of its part instances that are objects are deleted with it. (p.202) An Association with aggregation = AggregationKind::composite likewise has a diamond at the corresponding end, but differs in having the diamond filled in.

<부록 6> 비즈니스 그룹 정보 클래스(재배치)



출처: UNECE, GSIM Version 2.0 - 그룹 다이어그램 [웹사이트] (2026.1. 조회) 재구성.

<부록 7-1> GSIM Information Flow in GSBPM Sub-process 2.2 (변수 설명 설계)

Process Input Specification	
Core Input type <i>Business Cases</i> containing proposed solutions and approved decisions regarding <i>Concept to be measured</i> Parameter Input type Process Support Input type Standards, policies or guidelines to be followed on how to design <i>Variables</i> to be collected and write metadata description Repository or inventory of • <i>Variable, Unit Type</i> (e.g., persons, households, enterprises), <i>Measurement Type</i> (e.g., currencies, capita) • <i>Value Domain, Measurement Unit</i> • <i>Statistical Classification, Code List</i> that are agreed and approved in the <i>Organisation</i> to be used as reference <i>Represented Variables, Process Methods, Statistical Classification or any other referential metadata available</i> from previous <i>Statistical Program Cycles</i> or similar <i>Statistical Programs</i> that can be used as reference	Concept 측정을 위해, 각종 저장소 및 유사 프로그램의 표현변수 등 참고
Process Step Design	
Process Method Review <i>Business Case</i> regarding <i>Concepts</i> and <i>Variables</i> Consult repository or inventory of <i>Variables, Statistical Classifications</i>, etc., relevant standards, guidelines and any available information from previous <i>Statistical Program Cycles</i> or similar <i>Statistical Programs</i> Design <i>Represented Variables (Variables, Value Domains)</i> Design <i>Statistical Classifications*</i> if needed Design <i>Process Methods</i> to be used for derivation and calculation of <i>Represented Variables</i> Prepare metadata descriptions of <i>Represented Variables, Process Methods</i>, etc. * Version local to the current business process, the maintenance of <i>Statistical Classification</i> that is used across the <i>Organization</i> takes place as a separate business process.	
Process Output Specification	
Core Output type <i>Statistical Program Designs</i> that containing • <i>Represented Variables (Variables, Value Domains)</i> to be used for collection and to be derived • <i>Process Methods</i> specifying how to derive and calculate the <i>Represented Variables</i> • <i>Statistical Classifications</i>: description of the statistical classification (geospatial) structure and content. • <i>Referential Metadata Set</i> describing any relevant information about <i>Process Methods, Statistical Classification</i>, etc. Process Metric type Quality measures regarding designing such as: • Percentage of <i>Represented Variables</i> and classifications that have metadata descriptions • Percentage of definitions and classifications associated to (key) <i>Represented Variables</i> and <i>Populations</i> that follow international and national standards Quality measures of the <i>Process Step</i> such as: • Time spent to complete the <i>Process Step</i> (derived from <i>Process Execution Log</i>) • Cost spent to complete the <i>Process Step</i> Process Execution Log type Execution logs such as • Time that <i>Process Step</i> started • Time that <i>Process Step</i> ended	

출처: UNECE(2022), Linking GSBPM and GSIM.

<부록 7-2> GSIM Information Flow in GSBPM Sub-process 5.5 (새로운 변수·단위 도출)

Process Input Specification	
Core Input type	
<i>Data Sets (Unit Data Sets)</i> : edited and imputed	
<i>Data Structures</i> specifying the structure of the <i>Data Sets</i>	
<i>Represented Variables</i>	
<i>Units</i>	
<i>Process Methods</i> specifying methodologies (e.g., aggregation, splitting, suppression, rounding) and associated <i>Rules</i> to be used to derive new <i>Represented Variables</i> and <i>Units</i> such as:	
• Arithmetic formulae • Derivation order	Process Methods 및 Rules 적용
<i>Statistical Classification</i> or <i>Code Lists</i> that are used for derivation	
Parameter Input type	
Process Support Input type	
Technical / methodological handbooks, policies or guidelines to be followed regarding derivation as well as quality management	
Process Step Design	
Process Method	
Review <i>Data Sets</i> and <i>Process Methods</i>	
Apply <i>Process Methods</i> and <i>Rules</i> to derive new <i>Represented Variables</i> and <i>Units</i>	
Calculate quality measures specified by <i>Process Methods</i>	
Update <i>Data Sets</i> and associated element in <i>Data Structure</i>	
Process Output Specification	
Core Output type	
<i>Data Sets</i> with derived <i>Represented Variables</i> and <i>Units</i>	
<i>Data Structure</i> specifying the structure of <i>Data Set</i>	
<i>Referential Metadata Set</i> : descriptions of the <i>Process Methods</i> used, quality information summarising <i>Process Metrics</i> or any other relevant information to be passed along with <i>Data Sets</i>	
<i>Derived Units</i>	
<i>Derived Represented Variables</i>	
Process Metric type	
Quality measures such as:	
• Counts of derived <i>Data Points</i>	
Quality measures about <i>Process Step</i> such as	
• Time spent to complete the <i>Process Step</i> (derived from <i>Process Execution Log</i>) • Cost spent to complete the <i>Process Step</i>	
Process Execution Log type	
Execution log such as	
• Time that <i>Process Step</i> started • Time that <i>Process Step</i> ended	
Any message or event generated from software used for calculation	

출처: UNECE(2022), Linking GSBPM and GSIM.

<부록 8> 캐나다 통계청 “Policy on standards”(표준에 관한 정책) 내용 전문

소개

정책은 캐나다 통계청 프로그램에서 사용되는 인구, 통계 단위, 개념, 변수 및 분류에 대한 표준 명칭과 정의의 사용을 검토, 문서화, 승인 및 모니터링하기 위한 프레임워크를 제공합니다. 특정 주제 영역에 대한 표준은 필요에 따라 이 정책에 따라 수시로 발표될 것입니다.

정책

캐나다 통계청은 통계 프로그램에서 사용되는 모집단, 통계 단위, 개념, 변수 및 분류에 대해 일관된 명칭과 정의를 사용하는 것을 목표로 합니다. 이를 위해 다음과 같은 조치를 취합니다.

1. 통계 자료에는 모집단, 통계 단위, 개념, 변수 및 사용된 분류 체계의 정의에 대한 쉽게 접근 가능한 문서가 함께 제공되거나 명시적으로 참조될 것입니다.
2. 관련 통계 단위, 개념, 변수 또는 분류 간에 명칭이나 정의에 불일치 또는 모호성이 발견되는 경우, 기관은 이러한 차이를 조화시키는 통계 단위, 개념, 변수 및 분류에 대한 표준 개발을 위해 노력할 것입니다.
3. 특정 주제 영역을 다루는 표준 및 지침은 수시로 발행될 것이며, 해당 표준 및 지침의 사용은 본 정책의 조항에 따라 규정됩니다.
4. 부서별 표준이 제정된 경우, 해당 프로그램 영역은 이 정책 조항에 따라 특별한 면제를 받지 않는 한 해당 표준을 준수해야 합니다.
5. 프로그램은 가능한 한 각 표준 분류의 가장 기본적인 수준 또는 가장 세부적인 수준의 정보를 수집하고 보관해야 합니다. 이는 집계에 있어 최대한의 유연성을 제공하고 필요에 따라 소급하여 재분류할 수 있도록 하기 위함입니다.
6. 프로그램이 부서 표준에 포함되지 않은 모집단, 통계 단위, 개념, 변수 또는 분류를 사용하거나 면제로 승인된 표준의 변형을 사용하는 경우, 이전에 정의된 표준과 구별하기 위해 해당 개체에 고유한 이름을 사용해야 합니다.
7. 캐나다 통계청의 자문 서비스를 이용하는 고객은 이 정책에 따라 발행된 기준 및 지침을 숙지하고 준수하도록 권장되어야 합니다.
8. 해당 기관은 프로그램에서 사용되는 명칭과 정의에 대한 데이터베이스를 구축하고, 이 데이터베이스를 통계 시스템의 사용자 및 기타 관계자들이 접근할 수 있도록 할 것입니다.

범위

이 정책은 수집, 도출 또는 구성 방식, 배포 매체 또는 자금 출처와 관계없이 배포되는 모든 데이터에 적용됩니다. 또한 이 정책은 캐나다 통계청에서 데이터를 수집 및 처리하는 단계에 있는 데이터에도 적용될 수 있습니다.

표준 개발 및 문서화를 위한 지침

A. 서론

본 지침은 모집단, 통계 단위, 개념, 변수 및 분류의 표준 명칭과 정의를 개발하고 문서화하는 데 필요한 요건과 지침을 제공합니다. 섹션 B에서는 용어를 정의하고, 섹션 C에서는 구체적인 지침을 제시합니다.

B. 용어

본 지침의 목적상 다음과 같은 용어가 사용됩니다.

모집단: 데이터세트가 참조하는 통계 단위들의 집합입니다.

개념: 측정하고자 하는 사회적 및/또는 경제적 현상을 나타내는 일반적이거나 추상적인 아이디어.

통계 단위: 데이터를 수집하거나 도출하는 데 사용되는 관찰 또는 측정 단위. 다음 목록은 정의된 표준 통계 단위의 예시를 제공합니다.

개인 인구조사 가족경제 가족가구주거지사업체 회사기업

변수: 변수는 통계 단위와 속성, 두 가지 구성 요소로 이루어져 있습니다. 속성은 통계 단위의 특징 또는 특성입니다.

분류: 분류는 변수가 가질 수 있는 값들을 체계적으로 그룹화하는 것으로, 상호 배타적인 클래스들로 구성되며, 모든 값들을 포괄하고, 종종 데이터를 집계하기 위한 계층적 구조를 제공합니다. 주어진 변수에 대해 하나 이상의 분류를 사용하여 데이터를 표현할 수 있습니다.

예시: 다음은 ‘사람의 나이’라는 변수의 예시입니다.

개념: 캐나다 통계청이 통계 자료 및 메타데이터를 정리하는 데 사용하는 주제를 기반으로, ‘인구 연령’ 변수는 ‘인구 및 인구 통계’ 개념 아래에 나열됩니다.

통계 단위 및 속성: 이 변수를 정의하는 통계 단위와 속성은 각각 ‘개인’과 ‘연령’입니다. ‘개인’은 개개인을 의미하며, 이는 대부분의 사회 통계 프로그램에서 사용하는 분석 단위입니다. ‘연령’은 관심 대상인 개인(또는 연구 대상)의 최근 생일 기준 나이(또는 명확하게 정의된 기준일 기준 나이)를 나타냅니다.

분류: 이 변수에 대한 데이터를 나타내는 데에는 다양한 분류법이 사용될 수 있습니다. 이러한 분류법에는 5년 단위연령 그룹 및 생애주기별 연령 그룹이 포함됩니다.

인구, 통계 단위, 개념, 변수 및 분류의 표준 명칭과 정의는 통합 메타데이터베이스(IMDB)에 저장됩니다. 변수의 경우, IMDB에 저장되는 변수 이름에는 통계 단위 및 속성 외에도 표현 유형이 포함됩니다. 예를 들어, 연령을 예로 들면, IMDB에 저장되는 변수의 전체 이름은 ‘Category of Age of Person’이 됩니다. 표현 유형이 ‘Category’라는 것은 해당 변수가 범주형 변수이며, 연령 그룹 분류로 표현된다는 것을 의미합니다.

C. 지침

각 표준은 다음과 같은 특징을 가져야 합니다.

- 해당 표준이 적절한 경우 다루는 개념을 설명하십시오.
- 적용되는 통계 단위를 식별하십시오.
- 표준에 포함된 각 변수의 이름과 정의를 제공하십시오.
- 각 변수에 대한 데이터 편찬(compilation) 및 배포에 사용될 분류 체계를 제공하십시오.

표준에는 항상 가장 세부적인 분류 수준이 포함됩니다. 권장 및 선택적 집계 구조도 포함될 수 있습니다.

가능한 경우 개념은 프레임워크와 관련하여 설명되어야 합니다.

모든 변수에는 두 공식 언어로 이름이 부여되어야 하며, 일단 부여된 이름은 다른 변수를 나타내는 데 사용할 수 없습니다. 변수는 속성과 해당 속성이 적용되는 통계 단위를 명시하는 설명과 함께 정의되어야 합니다. 또한, IMDB에서는 표현 유형도 정의되어야 합니다.

모든 분류에는 두 공식 언어로 명칭을 부여해야 하며, 일단 부여된 명칭은 다른 분류를 지칭하는 데 사용할 수 없습니다. 분류는 정의되어야 하며, 필요한 경우 제외 사항 목록과 설명 주석을 포함해야 합니다.

모든 분류에는 두 가지 공식 언어로 이름이 부여되어야 하며, 일단 부여된 이름은 특정 분류군(즉, 특정 분류와 그 모든 변형) 내에서 참조되는 변수에 대한 다른 그룹을 나타내는 데 사용할 수 없습니다. 분류는 정의되어야 하며, 필요한 경우 제외 사항 목록과 설명 주석이 제공되어야 합니다.

가장 빈번하게 사용되는 인구 집단에는 두 공식 언어로 명칭을 부여해야 하며, 일단 부여된 명칭은 다른 인구 집단을 지칭하는 데 사용할 수 없습니다. 이러한 인구 집단은 설명 주석과 함께 정의되어야 합니다.

모든 통계 단위는 두 공식 언어로 명칭을 부여받아야 하며, 일단 부여된 명칭은 다른 통계 단위를 지칭하는 데 사용할 수 없습니다. 통계 단위는 설명 주석과 함께 정의되어야 합니다.

표준에는 관련 국제적으로 인정된 표준과의 적합성 진술서 또는 해당 표준과의 차이점에 대한 설명과, 가능한 경우 참조된 표준과의 일치성에 대한 설명이 함께 제공되어야 합니다.

새로운 표준이 이전 표준을 대체하는 경우, 이전 표준과 새로운 표준 간의 일치표를 제시해야 합니다.

표준에는 해당 표준의 적용 의무 정도에 대한 명시가 포함되어야 합니다. 의무 정도는 정도가 높은 순서대로 다음과 같습니다.

부서 표준: 정책 위원회의 승인을 받은 표준으로서, 본 정책의 조항에 따라 명시적으로 면제를 받지 않는 한 적용이 의무적입니다.

권장 표준: 방법 및 표준 위원회에서 권장 표준으로 인정한 표준으로, 특정 기간의 시험 기간을 거치거나 거치지 않고 해당 기간이 경과한 후 부서 표준으로 선포될 수 있습니다.

프로그램별 표준: 통계 프로그램에서 채택하고 표준 부서에 등록하여 시간 경과에 따른 시계열의 일관성을 보장하는 표준입니다.

수정일: 2025년 10월 31일

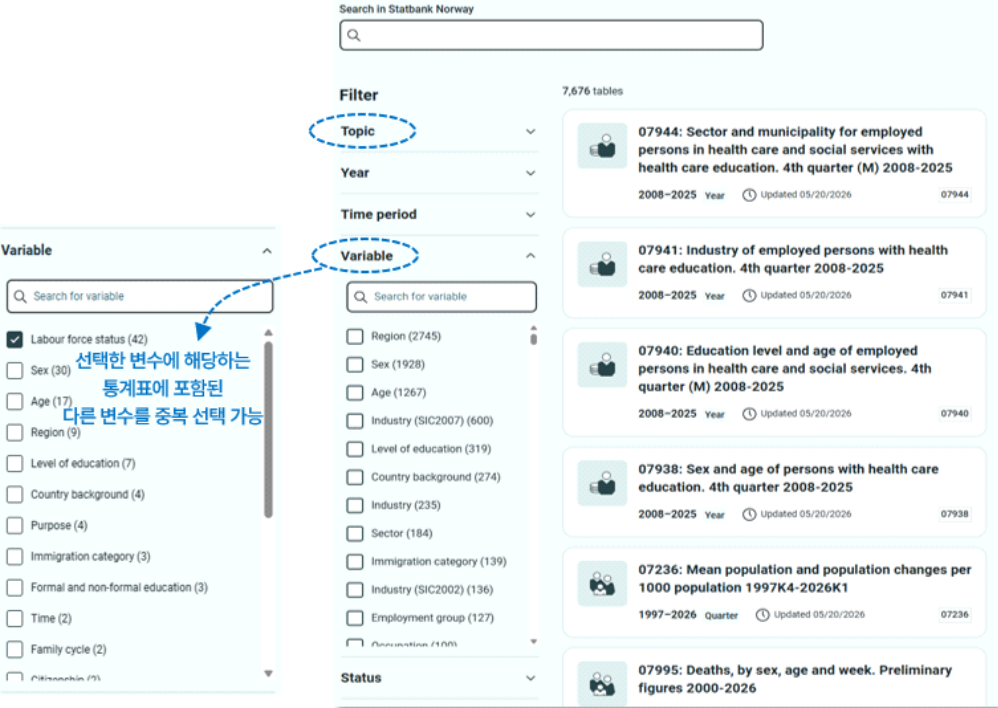
출처: Statistics Canada, Policy on standards (revised 2004) [웹사이트].

<부록 9> 노르웨이 통계청 변수 문서화 시스템 입력 항목

항목	필수여부	비고
1. Primary language	필수	노르웨이어(보크몰), 노르웨이어(뉘노르스크) 중 택일
2. Available languages	필수	(노르웨이어(보크몰), 노르웨이어(뉘노르스크), 영어 중 복수 선택
3. Identifier	자동생성	
4. Name	필수	자유 텍스트
5. Short name	-	자유 텍스트
6. Definition	필수	자유 텍스트
7. Owner division	자동생성	로그인 정보에서 자동 생성
8. Confidentiality	필수	값 목록에서 택일
9. Contact person	필수	기본값 = 항목 생성자
10. Valid from	-	
11. Valid to	-	
12. Statistical unit type	필수	값 목록에서 택일
13. Internal publication reference	-	내부 검색기능 가능여부
14. External publication reference	-	외부 검색기능 가능여부
15. Calculation	-	자유 텍스트
16. Comment for SSB users	-	자유 텍스트
17. Comment for external users	-	자유 텍스트
18. Classification?	-	기본값 = No
19. Classification	-	위에서 '예'를 선택시 분류DB에서 분류, 분류 버전 또는 분류 수준을 찾아 선택 가능
20. Subject area	필수	주제 분야 - 3레벨 계층 구조 값 목록에서 택일
21. Statistic	-	주제 분야에 따라 값 목록에서 택일
22. SSB source	-	SSB 출처 - 값 목록에서 택일
23. External source	-	외부 출처 - 값 목록에서 택일
24. Definition approved for Internet	체크박스	인터넷 공개 승인 정의
25. Definition approved for entire SSB	체크박스	전체 SSB 공개 승인 정의
26. Linked to other definitions in variable documentation system	체크박스	변수 문서 시스템의 다른 정의와 연결됨
27. Linked to file descriptions in our archive	체크박스	아카이브의 파일 설명에 연결됨
28. Linked to tables in our dissemination database	체크박스	배포 데이터베이스의 테이블에 연결됨
29. Used in our event history metadatabase	체크박스	이벤트 기록 메타데이터베이스에서 사용됨
30. Created	시스템생성	생성일
31. Created by	시스템생성	생성자
32. Last changed	시스템생성	최종 수정일
33. Last changed by	시스템생성	최종 수정자
34. Copied from identifier	-	다른 변수로부터 복사된 경우 자동으로 채워짐

출처: Statistics Norway, About SSBs variable system [웹사이트].

<부록 10> 노르웨이 통계청 StatBank의 Variable 필터 기능



The screenshot shows the StatBank Norway interface. On the left, there is a 'Variable' filter section with a search bar and a list of variables. A blue arrow points from the 'Variable' filter section to the search results. The search results are displayed in a list format, showing the variable name, the time period, and the number of matches. The 'Variable' filter section is highlighted with a blue circle, and the 'Variable' filter is highlighted with a blue circle. The search results are sorted by the number of matches, and variables with no matches will disappear as you make new filter selections.

The **Variable filter** includes classification variables, but not contents variables (i.e., what is measured or counted). Selecting multiple variables will return tables that include all selected variables. This helps narrow down the results quickly. The list is sorted by number of matches, and variables with no matches will disappear as you make new filter selections.

Above the result list, you'll see the filters and search terms you've used. You can easily remove one or more selections to adjust your search. These filter choices are also reflected in the URL.

출처: Statistics Norway, How to use StatBank Norway [웹사이트].

<부록 11> 미국 빈출 변수 사례(CPS 공공용 MD)

Frequently-used variables		
There are a number of variables that are commonly used by researchers. These define critical demographic concepts, such as race, sex, and age, as well as important details of a person's labor force status. Researchers also often rely on geographic variables that identify household location. The table below includes some of the more commonly used variables.		
Characteristic	Variable Name	Notes
Demographic Variables		
Type of person	PRPERTYP	Identifies whether the person is a child, adult, or serving in the U.S. Armed Forces
Age	PRTAGE	Called PEAGE before January 2015
Sex	PESEX	
Race	PTDTRACE	
Hispanic ethnicity	PEHSPNON	
Detailed Hispanic ethnicity	PRDTHSP	
Educational attainment	PEEDUCA	
Certification and licensing status	PECERT1	
Disability status	PRDISFLG	
Veteran status	PEAFEVER	
Marital status	PEMARITL	
Foreign-born status	PENATVTY	
Labor Market Variables		
Labor force status	PEMLR	Defines employed, unemployed, and not in the labor force
Full- or part-time work status	PRWKSTAT	Identifies full- and part-time status as defined by actual hours of work. Also identifies usual and actual full- and part-time status by economic and noneconomic reasons. (Can also be used to identify employed, unemployed, and not in the labor force.)
Weekly earnings	PRERNSA	Calculated for all workers. Collected for one-quarter of the sample.
Hourly earnings	PRERNHLY	Only calculated for hourly-paid workers. Collected for one-quarter of the sample.
Multiple job status	PRSMJ	
Usual weekly hours worked	PEHRUSLT PEHRUSL1 PEHRUSL2	PEHRUSLT describes all jobs combined, PEHRUSL1 describes the main job, and PEHRUSL2 describes the other job(s).
Actual hours worked during the survey reference week	PEHRACTT PEHRACT1 PEHRACT2	PEHRACTT describes all jobs combined, PEHRACT1 describes the main job, and PEHRACT2 describes the other job(s). The survey reference week is generally the week containing the 12 th of the month.
Occupation (main job)	PEIO1OCD PRDTOCC1 PRMJOC1 PRMJOCGR	These variables have varying levels of detail. PEIO1OCD has the most detailed occupation breakout, PRDTOCC1 has 22 categories, PRMJOC1 has 10 categories, and PRMJOCGR has 6 categories. Occupation of second job is available in PEIO2OCD.
Class of worker (main job)	PEIO1COW PRCOW1 PRDTCOW1	Class of worker for second job is available in PEIO2COW, PRCOW2, and PRDTCOW2.
Industry (main job)	PEIO1ICD PRDTIND1 PRIMIND1 PRMJIND1	These variables have varying levels of detail. PEIO1ICD has the most detailed industry breakout, PRDTIND1 has 51 categories, PRDTIND1 has 21 categories, PRIMIND1 has 21 categories, and PRMJIND1 has 13 categories. Industry of second job is available in PEIO2ICD.
Geographic Variables		
State	GESTFIPS	
Metropolitan Statistical Area	GTCBSA	
County	GTCO	

출처: U.S. Census Bureau, CPS PUMF Documentation.

<부록 12> 미국 빈출 변수 사례(ATUS 공공용 MD)

Frequently Used ATUS Variables			
The ATUS files have many variables and users may sometimes have difficulty determining which variables to use. This document attempts to identify those that are most frequently used. Unless noted otherwise, these variables appear on both the single-year and multi-year ATUS data files. All variables are defined in the ATUS Data Dictionaries, available at: www.bls.gov/tus/dictionaries.htm .			
Characteristic	Variables	Location	Notes
Demographic			
Age	TEAGE	Activity Summary file Roster file	
Sex	TESEX	Activity Summary file Roster file	
Race	PTDTRACE	Activity Summary file ATUS-CPS file	Not collected in ATUS but available from CPS interviews (conducted 2-5 months prior to ATUS interview).
Hispanic origin	PEHSPNON	Activity Summary file ATUS-CPS file	Not collected in ATUS but available from CPS interviews (conducted 2-5 months prior to ATUS interview). PEHSPNON identifies whether or not an individual is Hispanic.
Detailed Hispanic origin	PRDTHSP	ATUS-CPS file	Not collected in ATUS but available from CPS interviews (conducted 2-5 months prior to ATUS interview). PRDTHSP identifies the Hispanic origin of individuals (such as Mexican or Cuban).
Relationship to respondent	TERRP	Roster file	
Presence of household children < 18	TRHHCHILD	Respondent file	
Number of children < 18 living in the household	TRCHILDNUM	Activity Summary file Respondent file	
Presence of own nonhousehold children < 18	TRNHCHILD	Respondent file	
Presence of own household children < 18	TROHHCHILD	Respondent file	
Age of youngest household child < 18	TRYHHCHILD	Activity Summary file Respondent file	
Job and income variables			
Labor force status	TELFS	Activity Summary file Respondent file	
Weekly earnings	TRERNWA	Activity Summary file Respondent file	This is the most commonly used earnings variable.
Hourly earnings	TRERNHLY	Respondent file	
Job paid hourly	TEERNHRY	Respondent file	
Multiple job status	TEMJOT	Activity Summary file Respondent file	
Full-time or part-time status	TRDPFTPT	Activity summary file Respondent file	
Usual weekly hours of work (total)	TEHRUSLT	Activity Summary file Respondent file	TEHRUSLT identifies usual weekly hours worked at all jobs.
Usual weekly hours of work (main and other jobs)	TEHRUSL1 TEHRUSL2	Respondent file	TEHRUSL1 identifies usual weekly hours worked at main job and TEHRUSL2 identifies usual weekly hours worked at other job(s).
Occupation (main job)	TEIO1OCD TRDTOCC1 TRMJOCOC1 TRMJOCGR	Respondent file	These variables have varying levels of detail. TEIO1OCD has the most detailed occupation breakout, TRDTOCC1 has 22 categories, TRMJOCOC1 has 10 categories, and TRMJOCGR has 6 categories. The ATUS occupation classification system was updated in January 2020. See the ATUS changes documentation for more information about changes to the occupation codes over the years: www.bls.gov/tus/other-documentation/changes.pdf . Information on occupation of other job(s) is not collected in ATUS. For this information, analysts may use PEIO2OCD on the ATUS-CPS file. Occupation data in the CPS was also updated in 2020. This variable was collected 2 to 5 months prior to the ATUS interview and could be out of date.
Class of worker (main job)	TEIO1COW	Respondent file	Identifies whether respondent is self-employed, employed by government, a private firm, etc. Information on class of worker for other job(s) is not collected in ATUS. For this information, analysts may use PEIO2COW on the ATUS-CPS file. This variable was collected 2 to 5 months prior to the ATUS interview and could be out of date.
Industry of employment (main job)	TEIO1ICD TRDTIND1 TRIMIND1 TRMJIND1	Respondent file	These variables have varying levels of detail. TEIO1ICD has the most detailed industry breakout, TRDTIND1 has 51 categories, TRIMIND1 has 21 categories, and TRMJIND1 has 13 categories. The ATUS industry classification system changed in January 2020. See the ATUS changes documentation for more information about changes to the industry codes over the years: www.bls.gov/tus/other-documentation/changes.pdf . Information on industry of other job(s) is not collected in ATUS. For this information, analysts may use PEIO2ICD on the ATUS-CPS file. Industry data in the CPS was also updated in 2020. This variable was collected 2 to 5 months prior to the ATUS interview and could be out of date.
Family income	HUFAMINC HEFAMINC	ATUS-CPS file	Not collected in ATUS but available from CPS interviews (conducted 2-5 months prior to ATUS interview and possibly out of date at the time of the ATUS interview). HEFAMINC is an edited version of HUFAMINC that contains some allocated or imputed values. At the time HEFAMINC was added to the ATUS-CPS file, the variable HUFAMINC was dropped. See the ATUS changes documentation for more information: www.bls.gov/tus/other-documentation/changes.pdf . Caution should be used with these variables. Income is a sensitive question and there is a high level of nonresponse. HEFAMINC has an allocation rate of up to 20 percent.

출처: U.S. Census Bureau and U.S. Bureau of Labor Statistics(2026), ATUS Frequently Used Variables.

<부록 13> 미국 업무 프로세스 단계를 반영한 변수명 부여 사례(CPS)

< 변수명 부여 규칙 >

- 첫 번째 문자는 변수 수준, 두 번째 문자는 변수 유형, 이후 문자열은 변수 설명
- 변수 수준: 개인(P), 가구(H), 주소(G)
- 변수 유형: 원자료(U, Unedited), 내검자료(E, Edited), 재코딩(R, Recode),
상한처리(T, Topcoding), 가중치(W, Weight), 할당유형*(X, Allocation flag)
- * 할당값은 원자료가 결측, 모름, 불응 및 모순값일 때 부여

첫 번째 숫자(할당방식) \ 두 번째 숫자(초기값)		0	1	2	3
		value	blank	don't know	refused
공백	no change	0	1	2	3
1	changed to a value	10	11	12	13
2	changed to a longitudinal value	20	21	22	23
3	changed to an allocated longitudinal value	30	31	32	33
4	changed to an allocated value	40	41	42	43
5	changed to a blank	50	-	52	53

< 변수명 사례 >

NAME	SIZE	DESCRIPTION	LOCATION	비고
PEMARITL	2	Marital status	125 - 126	개인 / 내검자료
PXMARITL	2	Allocation flag	661 - 662	개인 / 할당유형
HUFINAL	3	Final outcome code	24 - 26	가구 / 원자료
HRHTYPE	2	Household type	61 - 62	가구 / 재코딩
PTHR	1	Hourly pay - Top code	524 - 524	개인 / 상한처리
PWSSWGT	10	Final weight	613 - 622	개인 / 가중치

< 변수값 예시(가상) >

- PUSEX = blank
- PESEX = 1 (MALE)
- PXSEX = 41 (blank to allocated value)

출처: U.S. Census Bureau, CPS PUMF Documentation.

<부록 14> GSIM 주요 정보 클래스별 정의 및 설명자료 (개념 그룹과 그 외 일부)

Concept	○ 특성(characteristic)으로 구별되는 사고의 단위 - '정의'를 통해 표현 ○ 최상위 수준의 공통성을 표현하는 것 외에도, 다양한 형태로 사용됨 - 개념의 하위유형: 측정할 특성(characteristic), 범주(category), 대상(단위유형, 모집단, 유니버스, ...)
용어집_정의	○ unit of thought differentiated by characteristics
용어집_속성	○ Definition(1..*): 관련 개념들과 구별되는 서술적 진술을 통해 개념을 표현하는 것
이용자안내	58. '개념'은 통계 수명주기 전반에 걸쳐 다양한 방식(개념의 하위유형)으로 사용됨 - 측정할 특성(characteristic) → 예) 성별(gender), 생산 가치 - 단위 유형(unit type), 모집단(population), 유니버스(universe), 사건(events), 엔티티(entities) 등 정보의 대상 → 예) 네덜란드 성인 모집단, 또는 단위 유형 혹은 사람에 기반한 대졸 성인 전체 - 개념의 세부 정의를 위한 범주 → 예) 성별이라는 개념에 대한 남성 및 여성
웨비나	○ Used to define a variable (e.g. marital status, age) * Variable: Mapping between a set of units (the sample) and a Value Domain ○ Concepts are characteristics of the unit classes - People have a marital status - People have an age ○ All marital status variables share the same concept ○ All age variables share the same concept
DDI-CDI	○ 어떤 주제에 관한 변수들은 모두 그 주제를 공통으로 공유합니다. 예를 들어, 연구 자료실의 혼인 상태에 관한 데이터세트에서 사용되는 모든 변수들은 동일한 개념을 공유합니다. 각 변수에서 측정된 혼인 상태 자체는 공통점이 거의 없을 수 있지만, 혼인 상태 그 자체, 즉 사회나 문화 전반에 걸쳐 혼인 상태가 존재한다는 사실은 공통적인 특징입니다. 이러한 공통성을 표현하는 개념이 바로 최상위 수준의 목적입니다. 이 단계의 개념은 매우 일반적입니다. 왜냐하면 이 개념을 사용하는 각 변수에 연결된 보다 세분화된 버전의 모든 가능한 변형을 고려해야 하기 때문입니다. (예) Marital status (정의: Category of current marital arrangement) (예) Age (정의: Whole number of years of operation) (예) Velocity (정의: Change of position per unit of time) ○ 개념이란 간단히 말해 아이디어에 대한 형식적인 정의이며, 특징에 따라 구별되는 사고의 단위입니다. 일반적으로 이는 공식적으로 정의된 용어(최소한 개념의 이름 및 정의 속성)로 표현됩니다. ○ 개념은 다양한 데이터를 탐색하고 비교하는 데 매우 중요한데, 이러한 탐색과 비교를 자동화된 방식으로 수행하기 위해서는 개념이 수행하는 다양한 역할을 이해하는 것이 중요합니다. 개념은 변수 정의와 관련될 뿐만 아니라, 위에서 설명한 것처럼 유니버스와 유닛 타입도 정의합니다. 또한, 개념은 값 도메인을 표현할 때 범주로 자주 사용됩니다. ○ '성별'이라는 개념변수가 있고, 이 변수가 '남성', '여성' 등 여러 범주 중 하나로 표현되는 데이터세트를 생각해 보겠습니다. 이 데이터세트는 변환 과정을 거치면 '남성'이라는 동일한 개념을 범주로 사용하는 '남성 여부'라는 변수가 있는 데이터세트와 비교 가능할 수 있습니다. 이 변수는 '예' 또는 '아니오'라는 이진 값으로 표현됩니다. 동일한 개념 '남성'이 두 데이터세트에서 동일한 역할을 한다는 사실은 두 데이터세트가 비교 가능하거나 연구 질문에 답하는 데 유용할 수 있음을 시사합니다. DDI CDI 모델은 이러한 상황을 기계가 감지하고 처리할 수 있는 공식을 제공합니다. ○ 개념을 범주로 사용하는 것(The use of Concepts as Categories)은 중요합니다. 왜냐하면 이를 통해 개념이 모집단 내 특정 단위와 어떻게 연관되는지 알 수 있기 때문입니다. 변수의 표현 방식이 다르더라도 동일한 개념이 개념값으로 존재할 수

	있습니다. 위에서 제시된 혼인 상태 예시가 그 예입니다. 동일한 범주 집합, 즉 동일한 개념이 서로 다른 코드를 사용하여 표현될 수 있습니다. 이러한 수준의 세부 정보는 어떤 것이 비교 가능한지 아닌지를 파악하기 위해 데이터를 처리할 때 다시 한번 중요한 역할을 합니다. ○ 범주를 사용하면 개념이 DDI-CDI모델의 두 가지 중요한 구조와 연결된다는 점에 주목할 필요가 있습니다. 데이터를 설명할 때, 모델은 변수로 설명되는 데이터에서 개념적 값으로 자주 사용되는 코드의 의미를 제공합니다. 또한 분류 항목(분류의 노드)에 대한 정의도 제공합니다. 범주는 분류 또는 코드목록의 맥락에서 사람 사용자뿐 아니라 기계에게도 코드의 의미를 알려주는 데 추가적인 가치를 지닐 수 있습니다. ○ 개념은 DDI DI 모델에서 여러 다른 역할에서도 사용되지만, 앞서 언급한 용도는 변수 연쇄 작용 방식을 이해하는 데 중요합니다.
ISO 11179-1	○ 3.2.8 concept : 고유한 특성들의 조합으로 생성된 지식 단위 - 개념은 반드시 특정 자연어에 국한되는 것은 아니나, 사회적·문화적 배경의 영향을 받으며 이는 종종 다른 범주화로 이어짐 - 개념은 표현방식과 독립적임 ○ 객체 클래스(object class)는 '일반 개념(General Concept)'일 수 있다. 이는 해당 객체 클래스에 대응하는 객체의 전체 집합이 둘 이상의 구성원을 가질 때 발생하며, 마이크로데이터(Record-level data)가 이런 방식으로 기술된다. - 반면, 객체 클래스는 '개별 개념(Individual Concept)'일 수도 있다. 이는 객체 전체 집합이 단 하나의 구성원만 가질 때 발생한다. 예를 들어 '모든 사람의 집합'이나 '서비스업 사업체들의 집합'과 같이 단일 객체에 대응하는 개념이 이에 해당한다. 집계 데이터(Aggregate data)는 이런 방식으로 기술되며, 이러한 객체 클래스의 속성 예시로는 각각 '평균 소득'이나 '총 수익' 등이 있다.
OECD 용어집	○ A concept is a unit of knowledge created by a unique combination of characteristics [ISO 1087-1:2000, 3.2.1] 개념은 고유한 특성들의 조합으로 생성된 지식의 단위임 ○ Context 개념은 공통점을 가진 것으로 여겨지는 일련의 행동, 태도 또는 특성에 대한 추상적인 요약, 일반적인 개념, 지식 등을 의미함. 개념은 정확한 의미를 제시/전달하고, 현상을 분류/해석/구조화해하는데 도움을 주기 위해 사용됨. 둘 이상의 개념 간 의미적 연결을 개념 관계라고 함. 둘 이상의 개념 간의 관계 유형에 대한 설명을 개념관계 유형 설명이라고 함 - Concepts are abstract summaries, general notions, knowledge, etc., of a whole set of behaviours, attitudes or characteristics which are seen as having something in common. - Concepts are used to assist in presenting/ conveying precise meaning, categorising, interpreting, structuring and making sense of phenomena (such as classifications) ('United Nations Glossary of Classification Terms' prepared by the Expert Group on International Economic and Social Classifications, unpublished on paper, available at: http://unstats.un.org/unsd/class/family/glossary_short.htm). - A semantic link among two or more concepts is a concept relationship. - A description of the type of relationship among two or more concepts is a concept relationship type description. (ISO/IEC International Standard 11179-3 'Information technology- Metadata registries (MDR)-Part 3: Registry metamodel and basic attributes', February 2003). ○ Source: ISO/IEC FDIS 11179-1 'Information technology - Metadata registries - Part 1: Framework', March 2004 ○ See also: Subject field; Data element concept; Structural definition; Characteristic; Key family; Dimension; Metadata structure definition; Definition; ISO / IEC 11179; Ontology; Structural definition; Statistical concept; Compiling agency French equivalent: Concept

SDMX 용어집(2.1)	○ Definition - Unit of thought created by a unique combination of characteristics. ○ Context - At an abstract level, a Concept is defined in the Generic Statistical Information Model (GSIM) as a 'unit of thought differentiated by characteristics'. Concepts are used in different ways throughout the statistical lifecycle, and each role of a Concept is described using different information objects (which are subtypes of Concept). A Concept can be used in these situations: (a) As a characteristic. The Concept is used by a Variable to describe the particular characteristic that is to be measured about a Population. For example, to measure the Concept of gender in a population of adults in Spain, the Variable combines this Concept with the Unit Type 'person'. (b) As a Unit Type or a Population. To describe the set of objects that information is to be obtained about in a statistical survey. For example, the Population of adults in the Netherlands based on the Unit Type of persons. ◎ Category to further define details about a Concept. For example, Male and Female for the Concept of Gender. Codes can be linked to a Category via a Node (i.e. a Code Item or Classification Item), for use within a Codelist or Statistical Classification. - In SDMX the Concept can be given a core Representation such as a reference to a Codelist for an enumerated Representation or other values such as 'integer' or 'string' for a non-enumerated Representation. This Representation can be overridden in the data structure when the Concept is used as a Dimension or Attribute. A Concept with a core Representation could be regarded as a represented variable. ○ Source - United Nations Economic Commission for Europe (UNECE), Generic Statistical Information Model (GSIM) Specification (Version 1.1, December 2013) (https://statswiki.unece.org/display/gsim/)
------------------	---

Conceptual Variable	○ 개념(Concept)의 의미와 단위유형(Unit Type)의 결합 - 측정하고자하는 특성을 정의하기 위한 역할 ○ 개념을 어떤 특성으로 측정할지에 따라, 즉 개념 도메인 구성에 따라 개념변수는 달리 정의됨 - (예) {미혼/유배우/사별/이혼}으로 측정하는 개념변수 ≠ {미혼/기혼}으로 측정하는 개념변수
용어집_정의	○ use of a Concept as a characteristic of Unit Type intended to be observed
용어집_설명	○ 개념변수는 개념의 의미와 단위 유형을 결합하여 측정하고자 하는 특성을 정의합니다. (combines the meaning of a Concept with a Unit Type, to define the characteristic that is to be measured) - (예) Sex of person, Number of employees, Value of production
이용자안내	61. 특성의 역할을 수행하는 개념과 단위유형의 결합 (The association of Unit Type and a Concept playing the role of a characteristic) - (예) 유니버스가 성인인 경우, 단위유형인 '개인'과 '교육수준' 개념을 결합한 개념변수를 생각해 볼 수 있음 62. 개념변수는 결과값 표현에 대한 정보를 포함하지 않음 (표현변수에서 포함) - 이렇게 개념변수와 표현변수를 구분함으로써, 표현방식이 다르더라도 측정대상이 개념적으로 동일한 경우 개념변수의 정의를 재사용할 수 있음
GSIM 1.1 명세	○ GSIM 1.1에서는 클래스명이 'Conceptual' 수식어 없이 'Variable'이었음
GSIM 1.2 용어집	○ Variable 정의: The use of a Concept as a characteristic of a Population intended to be measured.

	Variable 설명: The Variable combines the meaning of a Concept with a Unit Type, to define the characteristic that is to be measured.
웹비나	○ Concept level + the Conceptual Domain ○ 개념 도메인이 달라지면 개념변수도 달라짐 (When the CD changes, the Conceptual Variable changes) ○ With marital status, consider - CD4: <single> <married> <divorced> <widowed> - CD2: <single> <married> → CD4 ≠ CD2 ((CD4의 싱글은 결혼경험이 없는 사람, CD2의 싱글은 현재 배우자가 없는사람)) ○ 개념 도메인이 다르면 개념변수를 다르게 특징지음(These 2 CDs characterize marital status differently) - 혼인상태라는 개념은 여전이 같은 의미임 (Even though marital status still means the same thing) - 단, 혼인상태가 가지는 개념 도메인이 다르게 해석됨 (But marital status CVs are interpreted slightly differently) ○ Age - CD = {0 ≤ whole number} - We don't know right now if the number represents years · The number could represent days, for instance ○ Remember - - numbers(수) are concepts - numerals(숫자) are representations ○ 개념 도메인은 데이터의 기본 개념을 열거함 (CDs list the concepts underlying the data) ○ Variable: Mapping between a set of units (the sample) and a Value Domain - 수학에서의 함수와 유사하지만 완전히 같지는 않음 (Similar to but not quite the same as a function in mathematics) - Domain = sample; Range = Value Domain ○ Variable Cascade - Means of describing variables - Contains 4 levels: Concept / Conceptual Variable / Represented Variable / Instance Variable - 이러한 레벨들을 '변수'로 혼동하면 안됨 (Note that these levels are not to be confused with variables) - 데이터세트의 변수를 설명하기 위한 부품임 (Part of the description of variables found in data sets) - 각 계층은 변수 설명을 위한 재사용 가능하고 분리가능한 부분들임 (Levels represent reusable and separable parts of descriptions)
DDI-CDI	○ 개념변수는 변수를 설명하는 데 사용되는 대부분의 개념이 적용되는 수준입니다. 주요 개념은 특정 유형의 단위에 대한 개념의 특수한 적용이며, 측정 또는 설명의 특성을 나타냅니다. - 측정 대상 개념(혼인 상태)과 개별 측정 항목(혼인 상태의 종류) 모두 개념입니다. - 측정에 사용되는 값들은 DDI CDI에서 '실질적 값(substantive values)'이라고 합니다.

Represented Variable	○ 측정하고자 하는 모집단의 특성과 그 측정값에 대한 표현방식의 조합 - Conceptual Variable + Value Domain (substantive) ○ 표현방식 - 열거형(enumerated) 값 도메인 → 코드목록, 통계분류체계 등 - 기술형(described) 값 도메인 → 값에 대한 설명
----------------------	---

	○ 포함정보 - 자료유형 → 명목, 서열, 구간 등 - 측정단위 → (예) 년, 톤, 건수, 달러 등															
용어집_정의	○ combination of a characteristic of a population to be measured and how that measure will be represented															
용어집_설명	○ 이 측정 방법은 양적, 범주형 및 서술형 개념변수에 적용됩니다. - (예) (직원 수, 정수) 쌍에서 '직원 수'는 모집단의 특성(개념변수)이고 '정수'는 해당 측정값을 표현하는 방식(실질적 값 영역)임 - (예) 개념변수가 '산업'이고 실질적 값 영역이 'NACE 2007 레벨 1'이라면, 쌍은 (산업, NACE 2007 - 레벨 1)이 됨 - (예) 표현변수 '성별[1,2,3]'은 개념변수(성별)와 표현방식(1=남성, 2=여성, 3=기타)으로 구성됨															
이용자안내	63. GSIM은 객체(object)의 의미와 구체적 자료표현을 이해할 수 있도록 지원함 - 이에 따라, 정보를 개념적으로 설명하는데 사용되는 클래스와 표현방식을 나타내는 클래스로 구분함 64. 표현변수는 Value Domain과의 연계를 통한 결과값 표현 방법에 대한 정보를 나타냄 - 예) 개념변수가 '개인의 교육수준'인 경우, Code List(ISCED: 1.초, 2.중, 3.고, 4.대졸이상)와 연계될 수 있음 - 개념 도메인은 개념변수와 연계되는 반면, 값 도메인은 표현변수와 연계됨 65. 표현변수 표현방식 - 열거형(Enumerated) 값 도메인: 코드목록이나 통계분류체계의 조합으로 표현 - 기술형(Described) 값 도메인: 명시적 열거 대신, 값에 대한 설명 (예: 0~1 사이 모든 실수)으로 표현															
웹비나	○ Conceptual Variable + Value Domain (substantive) ○ Includes - 측정단위 (Unit of measure if applicable) - Intended datatype - Expression of how data are intended to be used - Not how they are stored, or what they look like - Example - Marital status categories represented with numerals - NOT numeric - Still nominal - unordered set of categories - Ranges for counts, measures, scales, and date/times ○ Value Domains for Marital Status<table><tr><td>VD4-A (CD4)</td><td>VD4-N (CD4)</td><td>VD2 (CD2)</td></tr><tr><td><s, single></td><td><1, single></td><td><s, single></td></tr><tr><td><m, married></td><td><2, married></td><td><m, married></td></tr><tr><td><d, divorced></td><td><3, divorced></td><td></td></tr><tr><td><w, widowed></td><td><4, widowed></td><td></td></tr></table> ○ Value Domain for Age VD = { 0 <= whole number, represented in Arabic numerals } ○ Include bounds of ranges	VD4-A (CD4)	VD4-N (CD4)	VD2 (CD2)	<s, single>	<1, single>	<s, single>	<m, married>	<2, married>	<m, married>	<d, divorced>	<3, divorced>		<w, widowed>	<4, widowed>	
VD4-A (CD4)	VD4-N (CD4)	VD2 (CD2)														
<s, single>	<1, single>	<s, single>														
<m, married>	<2, married>	<m, married>														
<d, divorced>	<3, divorced>															
<w, widowed>	<4, widowed>															
DDI-CDI	○ Represented Variable 수준에서 가장 중요한 추가 사항은 실체적 범주(substantive categories)입니다. 예시에서 다음과 같은 명칭이 추가될 수 있습니다. ▪ <s, single> ▪ <m, married> ▪ <d, divorced> ▪ <w, widowed> ○ 이러한 명칭들의 집합을 실질적 값 도메인(Substantive Value Domain)이라고 합니다. 이러한 개념들은 변수의 주제와 관련되어 있으며, 처리 과정과는 관련이 없습니다.															

	하나의 실질적 값 도메인은 여러 표현변수(Represented Variable)에서 사용될 수 있으므로, 이를 식별하고 관리하는 것이 중요합니다.												
Instance Variable	- 표현변수를 데이터세트에서 구현한 것으로, 데이터베이스의 컬럼에 해당함 - Represented Variable + Value Domain (sentinel) - 포함정보 - 어플리케이션(R, SAS 등) 데이터타입 - 센티널 값 도메인 (결측값, 응답거절, 에러값 등 자료처리와 관련된 값)												
용어집_정의	use of a Represented Variable within a Data Set												
용어집_설명	- 인스턴스 변수는 수집된 데이터의 실제 인스턴스를 설명하는 데 사용됩니다. 다음은 세 가지 예입니다. - 성별: Dan Gillman은 <m, 남성> 성별, Arofan Gregory는 <m, 남성> 성별, ... - 직원 수: Microsoft는 직원수 90,000명, IBM은 직원 수 433,000명, ... - 기부금: Johns Hopkins는 기부금 <3, \$1,000,000 이상>, Yale은 기부금 <3, \$1,000,000 이상>, ... 데이터 출처 정보도 포함가능												
GSIM e-training	Instance Variable <table border="1" style="margin-right: 10px;"> <thead> <tr><th>Unit</th><th>Gender</th></tr> </thead> <tbody> <tr><td>Dan Gillman</td><td>m</td></tr> <tr><td>Arofan Gregory</td><td>m</td></tr> </tbody> </table> Here are 2 examples: 1) Gender: Dan Gillman has gender <m, male>, Arofan Gregory has gender<m, male>, etc. 2) Number of employees: Microsoft has 90,000 employees; IBM has 433,000 employees, etc. <table border="1" style="margin-top: 10px;"> <thead> <tr><th>Unit</th><th>Number of Employees</th></tr> </thead> <tbody> <tr><td>Microsoft</td><td>90,000</td></tr> <tr><td>IBM</td><td>433,000</td></tr> </tbody> </table>	Unit	Gender	Dan Gillman	m	Arofan Gregory	m	Unit	Number of Employees	Microsoft	90,000	IBM	433,000
Unit	Gender												
Dan Gillman	m												
Arofan Gregory	m												
Unit	Number of Employees												
Microsoft	90,000												
IBM	433,000												
이용자안내	67. 인스턴스 변수는 Data Point를 통해 Data Set에 연결된 표현변수의 특수화로, 데이터베이스의 컬럼에 해당함 (예) 'ISCED 기준 교육정도'은 인스턴스 변수에 의해 설명되는 데이터 컬럼으로, 이는 '교육정도'라는 개념변수와 ISCED로 표현되는 값 도메인이 결합된 것임												
웹비나	- Represented Variable + Value Domain (sentinel) - Includes - Application datatype (what data look like and how the are stored) - Sentinel Value Domain - Allowed values representing processing needs - Missing, Refused, other (e.g., processing error) - Depends on applications (e.g. R, SAS, SPSS, and Stata represent missing and refused differently) - Example <m, missing> <r, refused> <p, processing error>												
DDI-CDI	- 인스턴스 변수는 데이터세트에서 사용되는 변수입니다. 각 데이터세트마다 새로운 인스턴스 변수가 생성됩니다. 주요한 추가 사항은 센티널 범주(sentinel categories)에 명칭(designation)을 부여하는 것입니다. 또한 센티널 값(명칭) 목록은 SentinelValueDomain이라는 단일 집합에서 관리됩니다. 실질적인 값 도메인과 센티널 값 도메인을 분리함으로써 메타데이터 관리 부담이 줄어듭니다. 한 종류의 값 도메인에서 필요한 변경 사항은 다른 도메인에 영향을 미치지 않습니다. 센티널 값 도메인의 지정 예시는 다음과 같습니다. <m, missing> <r, refused> - 인스턴스 변수는 데이터세트의 데이터와 연결되어 있으므로 해당 변수에 사용되는 데이터의 데이터 형식 또한 필요합니다.												

HLG-MOS (2026)	<센티널 값에 대한 사업 타당성 분석 (제안된 활동일자: 2026년 1~11월)> ○ 값 도메인은 변수가 할당할 수 있는 허용 가능한 값을 나타냅니다. 값에는 두 가지 유형이 있습니다. - 하나는 변수와 의미적으로 연결된 값, 즉 주제의 구분을 나타내는 값이고, 다른 하나는 데이터의 수집 및 처리와 연결된 값입니다. 이 두 가지 유형은 각각 실질적 값과 센티널 값이라고 합니다. GSIM에서는 이러한 값들을 정의하고 모델링합니다. ○ 가장 많이 사용되는 센티널 값은 결측값과 거부된 데이터에 대한 값입니다. - 이러한 값은 매우 흔하게 사용되므로 주요 통계 패키지(R, SAS, SPSS, Stata) 모두에서 이를 표현하는 기능을 제공합니다. 흥미롭게도 이러한 표현 방식(코드)은 패키지마다 다릅니다. ○ 센티널 값은 처리 과정에서도 발생할 수 있습니다. - 예를 들어, 일부 데이터는 계산하는 데 여러 단계가 필요할 수 있습니다. 중간 데이터세트에는 아직 계산되지 않은 데이터에 대한 값이 포함되어야 하며, '결측'이나 '거부'로는 해당 상황의 정확한 의미를 파악할 수 없습니다. 미완성이거나 진행 중이라는 표현이 가능합니다. ○ 복잡한 설문조사에서는 응답자가 요청된 정보를 가지고 있지 않거나 찾을 수 없는 경우가 있을 수 있습니다. - 이러한 경우 '알 수 없음'은 유용한 기준값이 될 수 있는데, '결측'이나 '응답 거부'로는 해당 상황을 제대로 포착할 수 없기 때문입니다. ○ 각 통계 패키지마다 결측값과 거부된 값에 대해 서로 다른 코드를 사용하기 때문에, 표준적인 센티널 값 코드목록을 개발하는 것은 비효율적입니다. 따라서 목표는 센티널 값을 정의하는 공통 범주 집합을 개발하는 것이 실현 가능하고 유용한지 여부를 확인하는 것입니다. ○ 따라서 이 문제를 조사하기 위한 그룹을 구성하자는 제안입니다. ○ 이 그룹의 결과물은 누락 및 거부에 대한 다양한 인코딩을 포함하는 범주 집합이 될 것입니다. 다른 범주를 나타내는 코드 사용에 대해서는 합의가 이루어지지 않을 것으로 예상하지만, 가능성은 있습니다. 그룹은 범주가 얼마나 보편적으로 사용되는지를 기준으로 포함 여부를 결정해야 합니다. 유사한 범주 간의 조화는 기대되며 바람직합니다.
-------------------	--

Measurement Type	○ 측정값의 유형(예: 질량 또는 통화)을 정의 - 서로 변환 가능한 모든 측정단위를 그룹화함
용어집_정의	○ defines the type of a measure e.g. mass or currency.
용어집_설명	○ 측정 유형은 서로 변환 가능한 모든 측정단위를 그룹화함 ○ 각 측정 유형은 서로 다른 측정단위 간의 변환에 사용할 수 있는 표준 측정단위를 최대 1개 가질 수 있음 - 단, 특정 측정 유형(예: 통화)에 대해 반드시 표준 측정단위가 있어야 하는 것은 아님

Measurement Unit	○ 공식적인 측정단위의 관점에서 측정값에 부여되는 척도(metric)
용어집_정의	○ metric for a measurement in terms of an official unit of measurement
용어집_속성	○ Abbreviation (0..*) (String): 측정단위의 약어 (예) 킬로그램의 경우 kg ○ Conversion Rule (0..1) (String): 표준 측정단위로의 변환 규칙 (단, 존재하는 경우)
용어집_설명	○ 측정단위는 관례나 법률에 의해 정의되고 채택된, 동일한 종류의 양을 측정하는 표준으로 사용되는 양의 명확한 크기입니다.

	○ 측정단위는 무게, 높이, 통화, 시간 등 다양한 측정 유형을 기반으로 할 수 있습니다. ○ 측정단위는 동일한 측정 유형 (예: 길이) 을 나타내는 경우 서로 변환할 수 있습니다(예: 킬로미터를 미터로 변환). ○ 변환 규칙 속성을 사용하여 곱셈 계수를 포함할 수 있습니다. 예를 들어, 비표준 측정단위 '1000 kg'은 표준 측정단위 'kg'에 1000을 곱한 값입니다.
--	--

Datum	○ 측정값
용어집_정의	○ value
용어집_설명	○ 데이터(Datum)는 획득되거나 도출된 데이터의 실제 인스턴스입니다. - 데이터 포인트(Data Point)를 채우는 값이며, 테이블의 셀에 있는 값입니다.

Conceptual Domain	○ 유효한 개념의 집합
용어집_정의	○ set of valid Concepts
용어집_속성	○ Sentinel(1.1)(Boolean): 참이면 해당 도메인은 센티널 도메인(즉, 처리 수명 주기의 상태를 나타내는 데 사용되는 값(예: 누락된 데이터))이고, 그렇지 않으면 실질적 도메인(즉, 관심 있는 단위의 관측값을 나타내는 데 사용되는 값)입니다.
용어집_설명	○ 개념 도메인은 열거형 또는 표현식을 통해 설명할 수 있습니다.
웨비나	○ Conceptual domain is source of meanings of allowed values ○ Correspondence between similar Value Domains ○ Example of VD equivalence using a marital status CD - CD: {single, Married, Divorced, Widowed} - VD: {<s, single>, <m, married>, <d, divorced>, <w, widowed>}

Described Conceptual Domain	○ 표현식으로 정의된 개념 도메인
용어집_정의	○ Conceptual Domain defined by an expression
용어집_설명	○ (예) 0과 1 사이의 모든 실수

Enumerated Conceptual Domain	○ 범주 목록으로 표현된 개념 도메인
용어집_정의	○ Conceptual Domain expressed as a list of Categories
용어집_설명	○ (예) '성별' 범주: '남성', '여성'

Value Domain	○ 개념변수에 허용되는 값들의 집합
용어집_정의	○ set of permissible values for a Conceptual Variable
용어집_설명	○ 값_도메인은 열거형 또는 표현식으로 설명할 수 있습니다.
이용자안내	66. 값 도메인은 자료유형과 측정단위 정보를 포함함 - 자료유형은 해당 자료Datum(명목, 서열, 구간 등)에 허용가능한 연산 정보를 포함 - 측정단위는 값 도메인의 측정내용을 구체화(예: 년, 톤, 건수, 달러 등) (예) 성별 코드는 명목형 통계 자료로, 연령 값은 구간 자료로 이어짐
웨비나	○ A set of allowed values (meanings and their representations) (예) Marital status categories → <s, single>;<m, married>;<w, widowed>

	Age range in years $\rightarrow \{n \mid 100 > n \geq 0\}$ in Arabic numerals - the allowed values (meaning and the representation) for variables (예) Marital status {s, single}; {m, married}; {d, divorced}; {w, widowed} Age {0 ≤ whole number, and in Arabic numerals} - They come in 2 type systems (Usage and Structure) 1) Substantive versus Sentinel (Usage) - Substantive: allowed values based on subject matter (예) marital status classification; age range - Sentinel: allowed values based on processing (예) missing, refused, or other processing needs 2) Enumerated versus Described (Structure) - Enumerated - Each allowed value is listed - An exhaustive list - Described - rule or range - Rule: text strings for qualitative data (예) names (of people or businesses); identifiers - Range: dates and times (begin to end) - Range for quantity - numeric (min to max) - Integer for counts - Real for inexact quantities - Scaled for currencies
ISO 11179-1	- 값 도메인(Value Domain)은 데이터 요소에 허용된 값들의 집합이다. - 데이터 타입(Datatype)과 필요한 경우 측정단위(Unit of Measure)가 값 도메인과 결합된다. - 예를 들어 '연간 가구 소득'을 나타내는 데이터 요소는 아라비아 숫자로 된 비음수 정수 집합(단위는 달러)을 유효한 값의 집합으로 가질 수 있다. 이것이 그 데이터 요소의 값 도메인이다. '통화(Currency)'는 표현 클래스(Representation Class)가 될 수 있고, '달러(Dollars)'는 측정단위가 된다. ** (nk) 표현클래스는 GSIM 에서는 측정 유형 3.3.7 value meaning: 값(value)이 담고 있는 의미론적 내용(semantic content) - 값 도메인의 표현 방식에 독립적임 3.3.9 value domain (VD): 허용가능한 값들의 집합 - 값 도메인은 표현방식(representation)을 제공할 뿐, 어떤 개념과 연결되는지 혹은 그 값이 무엇을 의미하는지는 관련이 없음 - 허용가능한 값은 열거형, 서술형, 혹은 둘의 혼합형 모두 가능함 3.3.10 permissible value: 해당 값의 의미를 나타내는 지정(designation) - 지정(designation)으로서, 값(value)은 기표(sign)이며, 값의 의미(value meaning)는 개념(concept)임 3.3.11 conceptual domain (CD): 값의 의미(value meanings)들을 열거형 집합이나 하위 개념에 대한 설명, 혹은 이 두가지 모두로 묶어서 표현한 개념(concept)

Described Value Domain	표현식으로 정의된 값_도메인
용어집_정의	Value Domain defined by an expression
용어집_속성	Data Type(1..1)(String)
용어집_설명	(예) 0과 1 사이의 모든 십진 실수 (demical 이라는 물리적 표현 형식 추가)

Enumerated	범주 및 관련 코드목록으로 표현된 값_도메인
------------	--------------------------

Value Domain	
용어집_정의	○ Value Domain expressed as a list of Categories and associated Codes
용어집_설명	○ (예) 성별 코드: <m, 남성>; <f, 여성>; <o, 기타>.

Substantive Value Domain	
용어집_정의	○ Value Domain containing substantive values, where a substantive value is subject matter related
용어집_설명	○ 실질적 값은 성별 분류에서 <f, 여성>과 같이 주제 내용을 나타내는 값입니다. 가능한 값의 범위와 의미는 실질적 값 영역과 연관된 개념 영역의 틀 안에서 정의됩니다. 예를 들어, <0, 유아>, <1, 초등>, <2, 중등학교>, <3, 고등학교>, <4, 고등교육(대학 교육 제외)>, <5, 대학교 1단계>, <6, 대학교 2단계>와 같이 값의 범위와 의미는 교육 수준별 범주 내에서 정의됩니다.

Sentinel Value Domain	
용어집_정의	○ Value Domain containing sentinel values, i.e. processing-related special values
용어집_설명	○ 센티널 값은 처리 목적으로 사용되며, 누락 또는 거부와 같이 주제 내용이 없는 값입니다. - 센티널 값 도메인은 열거(목록화)하거나 설명할 수 있습니다. 값 도메인은 센티널 값에 대한 범주 목록 또는 해당 설명으로 표현됩니다. 가능한 값의 범위와 의미는 센티널 값 도메인과 연결된 개념 도메인 내에서 정의됩니다. - 센티널 값을 실질적인 값과 분리하면 유지 관리해야 하는 값 도메인, 표현변수 및 인스턴스변수의 수를 크게 줄일 수 있습니다. - 대부분의 코드목록에 나타나는 개념에는 일반 코드를 사용하는 것이 좋습니다. 예를 들어 <S_X, 지정되지 않음>, <S_Z, 해당 없음>, <S_R, 거부>, <S_U, 알 수 없음>과 같습니다.

Unit	
용어집_정의	○ 정보 수집대상인 동시에 궁극적으로 통계 자료가 작성되는 개체(entity) - 개별 개체를 의미함 (예. '홍길동', '이순신' 등)
용어집_설명	○ (예) Individual US person (e.g., Arofan Gregory, Dan Gillman, Barack Obama, etc.) ○ (예) Individual US computer companies (e.g., Microsoft, Apple, IBM, etc.) ○ (예) Individual US universities (e.g., Johns Hopkins, University of Maryland, Yale, etc.)
이용자안내	60. 단위(Unit)는 통계 프로세스에서 관심의 대상이 되는 엔티티(entity)임 - 단위에 대해 자료가 수집됨 - 단위는 정보를 얻을 수 있는 모집단과 연관된 개별 개체임 - 예) 특정 개인, 특정 사업체, 특정 기업체, 특정 대학교
웨비나	○ The objects on which we collect data
MCV(2009) Unit 정의	○ Observation unit - Definition: Those entities for which information is received. - Context: During the collection of data, this is the unit for which data is recorded. It should be noted that this may, or may not be, the same as the reporting unit. - Source: SDMX (2009) ○ Statistical unit - Definition: Entity for which information is sought and for which statistics are ultimately

	compiled. - Context: The statistical unit is the object of a statistical survey and the bearer of statistical characteristics. These units can, in turn, be divided into observation units and analytical units. Statistical units for economic statistics are defined on the basis of three criteria: 1) Legal, accounting or organisational criteria; 2) Geographical criteria; 3) Activity criteria. Statistical units comprise the enterprise, enterprise group, kind-of-activity unit (KAU), local unit, establishment, homogeneous unit of production, persons, households, geographical areas, events etc. In other statistical domains, statistical units can include persons, households, geographical areas, events etc. Statistical units can be categorised into basic statistical units, i.e. those for which data is collected, and derived statistical units, i.e. those which are constructed during the statistical production process. A basic statistical unit is the most detailed level to which the obtained characteristics can be attached. - Source: SDMX (2009)
--	--

Unit Type	○ 단일 특성을 기준으로 분류된 단위들의 집합 또는 그룹 - 단위들의 집합이 갖는 단일 특성에 해당하는 추상적 개념 (예. '개인', '기업체' 등)
용어집_정의	○ class or group of Units based on a single characteristic
용어집_설명	○ 단위 유형은 시간이나 지리적 범위를 명시하지 않고 단일 특성을 기준으로 단위들의 집합 또는 그룹을 설명하는 데 사용됩니다. ○ 예를 들어, '사람'이라는 단위 유형은 '사람'이라는 특성을 가진 단위들을 하나의 그룹으로 묶은 것입니다. ○ 이는 통계 자료 배포에 사용되는 단위 유형 뿐만 아니라 통계 처리 과정 전반에 걸쳐 적용됩니다. 예를 들어, 행정자료를 사용할 때 재정 단위(fiscal unit)를 활용할 수 있습니다.
이용자안내	60. 단위 유형(Unit Type)은 단위를 추상적 개념으로 구분함 (A Unit Type is a way of identifying an abstract conceptualisation of Unit)
웨비나	○ 기관에서 데이터를 수집하는 기본유형 (Fundamental class of units for which an organization acquires data) ○ 표본은 단위유형에 해당하는 단위들의 부분집합임 (Sample is a subset of the set of units corresponding to a unit type) (예) US Bureau of Labor Statistics → People, Households, Business establishments (예) US Department of Labor (not BLS) → Business establishments, Inspections, Enforcements, Program recipients (예) US Fish and Wildlife Service → animals (Fishes, Mammals, Birds) ○ Note, people are animals, but the organizational context matters

Universe	○ 시간이나 지리적 구분 없이, 주제적 특성에 따른 단위 유형의 특수화 - 즉, 주제적 특성에 따른 개념적 집단 (저자 정리: Unit Type에 맥락을 부여하는 것으로 이해)
용어집_정의	○ specialized Unit Type, but not by time or geography
용어집_설명	○ 유니버스에 대한 설명은 일반적으로 '대학 학위 소지자'와 같이 포괄적인 용어로 작성됩니다. ○ 때때로 유니버스는 '대학 학위 소지자를 제외한 모든 사람'과 같이 제외 대상을 명시하여 정의되기도 합니다.

	○ 두 경우 모두 대학 학위 소지자라는 조건을 추가하면 특정 집단으로 분류되는데, 이는 유니버스 유형을 정의하는 기준이 됩니다.
이용자안내	60. Universe는 단위 유형(Unit Type)의 구체화임. (단, 시·공간적으로는 구체화하지 않음)
웹비나	○ 주제적 특성에 따른 단위유형의 특수화 (Specializaion of a Unit Type by subject matter charateristics) (예) Women ← People (단위유형); female, 18세이상 (특성) (예) Mines ← 사업체 (단위유형); 광물 채굴 (특성)

Population	○ 정의된 분류의 사람, 객체, 사건의 전체 구성원 - 시·공간적 조건이 부여된 유니버스의 특수화 - (예) Unit Type: 사람 (가장 넓은 개념) - Universe: 한국의 성인 (조건이 붙은 개념) - Population: 2025년 인구주택총조사 기준 한국 성인 전체 (시점과 범위가 확정된 최종 개념)
용어집_정의	○ total membership of a defined class of people, objects or events
용어집_속성	○ Geography(0..1)(String): 모집단의 지리적 범위 ○ Reference Period(0..1)(Date): 모집단의 시점(기간)
용어집_설명	○ 모집단은 시간 및 지리적 경계와 같은 특성을 기준으로 사람, 사물 또는 사건 집단의 전체 구성원을 나타내는 데 사용됩니다. (예) 1956년 11월 13일 유럽의 성인 인구 (예) 2012년 말 미국의 컴퓨터 회사 (예) 2023년 1월 1일 전 세계 대학교
이용자안내	60. 모집단은 사용되는 프로세스 단계에 따라 여러 종류가 있음(목표 모집단, 조사 모집단, 표본추출틀, 분석 모집단 등) - 모집단은 유니버스에 대한 구체화임
웹비나	○ 시간과 공간 조건으로 유니버스를 특수화 (specialization of a universe by time and location) (예) 2025년 미국 Women (예) 2020년 콜로라도 Mines ○ Unit Type / Universe / Population - 상위단계에서 하위단계로 구체화되는 '계층적 구체화' 구조 (Specialization path (Unit Cascade in DDI)) - 변수별 측정대상과 연결됨 (Tied to descriptions of variables)
DDI-CDI	○ Universe에 시간과 공간 특성이 추가된 것으로, (population의 모든 구성원을 측정할 수 있는 경우, 대표 표본이 되는) 단위들의 선택으로 표현됨 - The Population is a Universe with the additional qualifying characteristics of time and space added, expressed as a selection of Units (the represented sample in cases where not every member of a population is measured).

Category	○ 특성을 외연적으로 정의하고 측정하는 역할을 하는 개념 - (예) 개념: Gender → 범주: Male, Female
용어집_정의	○ Concept whose role is to extensionally define and measure a characteristic
용어집_설명	○ 성별 개념의 범주는 남성, 여성입니다. * 참고: 외연적 정의는 하나의 기준 또는 하위 분류 아래에 모든 하위 개념을 열거하여 개념을 설명하는 것임. 예를 들어, 주기율표의 비활성 기체는 헬륨, 네온, 아르곤, 크립톤, 제논, 라돈을 포함하는 원소 집합으로 외연적으로 정의됨(ISO 1087-1). ○ 유의어: class
이용자안내	58. 개념은 3가지 상황에서 사용될 수 있는데, - 그 중 하나는 개념의 세부 사항을 더 자세히 정의하기 위한 '범주'임. (예) '성별(gender)' 개념에 대한 '남성(Male)', '여성(Female)'

	- '코드목록'이나 '통계분류'에서 사용하기 위해 노드(코드항목, 분류항목)를 통해 코드를 범주에 연계할 수 있음. 68. 범주는 특성을 외연적으로 정의하기 위한 역할을 하는 개념의 특정한 유형임. - 범주가 사용되는 방식은 3가지로서, GSIM에서는 노드의 3가지 하위유형인 범주항목, 코드항목, 분류항목으로 설명함. - 범주들은 사용되는 방식에 따라 3가지 유형의 노드세트(항목세트, 코드목록, 통계분류)에 포함됨.
웨비나	○ 범주는 특성을 외연적으로 정의하는 역할을 하는 개념의 특수한 유형으로, 노드에 포함됨 (A Category is a particular type of concept, whose role is to define a characteristic extensionally, contained in Node)

Node	○ 범주와 관련 속성의 결합 - 총 3가지의 유형으로 사용됨: 범주항목, 코드항목, 분류항목
용어집_정의	○ combination of a Category and related attributes
용어집_속성	○ 집계유형(Aggregation Type)(0..1)(String): 노드 간의 부모/자식 관계를 정의할 때는, 전체(whole)/부분(part) 관계인지, 상위(super)/하위(sub) 관계인지를 규정해주어야 함 - (저자 정리) 전체/부분: 물리적 구성 관계 예) 전국/시도 상위/하위: 개념적 상속 관계 예) 사람/남성, 산업/광업
용어집_설명	○ 노드는 해당 범주가 사용되는 상황을 정의하기 위해, 범주항목, 코드항목 또는 분류항목으로 생성됩니다.
이용자안내	58. 개념은 3가지 상황에서 사용될 수 있는데, - 그 중 하나는 개념의 세부 사항을 더 자세히 정의하기 위한 '범주'임. (예) '성별(gender)' 개념에 대한 '남성(Male)', '여성(Female)' - '코드목록'이나 '통계분류'에서 사용하기 위해 노드(코드항목, 분류항목)를 통해 코드를 범주에 연계할 수 있음. 68. 범주는 특성을 외연적으로 정의하기 위한 역할을 하는 개념의 특정한 유형임. - 범주가 사용되는 방식은 3가지로서, GSIM에서는 노드의 3가지 하위유형인 범주항목, 코드항목, 분류항목으로 설명함. - 범주들은 사용되는 방식에 따라 3가지 유형의 노드세트(항목세트, 코드목록, 통계분류)에 포함됨.
웨비나	○ 노드는 범주와 관련 속성의 조합으로, 해당 범주가 사용되는 상황을 정의하는데 사용됨(3가지 하위 유형) - A Node is a combination of a Category and related attributes for the purpose of defining the situation in which the Category is being used (3 subtypes): - 범주항목은 표현(코드)없이 범주의 의미만 포함함 Category item contains the meaning of a category without a representation (Code) (예) Growing of rice - 코드항목은 표현(코드)와 함께 범주의 의미를 포함함 Code item contains the meaning of a category with a representation (Code) (예) A0112 Growing of rice - 분류항목은 특정 레벨에서 범주와 표현을 결합함 Classification item combine a category at a certain level with a representation (Code) (예) 01.12 Growing of rice

Code	○ 범주에 대한 명칭(designation)임
용어집_정의	○ Designation for a Category
용어집_설명	○ 코드는 코드목록 내에서 고유합니다. (예) M(남성), F(여성)
OECD 용어집	○ A language-independent set of letters, numbers or symbols that represent a concept whose meaning is described in a natural language. Context: A code normally consists of one or more alphabetic, numeric or alpha/numeric characters. Source: Statistical Data and Metadata Exchange (SDMX) - BIS, ECB, Eurostat, IBRD,

	IMF, OECD and UNSD - Metadata Common Vocabulary See also: Code list; Coding
--	--

Designation	○ 개념을 표기하는 위해 부여된 기호적(sign) 명칭
용어집_정의	○ association of a Concept with a sign that denotes it
용어집_설명	○ 식별을 위해 객체(object)에 부여하는 이름입니다.
ISO 11179-1	○ 실제 객체 클래스나 속성을 그 명칭(Name)과 구별하는 것은 매우 중요하다. 이는 개념(Concept)과 그에 대한 지정(Designation) 사이의 구분이다. 객체 클래스와 속성은 개념이며, 그 명칭은 지정이다. - 사람들은 단어(지정)를 통해 개념을 전달하기 때문에 개념과 그 지칭을 혼동하기 쉽다. 예를 들어, 대부분의 사람들은 '소득(Income)'이라는 단어를 읽고 자신이 모호함 없이 해석했다고 확신할 것이다. 그러나 '소득'이라는 지정이 모든 독자에게 동일한 개념을 전달하지 않을 수 있으며, 더 중요한 것은 '소득'이라는 각각의 사례가 모두 동일한 개념을 지칭하는 것도 아니라는 점이다. - 모든 아이디어가 자연어(Natural language)로 쉽게 표현되는 것도 아니다. 예를 들어 '최근 12개월 이내에 적어도 한 번의 생존 출아를 경험한 15세에서 45세 사이의 여성'은 유효한 객체 클래스이지만 영어로 쉽게 명명하기 어렵다. 어떤 아이디어는 특정 언어에서 다른 언어보다 더 쉽게 표현되기도 한다. 예를 들어 독일어 'Götterdämmerung(신들의 황혼)'은 영어로 간단히 대응되는 단어가 없다.

Category Item	○ 범주세트에서만 사용되는 노드 유형으로, 특정 표현없이 범주의 의미만을 담고 있음 - 1개의 범주항목은 1개의 범주에 대한 의미만 포함함
용어집_정의	○ type of Node exclusive to a Category Set that contains a single Category
용어집_설명	○ 범주항목은 특정 표현 없이 범주의 의미만을 담고 있습니다. (예) '남성'
이용자안내	69. 범주세트는 범주항목 형태로 범주들을 포함한다. 범주항목은 어떠한 관련 표현방식 없이 범주의 의미만을 포함한다. (예) An example of a Category Set is: Male, Female

Code Item	○ 코드목록에서만 사용되는 노드 유형으로, 범주의 의미와 이를 표현하는 코드의 결합 - 1개의 범주항목은 1개의 범주에 대한 의미만 포함함
용어집_정의	○ type of Node exclusive to a Code List that combines a Category with a Code that represents it
용어집_설명	○ 코드항목은 포함된 범주의 의미와 코드 표현을 결합한 것입니다. - 코드들은 해당 코드목록 내에서 고유합니다. (예) M(남성), F(여성)
이용자안내	70. 코드목록은 코드항목 형태로 범주들을 포함한다. 코드항목은 코드 표현과 범주의 의미를 함께 포함한다. (예) An example of a Code List is: 1-Male, 2-Female

Classification Item	○ 통계분류에서만 사용되는 노드 유형으로, 특정 레벨의 범주와 이를 표현하는 코드의 결합 - 1개의 범주항목은 1개의 범주에 대한 의미만 포함함
용어집_정의	○ type of Node exclusive to a Statistical Classification that combines a Category at a certain Level with a Code that represents it.
용어집_속성	○ Case Laws(0..*)(MultilingualText) - 분류항목에 대한 분류사례 식별자 (Refers to identifiers of one or more case law rulings related to the Classification Item.) ○ Case Law Descriptions(0..*)(MultilingualText) - 분류사례 설명 (Refers to descriptions of the case laws.)

	○ Case Law dates(0..*)(Date) - 분류사례 날짜 (Refers to date of case laws.) ○ Generated(0..1)(Boolean) - 분류 레벨의 완전성을 유지하기 위해 생성된 분류항목인지 여부 (Indicates whether or not the item has been generated to make the level to which it belongs complete.) (저자 정리: 기타 항목 혹은 가상 항목) ○ Linked Items(0..*)(String) - 연계표(correspondence tables)를 통해 현재 항목과 연결된 다른 버전 혹은 다른 변형 분류의 항목 (Items of other classification versions or variants with which the item is linked, either as source or target, through Correspondence Tables.) (저자 정리: 9차 산업분류와 10차 산업분류, 한국산업분류와 국제산업분류)
용어집_설명	○ 분류항목은 해당 범주의 내용과 경계를 정의합니다. - 하나의 단위는 통계분류의 각 수준에서 오직 하나의 항목으로만 분류될 수 있습니다. - 범주들은 하위 모집단(sub-population)을 생성하는 데 사용되며, 통계분류에 포함될 때 상호배타적(mutually exclusive)이어야 합니다.
이용자안내	71. 통계분류는 분류항목이 코드 표현과 범주의 의미를 결합한다는 점에서 코드목록과 유사하다. 하지만 통계분류의 내용은 추가적인 기준을 충족해야 한다. 즉, 분류항목은 해당 통계분류 내에서 자신이 속한 레벨을 기준으로 반드시 상호 배타적(mutually exclusive)인 동시에 포괄적(jointly exhaustive)이어야 한다. (예) An example of a Statistical Classification is: 국제표준산업분류(ISIC), 표준직업분류(SOC), 북미산업분류체계(NAICS), 목적별 개별소비지출분류(COICOP)

Node Set	○ 노드들의 집합 - 개념체계의 일종
용어집_정의	○ set of Nodes
용어집_설명	○ Node Set is a kind of Concept System. Here are 2 examples: 1) Sex Categories - Male - Female - Other 2) Sex Codes - <m, male> - <f, female> - <o, other>
웨비나	○ A Node Set = a kind of Concept System (3 subtypes) - Category Set for grouping Categories via Category Items (예) Subject fields (tourism, population) - Code List for grouping pairs of Categories and their Codes via Code Items (예) Code List of Activities (A0112 Growing of rice) - Statistical Classifications, hierarchically organized set of mutually exclusive and jointly exhaustive Categories that share the same or similar characteristics (예) NACE Rev.2.1 (01.12 growing of rice)
이용자안내	73. 통계분류, 코드목록, 범주세트의 공통점은 노드세트의 하위 유형으로써 그들의 연결을 상속하는 것이다. 유사하게, 이들 3가지 그룹을 구성하는 3가지 항목 유형도 노드의 하위 유형이다.

Category Set	○ 범주항목을 통해 범주들을 그룹화하는 노드세트의 유형
용어집_정의	○ type of Node Set for grouping Categories via Category Items
용어집_설명	○ 범주세트 내의 범주들은 일반적으로 지정된 명칭(Designations)(코드(Codes))가 없습니다. (예) 남성, 여성

Code List	○ 코드항목을 통해 범주와 그 코드 쌍을 그룹화하는 노드세트의 유형
용어집_정의	○ type of Node Set for grouping pairs of Categories and their Codes via Code Items
용어집_설명	○ 유사한 코드목록들은 (노드세트에서 상속받은 '관련됨(related to)' 관계를 통해) 함께 그룹화될 수 있음. ○ 코드목록은 열거형 값 도메인에서 사용할 수 있는 허용값들의 집합을 사전에 정의하여 제공함. (A Code List provides a predefined set of permissible values for an Enumerated Value Domain.)
이용자안내	72. 코드목록은 통계분류와 동일한 기준이 요구되지 않음 - 코드목록은 결측값 포함과 같이 코드목록의 특정 용도를 지원하기 위한 추가 코드항목을 포함할 수 있음

Statistical Classification	○ 동일·유사한 특성을 공유하는 범주들을 상호 배타적이고(중복없고) 포괄적으로(누락없이) 계층화한 집합으로, 관심 모집단 내의 객체나 단위를 의미있게 그룹화(분류)하는데 사용됨
용어집_정의	○ hierarchically organised set of mutually exclusive and jointly exhaustive Categories that share the same or similar characteristics, used for meaningfully grouping the objects or units in the population of interest
용어집_속성	○ Changes from Base Statistical Classification (0..1) (MultilingualText) - 변형(variant)과 기본 통계분류 간의 관계를 설명하며, 여기에는 재분류, 집계 추가 및 확장 내용이 포함됨. ○ Changes from Previous Version or Update (0..1) (MultilingualText) - 이전 버전 또는 업데이트와의 변경사항에 대한 요약 설명임. 구체적인 변경 내용은 분류항목의 속성에 기록됨. ○ Copyright (0..*) (String) - 통계분류에는 저작권이 제한될 수 있음. 이러한 통계분류는 다운로드가 제한됨. 저작권 소유자를 명시하기 위해 공식 간행물에 표시해야하는 저작권 문구를 참고 ○ Current (0..1) (Boolean) - 통계분류가 현재 유효한지 여부 ○ Derived From (0..1) (String) - 해당 분류의 원천이 되는 분류버전을 나타냄 (저자 정리: KSIC는 ISIC로부터 파생) * 통계분류는 다른 분류시리즈의 분류 버전으로부터 파생될 수 있는데, 파생된 통계분류는 원래 분류 버전의 구조를 계승하면서 더 자세한 내용을 추가하거나, 원래 분류 버전의 상당부분을 사용하되 다른 구조로 재배열할 수 있음 ○ Floating (0..1) (Boolean) - 플로팅 분류인지 여부를 나타냄. 플로팅 통계분류는 모든 분류항목에 대해 유효기간을 정의해야하며, 이를 통해 다양한 시점에서 항목의 구조와 내용을 표시할 수 있음 ○ Introduction (0..1) (MultilingualText) - 소개에서는 통계분류에 대한 자세한 설명 및 생성·변경 배경, 분류 변수 및 분류 대상/단위, 분류 규칙 등을 제공 ○ Languages Available (0..*) (String) - 지원가능한 언어를 나타냄. 제공되는 언어, 번역 완료여부(전체/부분), 그리고 각 언어로 제공되는 부분 등을 표시 ○ Legal Base (0..*) (MultilingualText) - 통계분류가 법률이나 기타 공식 협약에 의해 규정되어있는지 나타냄

	○ Name Types (0..*) (ControlledVocabulary) - 통계분류에서 사용할 수 있는 대체 항목 이름의 유형 (정의된 이름유형 목록을 참조함) (저자 정리: 표준 명칭, 국제 명칭, 약칭 등?) ○ Predecessor (0..1) (String) - 버전이 관리되거나 업데이트되는 통계분류의 경우, 직전 통계분류를 명시 (저자 정리: 10차 분류의 successor = 9차 분류) ○ Successor (0..1) (String) - 차기 버전의 통계분류를 명시 (저자 정리: 9차 분류의 successor = 10차 분류) ○ Update (0..1) (Boolean) - 통계분류가 업데이트되었는지 여부 ○ Updates Possible (0..1) (Boolean) - 새로운 버전을 생성하지 않고 해당 분류 버전 내에서 업데이트 허용 여부 * 버전 내에서 허용가능한 구조적 변경 사항 명시: 분류항목의 추가, 재승인(revalidated), 무효화(invalidated) 등 * 버전 내에서 구조적 변경을 수반하지 않는 분류항목 이름 및 설명 노트와 같은 항목의 변경 가능 여부 명시
용어집_설명	○ 범주들은 관측단위로 구성된 특정 모집단의 하나 이상의 특성을 참조하여 정의됨. ○ 통계분류는 평면적이고 선형적인 구조를 가질 수도 있고, 하위 레벨의 모든 범주가 바로 위 레벨 범주의 하위 범주가 되는 계층적 구조를 가질 수도 있음. ○ 통계분류 내의 범주는 정보 모델에서 분류항목으로 표현됨.
이용자안내	74. 통계분류에 관한 구체적 내용은 'GSIM 통계분류모델'에서 확인할 수 있음 (GSIM v.1.1 하위 문서) 77. 통계분류는 범주의 의미와 코드 표현이 결합된 범주들의 집합임 - 통계분류는 통계 생산 및 보급에서 하나 이상의 표현변수에 할당될 수 있음 - 분류 구조의 각 레벨에 있는 범주들은 범위 내의 모든 객체/단위에 대해 상호배타적이고 전체 포괄적이어야 함 (예) 대분류(section), 중분류(division), 소분류(group) and 세분류(class) in ISIC rev 4. 78. 범주들은 특정 관심 모집단의 하나 이상의 특성을 참조하여 정의된다. - 통계분류는 평면적이고 선형적인 구조를 가질 수도 있고, 하위 레벨의 모든 범주가 바로 위 레벨 범주의 하위 범주가 되는 계층적 구조를 가질 수도 있다. 79. 분류항목은 범주의 의미와 그 표현(즉, 코드), 그리고 통계분류 기준을 충족하기 위한 추가 정보를 결합한 것이다. - 예를 들어, 'A-농업, 임업 및 어업'과 함께 이에 대한 포함 및 제외 사항에 대한 상세 설명문까지 함께 제공된다. 80. 통계분류는 크게 버전(version)과 변형(variant)로 나뉜다. - '변형' 형태는 '버전'을 기반으로 만들어지는데, 원본 통계분류의 표준 순서와 구조에 추가 사항이나 대안을 제공하기 위해 '버전'의 범주들을 분할, 통합(aggregation), 또는 재그룹화할 수 있음
웨비나	○ 통계분류와 관련 클래스에 대한 관점 (View of the Statistical Classification and its related management classes in GSIM) - GSIM 노드세트에 대한 특정 관점 (Particular view of the Node Set portion of GSIM) - 상세한 내용은 GSIM 통계분류모델 참고 (Further details can be found in the 'GSIM Statistical Classification Model')

Level	○ 계층적 노드세트 내에서, 최상위(root) 노드로부터 동일한 깊이(부모-자식 step)에 위치하면서, 하나의 통합 개념에 의해 정의되는 노드 집합
용어집_정의	○ set of Nodes in a hierarchical Node Set in which 1) each Node in the set is the same number of parent-child steps away from the root Node in the hierarchy, and 2) the set

	is defined by a unifying Concept 계층적 노드세트의 노드들의 집합은 다음과 같은 특징을 갖습니다. 1) 노드세트 내의 각 노드는 계층 구조의 최상위(root) 노드로부터 동일한 수의 부모-자식 단계(step)만큼 떨어져 있습니다. 2) 노드세트는 하나의 통합 개념에 의해 정의됩니다.
용어집_속성	○ Code Structure (0..1) (String) - 코드가 숫자, 문자 및 구분자를 사용하여 어떻게 구성되는지 명시 ○ Code Type (0..1) (ControlledVocabulary) - 해당 레벨의 항목 코드 유형을 명시: 문자형(Alphabetical), 숫자형(Numerical), 또는 혼합형(Alphanumeric) ○ Dummy Code (0..1) (String) - 연속된 상·하위 레벨이 하나 이상의 동일한 범주를 사용하는 경우, 상위 코드로부터 더미 코드를 생성하는 규칙 ○ Items (1..*) (MultilingualText) - 레벨을 구성하는 범주(분류항목)의 정렬된 목록 ○ Level Number (0..1) (Number) - 레벨에 부여된 번호로, 최상위(집계수준이 가장 높은) 레벨을 1로 시작하여 순차적으로 부여

Map	○ 원천 통계분류의 분류항목과 대상 통계분류의 해당 분류항목 간의 관계를 나타내는 표현
용어집_정의	○ expression of the relation between a Classification Item in a source Statistical Classification and a corresponding Classification Item in the target Statistical Classification
용어집_설명	○ 맵은 두 분류항목 간의 관계가 부분적인지 완전한지 명시해야 함 ○ 대응표의 관계 유형에 따라, 하나의 원본 항목 또는 대상 항목에 대한 여러 개의 맵이 존재할 수 있음 ○ 대응표와 맵의 사용은 모든 유형의 노드와 노드세트로 확장될 수 있음. 즉, 대응표를 통해 통계분류뿐 아니라, 코드목록, 범주세트의 항목 간에도 매핑이 가능함

Correspondence Table	○ 두 통계분류의 분류항목 간에 설정된 맵(Map)들의 집합
용어집_정의	○ set of Maps between the Classification Items of two Statistical Classifications
용어집_설명	○ 일반적으로 다음과 같은 경우가 이에 해당합니다. - 동일한 분류 시리즈 내의 두 버전 간 - 서로 다른 분류 시리즈의 통계분류 간 - 변형과 해당 변형의 기반이 되는 버전 간 - 특정 변형의 서로 다른 버전 간 ○ 첫 번째와 마지막 예에서 대응표는 시간 경과에 따른 비교를 용이하게 함 ○ 대응 관계는 양방향으로 표시함 ○ 통계분류는 노드 집합의 하위 유형으로, 통계분류와 대응표 간의 관계는 코드목록 및 범주집합과 같은 다른 노드세트에도 확장 적용할 수 있음
용어집_속성	○ Floating (0..1) (String): 대응표의 원천 또는 대상 통계분류가 유동 분류인 경우, 대응표의 날짜를 명시 ○ Relationship Type (0..1) (String): 원천과 대상 항목간 관계 유형 (1:1, 1:N, N:1 또는 M:N 등) ○ Source Level (0..1) (String): 대응 관계가 적용되는 원천 통계분류의 레벨을 지정. 만약, 지정되지 않으면 원천 통계분류의 모든 레벨이 매핑 대상이 될 수 있음 ○ Target Level (0..*) (String): 대응 관계가 적용되는 대상 통계분류의 레벨을 지정. 만약,

	지정되지 않으면 대상 통계분류의 모든 레벨이 매핑 대상이 될 수 있음
이용자안내	81. 대응표(Correspondence Table)는 매핑(Maps)의 집합이다. 이러한 매핑들은 두 분류항목에 공통이 되는 개념을 통해 한 통계분류의 분류항목을 다른 통계분류의 해당 분류항목과 연결한다. 예를 들어, ISIC rev 4와 북미산업분류체계(NAICS 2007(US)) 간의 관계를 나타내는 대응표에서, ISIC rev 4의 '0112 벼 재배(Growing of Rice)'는 NAICS의 '111160 벼 농사(Rice Farming)'와 '벼 재배(growing rice)'라는 공통 개념을 통해 매핑된다.

Classification Index Entry	○ 객체/단위의 유형 또는 객체 속성을 설명하는 단어 또는 짧은 텍스트(예. 지역명, 경제활동명, 직업명 등)와 해당 분류 항목의 코드와 함께 기술한 것
용어집_정의	○ word or a short text (e.g. the name of a locality, an economic activity or an occupational title) describing a type of object/unit or object property to which a Classification Item _applies, together with the code of the corresponding Classification Item_
용어집_속성	○ Coding Instructions (0..*) (String): 코딩 처리를 위해 필요한 추가 정보 (코딩이 하나 이상의 다른 요소에 의존하는 경우에 필요) ○ Text (1..*) (String): 객체/단위의 유형 또는 객체 속성을 설명하는 텍스트
용어집_설명	○ 분류항목은 노드의 하위 유형으로, 분류항목과 분류 색인 항목 간의 관계는 코드항목 및 범주항목과 같은 다른 노드 유형으로 확장될 수 있음 ○ 각 분류 색인 항목은 일반적으로 통계분류의 한 항목을 참조하는데, 통계분류의 모든 레벨의 분류항목과 연결될 수도 있지만 보통은 가장 낮은 레벨의 항목과 연결됨
이용자안내	83. 분류 색인 항목(Classification Index Entry)은 분류항목이 적용되는 개념의 유형을 설명하는 단어 또는 짧은 텍스트(예: 지역명, 경제 활동명 또는 직업)로, 해당 분류항목의 코드를 함께 나타낸다. 각 분류 색인 항목은 일반적으로 통계분류의 한 항목을 참조한다. 분류 색인 항목은 통계분류의 모든 레벨의 분류항목과 연관될 수 있지만, 보통은 최하위 레벨의 분류항목과 연관된다.

Classification Index	○ 분류 색인 항목들을 특정 순서(알파벳순, 코드순 등)로 정렬한 목록
용어집_정의	○ ordered list (e.g. alphabetical, in code order) of Classification Index Entries
용어집_속성	○ Coding Instructions (0..*) (String): 분류 색인의 모든 항목에 대한 공통적인 코딩 처리를 위해 필요한 추가 정보 ○ Corrections (0..1) (String): 분류 색인 내에서 발생한 수정사항에 대한 요약 설명으로, 분류 색인 항목에 연계된 항목 코드 변경사항 포함 ○ Languages Available (0..*) (String): 분류 색인에서 지원하는 언어. 언어별로 동일한 현상을 설명하는 단어(term)의 수가 다를 수 있으므로 언어별 항목(entry) 수는 다를 수 있으나, 다루는 현상의 범위는 같아야 함
이용자안내	82. 분류 색인(Classification Index)은 통계 자료원(설문조사 응답, 행정 기록 등)에서 발견되는 텍스트와 하나 이상의 통계분류 간의 관계를 보여준다. 분류 색인은 통계 프로그램의 관측치에 분류 항목 코드를 할당하는 데 사용될 수 있다.

Classification Series	○ 동일한 개념을 기반으로 하며, 버전이나 업데이트 관계를 통해 서로 연결된 하나 이상의 통계분류들의 집합
용어집_정의	○ ensemble of one or more Statistical Classifications, based on the same concept, and related to each other as versions or updates
용어집_속성	○ Context (0..1) (String): 분류 시리즈 설계 맥락(배경) ○ Keywords (0..*) (String): 분류 시리즈에 대한 핵심 키워드

	○ Objects/Units Classified (1..1) (String): 특정 속성에 따라 분류하고자 하는 객체/단위 유형 (저자 정리: 산업분류: 사업체, 직업분류: 개인, 질병분류: 질병, 행정구역코드: 행정동..) ○ Subject Areas (1..1) (String): 분류 시리즈가 적용되는 통계 분야
용어집_설명	○ 일반적으로 이러한 통계분류들은 동일한 이름을 사용 (예) ISIC 또는 ISCO
이용자안내	76. 분류 패밀리는 공통 개념(예: 경제 활동)에 기반한 관련 분류 시리즈 그룹입니다. - 분류 시리즈는 동일한 개념에 기반한 하나 이상의 통계분류의 집합입니다. 분류 시리즈 내의 통계분류는 버전 또는 업데이트 형태로 서로 연결됩니다. 일반적으로 이러한 통계분류는 동일한 이름을 갖습니다. (예) 국제표준산업분류(ISIC), 국제표준직업분류(ISCO)
웨비나	○ Ensemble of one or more Statistical Classifications, based on the same concept, and related to each other as versions or update ○ Relation between two ConceptSchemes (예) NACE Rev.2.1 <xkos:follows> NACE Rev.2 ○ Relation between a ConceptScheme and a Series item - The series are represented in the 'European Statistical Classification Series' dataset (예) NACE, CPA, PRODCOM - NACE Rev.2.1 <xkos:belongsTo> NACE - NACE Rev.2 <xkos:belongsTo> NACE

Classification Family	○ 공통 개념(예: 경제 활동)을 기반으로 한 분류 시리즈 그룹
용어집_정의	○ group of Classification Series based on a common Concept (e.g. economic activity)
용어집_설명	○ 분류 패밀리에 대한 국제 표준이 확립되지 않아, 분류 데이터베이스마다 분류 패밀리의 유형 및 이름이 서로 다를 수 있음
웨비나	○ 공통된 개념이나 도메인을 공유하는 관련 분류들의 그룹 (A group of related classifications sharing a common concept or domain) (예) ISIC(국제표준산업분류), NACE(유럽연합경제활동분류), NAICS(북미산업분류체계) → economic activities ○ ESS(유럽통계시스템) Classifications are linked to the EuroVoc thesaurus(유럽연합 다국어·다분야 유의어사전) concepts with <xkos:covers> ○ Recommended to use CSA (Classification of Statistical Activities) - not yet available in RDF format

Subject Field	○ 일련의 개념(Concepts)과 그 명칭(Designations)이 사용되는 지식 분야
용어집_정의	○ field of knowledge under which a set of Concepts and their Designations is used
용어집_설명	○ 유의어: subject area, theme ○ (예) 노동 시장, 환경 지출, 관광 등
이용자안내	59. 주제영역은 특정 지식 분야(예. 노동시장, 관광)에 기반한 개념체계를 그룹화합니다. A Subject Field groups Concept Systems based on their field of special knowledge (e.g., labour market, tourism).
ISO 11179-1	3.2.2 domain (유의어: subject field) - 특정 지식 영역 (field of special knowledge) - subject field(원문:domain)가 세분화될 경우, 세분화된 결과 역시 각각 하나의 subject field(원문:domain)가 된다. (Note 1 to entry: The borderlines and the granularity of a domain are determined from a purpose-related point of view. If a domain is

	sub-divided, the result is again a domain.) [SOURCE: ISO 1087:2019, 3.1.4]
--	---

Concept System	○ 개념들 간의 관계에 의해 구조화된 개념들의 집합
용어집_정의	○ set of Concepts structured by the relations among them
용어집_설명	○ (예1) 성별 개념: 남성, 여성, 기타 ○ (예2) ISIC (목록이 너무 길어서 모두 적지 못함)
이용자안내	59. (용어집 정의와 동일)

Data Point	○ 인스턴스 변수의 단일 값을 수용하는 데이터세트 내의 최소 구조 단위 - 테이블의 셀에 해당하는 구조적인 관점이므로, 저장되는 값(Datum) 없이 비어 있을 수도 있음
용어집_정의	○ container for a single value of an Instance Variable
용어집_설명	○ 데이터 포인트는 값을 저장할 수 있는 셀 또는 자리 표시자임(데이터 포인트는 비어 있을 수도 있음). ○ 데이터 구조의 한 필드(field)로서, 예를 들어 테이블의 셀에 해당함 ○ 데이터 포인트는 구조적인 관점이므로, 저장되는 값(Datum)과는 구별됨
이용자안내	87. 데이터세트(Data Set)은 데이터 포인트(Data Points)를 가진다. - 데이터 포인트는 데이터세트에서 데이터값(Datum)을 입력받을 수 있는 자리표시자(예. 테이블의 빈 셀)임

Data Set	○ 조직화된 데이터 집합
용어집_정의	○ organised collection of data
용어집_속성	○ Type (0..1) (ControlledVocabulary): 데이터세트 유형 (예: 단위 데이터세트, 차원 데이터세트)
용어집_설명	○ 유의어: Database, data file, file, table ○ 데이터세트는 관측 등록부(observation registers), 시계열 데이터, 종단 데이터, 설문조사 데이터, 직사각형 데이터세트, 사건 이력 데이터, 테이블, 데이터 테이블, 레지스터, 데이터 큐브, 데이터 웨어하우스/마트 및 행렬을 포함하여 매우 광범위하고 다양한 콘텐츠를 구성하는 데 사용할 수 있습니다. - 개인의 사회 보장 번호, 나이, 시민권 및 출생 국가를 포함하는 인구 단위 데이터세트 (마이크로데이터)의 예: (1212123, 48, American, United Kingdom), (1212111, 38, Hungarian, United Kingdom), (1212317, 51, Canadian, Mexico) - 국가 이름, 해당 연도 및 해당 국가의 인구(백만 단위)를 포함하는 인구 차원 데이터세트 (집계)의 예: (멕시코, 2021년, 130.3), (영국, 2021년, 67.33), (이탈리아, 2022년, 60.24)
이용자안내	67. 인스턴스 변수는 데이터 포인트를 통해 데이터세트와 연결된다. 87. 데이터세트(Data Set)은 데이터 포인트(Data Points)를 가진다(has). 88. 데이터세트는 등록부(Register)의 마이크로 수준 데이터, 시계열 데이터, 패널 데이터, 생존분석 데이터 등 다양한 형태로 존재함 - 데이터세트의 유형에 따라, 정의되어야 할 속성, 필요한 데이터구조 유형(단위/차원), 가용한 분석방법이 결정됨

	89. 데이터세트는 관점에 따라 단위 데이터, 차원 데이터가 될 수 있음 - 예를 들어, 국가 통계청에서는 차원 데이터로 간주하는 것을 Eurostat에서는 단위 데이터로 간주할 수 있음 - 이렇듯 데이터를 보는 관점은 개념적인 것이며, 특정 데이터 집합을 본질적으로 둘 중 하나로 간주할 필요는 없음 90. GSIM에서는 모든 데이터세트는 구조가 있어야 한다고 명시하고 있으나, 실제로는 구조를 알 수 없는 데이터(비정형데이터, 구조정보가 없는 데이터 등)가 존재할 수 있으며 앞으로 더 빈번하게 나타날 수도 있음 - 이러한 데이터를 활용하기 위해, 사용 시점에 어떤 형태로든 데이터구조를 정의되어야 함. 이를 위해 통계기관에서는 정보 세트(Information Set) 클래스의 새로운 하위 유형을 만드는 방식으로 GSIM을 확장할 수 있음
--	---

Statistical Programme	○ 주제영역의 맥락 내에서 특정 유니버스에 대한 통계 생산활동 집합 - GSBPM 4~7단계(수집/처리/분석/공표)에서 다루는 지속적인 활동에 해당 - 즉, 설계 및 지원(메타데이터 및 데이터 관리) 기능은 미포함
용어집_정의	○ set of activities to produce statistics on a given Universe within the context of Subject Fields
용어집_속성	○ Budget (0..1) (Number): 예산 소요액 ○ Date Ended (0..1) (Date): 통계 프로그램 종료일 ○ Date Initiated (0..1) (Date): 통계 프로그램 시작일 ○ Legal Framework (0..*) (String): 법적 근거 (예: 통계 프로그램에서 생산하는 통계에 대한 법적 근거) ○ Legislative Reference (0..*) (String): 입법 참고자료 (예: 국회 제출 문서) ○ Source of Funding (0..1) (String): 예산 출처 ○ Programme Status (1..1) (ControlledVocabulary): 프로그램 상태 (예: 신규 제안, 개발 중, 진행, 완료, 취소, 다른 기관으로 이전)
용어집_설명	○ 통계 기관에서 통계 생산활동이 수행되는 환경적 맥락을 제공합니다. 일반적으로 GSBPM 4~7단계에서 다루는 조사 또는 산출 시리즈와 같은 지속적인 활동을 의미함 - (예) 노동력 조사, 다목적 가구조사, 국민계정, 인구통계, 출입국통계 등 - 통계 프로그램 설계 및 통계 프로그램 주기 클래스와 연결되어 비즈니스 프로세스의 상세 설계 및 수행 정보를 보유함 - 다른 통계 프로그램의 산출물(예: 국민 계정)을 입력값(inputs)으로 사용할 수 있음. - 이러한 활동들은 모두 '산출물(Products)'을 생성하기 위해 수행됨 - 전통적인 접근 방식의 경우, 기관은 통계 수요를 접수하고 변경 정의서와 승인된 비즈니스 제안서(business case)를 작성함. 비즈니스 제안서에는 기존 통계 프로그램의 설계 또는 방법론에 대한 변경(이로 인해 새로운 통계 프로그램 설계로 이어짐), 하나 이상의 기존 통계 프로그램 변경(예: 통계 프로그램에 추가적인 목적 추가), 또는 새로운 통계 프로그램 생성 중 하나를 명시하게 됨 - 단, 메타데이터 관리, 데이터 관리(및 기타 전반적인 GSBPM 프로세스)와 같은 통계지원 기능이나 설계 기능은 포함하지 않음. 이러한 활동은 통계 지원 활동의 일환으로 수행됨
이용자안내	16. 통계 프로그램(statistical programme)이란 통계 생산을 위해 조직이 수행하는 일련의 프로그램들(예: 소매업조사, 노동력조사, 인구주택총조사 등)을 의미합니다. - 통계 프로그램은 대개 주기적으로 운영되며, 수집, 처리, 분석, 공표의 과정을 거치는 주기(cycle)를 반복합니다. 이러한 각 주기는 통계 프로그램 주기(Statistical Programme Cycle) 클래스로 표현됩니다. 통계 프로그램 주기는 특정 시점의 통계(예: 2022년 3월 소매업조사)를 생산하기 위해 반복되는 활동을 의미합니다.

Statistical	○ 특정 모집단(Population)을 대상으로 수행되는 통계 프로그램의 반복 회차
-------------	--

Programme Cycle	
용어집_정의	○ iteration of a Statistical Programme for a given Population
용어집_속성	○ Question Purpose (0..1) (MultilingualText): 질문의 목적에 대한 설명, 즉 질문이 특정한 기능을 수행할 것으로 예상되는지 여부 ○ Question Text (1..1) (MultilingualText): 언어야 할 정보를 설명하는 텍스트
용어집_설명	○ 특정 모집단(예: 특정 참조 기간, 지역)에 대하여, 관련 통계 프로그램 설계에 따라 통계 프로그램의 반복 실행 과정을 문서화합니다. 이 문서에는 주기에서 수행된 활동, 필요한 자원, 사용되는 프로세스, 그리고 통계 프로그램 설계에서 정의된 해당 주기 관련 방법론적 정보에 대한 설명이 포함됩니다.
이용자안내	16. 통계 프로그램(statistical programme)이란 통계 생산을 위해 조직이 수행하는 일련의 프로그램들(예: 소매업조사, 노동력조사, 인구주택총조사 등)을 의미합니다. - 통계 프로그램은 대개 주기적으로 운영되며, 수집, 처리, 분석, 공표의 과정을 거치는 주기(cycle)를 반복합니다. 이러한 각 주기는 통계 프로그램 주기(Statistical Programme Cycle) 클래스로 표현됩니다. 통계 프로그램 주기는 특정 시점의 통계(예: 2022년 3월 소매업조사)를 생산하기 위해 반복되는 활동을 의미합니다.

Question	○ 개념변수에 대한 응답을 이끌어내기 위해 사용되는 텍스트 - 응답값을 규정하기 위해 값 도메인과 연결됨 - 질문의 형태는 1차원(단일 질문) 혹은 2차원(복합 질문)으로 구성할 수 있음
용어집_정의	○ text used to elicit a response for a Conceptual Variable
용어집_속성	○ Reference Period End (1..1) (Date): 참조 기간의 종료일 ○ Reference Period Start (1..1) (Date): 참조 기간의 시작일
용어집_설명	○ 단일(single) 질문뿐 아니라 각각의 응답을 가진 하위 질문들이 연결된 복합(multiple) 질문일 수 있음 ○ 질문에 대한 응답 기준(criteria)을 규정하기 위해 값 도메인과 관련성이 있음 - 단일 질문은 하나의 값 도메인과 연관되나, 복합 질문은 하나 이상의 값 도메인을 갖게 됨 ○ 질문은 다양한 조사표에서 사용할 수 있도록 재사용성을 염두에 두고 설계해야 함 ○ 국가별 구현 사례에서, 질문 형태에 따른 세분화 유형: - 표 형태(QuestionGrid): 질문을 격자/표 형태로 모델링하는데 유용함. 실제로는 차원 정보, 라벨링 옵션, 그리고 그리드 내 하나 이상의 셀에 연결된 응답 영역을 제공하는 큐브 형태의 구조임. (예) 계열사별 매출액들을 입력받기 위한 2차원 분할표 - 항목 형태(QuestionItem): 1차원적인 단순한 질문 형태임. (예) '귀하의 연령은 몇 세입니까?'□
이용자안내	(부록2) ○ 질문은 복합 질문의 형태를 가질 수 있으며, 계층적일 수 있음. - 질문은 하나 이상의 개념변수와 연결되며, 해당 개념변수에 대한 질문의 응답값으로 할당할 수 있는 값을 제한하기 위해 값 도메인과 연관됨

Instance Question	○ 특정 조사표에서 질문을 사용하는 것 (즉, 추상적 질문을 실체화한 것)
용어집_정의	○ use of a Question in a particular Questionnaire
용어집_속성	○ Question Purpose (0..1) (MultilingualText): 질문의 목적에 대한 설명 ○ Question Text (1..1) (MultilingualText): 언어야 할 정보를 설명하는 텍스트
용어집_설명	○ 특정 조사표 구성요소(Questionnaire Component)에서 실제로 질문을 사용하는 것 - 조사표 구성요소의 특정 유형인 질문 블록(Question Block)에서 질문을 사용하는 것을 포함함

Questionnaire	○ 관찰 단위들로부터 정보를 이끌어내기 위한 '교환 수단'
용어집_정의	○ Exchange Instrument to elicit information from observation Units

용어집_속성	- Media (1..1) (String): 설문조사에 사용될 매체의 종류에 대한 설명 (인쇄물, 전자 매체 등) - Support Artifacts (0..*) (String): 설문지 사용을 지원하는 데 필요한 장치, 소프트웨어 프로그램, 저장 매체, 기기 또는 기타 도구 목록입니다. - Survey (0..*) (String): 설문지가 사용될 설문조사에 대한 정보입니다.
용어집_설명	- 조사표는 통계 기관이 정보를 수집하는 한 가지 방법(교환 도구)의 예입니다. - 각 수집 방식(예: 대면 조사, CAPI(컴퓨터 지원 개인 면접), 온라인 설문조사)은 조사표 사양으로부터 파생된 새로운 조사표로 해석되어야 함
이용자안내	(부록2) - 조사표는 사용될 조사의 목적에 맞게 하나 이상의 일반 유형으로 개발될 수 있음 - 각 개별(instance) 조사표는 조사표 명세를 참조하여 구성됨 - 특정 조사에 대한 표준 조사표 사양(레이아웃이 동일하거나 변경이 비교적 적은 형태)을 따를 수도 있고, 반대로 설문조사에 선택된 각 정보제공자에 맞게 조사표 사양을 맞춤화할 수도 있음

Exchange Instrument	정보를 교환하는 구체적이고 사용할 수 있는 도구
용어집_정의	concrete and usable tool to exchange information
용어집_속성	Direction (1..1) (String): 교환 수단의 방향: 수집 또는 배포
용어집_설명	- 교환 수단은 정보를 주고받는 도구로, 외부 및 내부 목적으로 사용됨 - 데이터 수집과 배포를 위해 다양한 교환 수단이 사용됨 - 정보 수신을 위한 교환 수단의 예 → 조사표(Questionnaire) - 정보 전송을 위한 교환 수단의 예 → 배포 수단(Dissemination Instrument) - 각 기관의 필요에 따라 모델에 추가적인 교환 수단을 추가할 수 있음
이용자안내	38. 통계기관이 데이터나 메타데이터를 정보제공자(예. 응답자, 행정자료 제공자)로부터 수집하고, 정보소비자(예. 정부, 기업, 국민)에게 보급할 때 교환 수단이 사용됨 39. 교환 수단은 데이터와 메타데이터가 교환되는 구체적이고 사용할 수 있는 도구임 - 유입 데이터를 설명하는 하위 클래스: Data Harvest, Questionnaire - 유출 데이터를 설명하는 하위 클래스: Dissemination Instrument (Product 형태) 40. 필요에 따라 교환 수단의 하위 클래스는 얼마든지 확장 가능 - 정보 교환은 기관 외부뿐 아니라 내부에서도 발생 (예) 행정자료 수집부서에서 구축한 API를 다른 부서가 활용하여 자료를 수집하는 것도 교환 수단의 한 형태임 41. 모든 교환 수단은 구축 및 갱신 전에 반드시 설계 단계를 거쳐야 함 - 이 설계에 대한 개요와 상세 설명은 교환 사양(Exchange Specification)으로 표현됨 - 여기에는 SDMX, 웹 서비스, 또는 대면 면접과 같이 정보 교환에 대한 통신 규약(Protocol) 정보가 포함될 수 있음 42. 교환은 제공 협약(Provision Agreement)에 명시된 방식에 따라 일관되게 수행됨 - 자료 수집시: 정보제공자로부터 데이터 및 참조 메타데이터를 받아서, 교환 수단을 통해 정보 세트(Information Set)를 생성함 - 자료 배포시: 정보소비자는 정보 세트를 소비함

Data Harvest	두 소스 간에 정보를 전달하는 교환 수단으로, 일반적으로 기계 대 기계 메커니즘을 통해 이루어짐
용어집_정의	Exchange Instrument to pass information between two sources, usually by a machine to machine mechanism
용어집_설명	(예) 웹 스크래퍼, API(예: 관리 기관으로부터 데이터를 가져오기 위한 API), 스캐너, 센서 등

Dissemination Instrument	○ 정보를 배포하는 교환 수단
용어집_정의	○ Exchange Instrument to disseminate information
용어집_설명	○ (예) 데이터 배포를 위한 API 또는 웹 서비스 등

Product	○ 대외적으로 공표할 수 있는 콘텐츠 패키지 - 출력 사용에 정의된 방식으로 제공됨(인쇄물과 같은 고정 사양 혹은 사용자가 실시간으로 설정할 수 있는 동적 사양) - 제공 협약을 근거로 정보소비자에게 전달되며, 제공 협약은 통계기관이 게시한 이용약관에 따른 목시적인 경우가 대부분임
용어집_정의	○ package of content that can be disseminated as a whole
용어집_설명	○ 산출물은 출력 사양에 정의된 방식대로 정보 세트를 결합하여 만든 최종 통계 결과물임 ○ 출력 사양은 정보 세트의 정보가 정보소비자에게 어떻게 제시되는지(프레젠테이션을 통해)를 정의함 - 인쇄물과 같은 고정된 사양뿐 아니라, 정보소비자가 직접 실시간으로 결정할 수 있는 동적 사양도 가능함 - 두 경우 모두, 출력 사양(Output Specifications)에는 산출물에 포함될 정보 세트 데이터 또는 참조 메타데이터를 명시함 ○ 통계 기관과 정보소비자 간의 제공 협약에 따라 정보소비자가 산출물을 이용함 사용을 규율함 - 제공 협약은 명시적 또는 묵시적일 수 있으며, 이는 양 당사자가 데이터를 교환하기로 합의하는 법적 또는 기타 근거를 제공함 - 많은 경우, 배포 제공 협약은 통계 기관이 게시한 이용약관에 묵시적으로 포함됨

Presentation	○ 데이터와 참조 메타데이터가 제시되는 방식
용어집_정의	○ way data and referential metadata are presented
용어집_설명	○ 정보 세트로부터 데이터와 참조 메타데이터를 제시함 - 출력 사양에서 프레젠테이션을 사용하여 산출물의 정보 세트를 어떻게 제시할지 지정할 수 있음 ○ 표, 그래프, 구조화된 데이터 파일 등 다양한 형태로 제공될 수 있음. 예를 들어, - 통계표: 데이터 구조(Data Structure)를 사용하여 행/열 제목 지정, 데이터세트로 셀을 채우고, 참조 메타데이터로 각주 생성 - 표준 데이터 파일 (예. SDMX) - 통계분류를 설명하는 PDF 문서 - 표준 형식으로 표현된 모든 구조화된 메타데이터 객체 (예. DDI 3.1 XML) - 산출물 또는 서비스 목록 (예: 제품 카탈로그 또는 WSDL(웹 서비스 설명 언어) 파일) - 통계분류, 개념변수 설명 등이 포함된 웹 페이지 등

출처: (용어집_정의) UNECE, GSIM Version 2.0 - 클래스 정의(Definition) [웹사이트];
(용어집_속성) UNECE, GSIM Version 2.0 - 클래스 정의(Attributes) [웹사이트];
(용어집_설명) UNECE, GSIM Version 2.0 - 클래스 정의(Explanatory Text) [웹사이트];
(이용자안내) UNECE(2024), GSIM 2.0 User Guide;
(웨비나) UNECE Statistics, GSIM 2.0 Webinar 2025 [YouTube] (2025.9. 게시);
(DDI-CDI) DDI Alliance(2025), DDI-CDI Specification Overview;
(ISO 11179-1) ISO/IEC(2023), ISO/IEC 11179-1:2023;
(OECD 용어집) OECD(2008), OECD Glossary of Statistical Terms;
(SDMX 용어집(2.1)) SDMX(2020), SDMX Glossary;
(GSIM 1.1 명세) UNECE(2013), GSIM Specification v1.1;
(GSIM 1.2 용어집) UNECE(2019), GSIM Glossary v1.2;
(GSIM e-training) UNECE(2019), GSIM Concept Group, GSIM e-training;
(HLG-MOS(2026)) UNECE HLG-MOS Supporting Standards Group(2026), Business Case for Sentinel Values;
(MCV(2009) Unit 정의) UNECE Statswiki(2009), Statistical Unit Definitions in the MCV.

Abstract**A Study on the Use of GSIM for the Systematic Management of Concepts in National Statistics****Neung-kyung Kim, Mi-kyung Bae**

To ensure the appropriate use of statistics, systematic conceptual information explaining the population, criteria, and methods used in statistical production is essential. This study proposed a management framework for conceptual information of national statistics by reviewing the Generic Statistical Information Model (GSIM), a reference framework for statistical information, and analyzing the practices of Statistics Canada and Statistics Norway.

GSIM covers the whole statistical production process and organizes its information classes into five groups: Concepts, Structures, Business, Exchange, and Base. Its Variable Cascade separates design from implementation, facilitating the reuse of common information and ensuring consistency across statistics. The two countries differed in their approaches to GSIM implementation. Canada emphasized the establishment and reuse of standards through its "Policy on Standards," whereas Norway managed variables, statistical classifications, and code lists in function-specific systems and linked them to the statistical tables. However, in both countries, descriptions of Conceptual Variables were found to include not only abstract definitions but also specific information, such as determination criteria and derivation methods.

For Korea, this study proposes Unit Type, Variable, and Code List as key management targets and recommends managing Variables at two levels: standard variables and programme-specific variables. It also proposes a delineation of roles between standards management divisions and statistical programme divisions, along with procedures for the development, revision, and approval of standards.

As this study was applied to selected statistical domains, further research is needed to expand its application to a wider range of statistics and verify the effectiveness of the proposed framework. In addition, this study proposes first structuring concepts, variables, and code lists for each statistical programme and then gradually standardizing them, starting with those with high levels of

commonality and utility. This study is expected to enhance the comparability and reusability of statistics and contribute to laying the foundation for data integration in the future.

As this study was applied to selected statistical domains, further research is needed to expand its application to a wider range of statistics and verify the effectiveness of the proposed framework. In addition, an approach could be considered that first structures concepts, variables, and code lists for each statistical programme and then gradually standardizes those with high levels of commonality and utility. The GSIM-based management framework is expected to enhance the comparability and reusability of statistics and serve as a foundation for the integration of statistics and data.

Key words: GSIM, Unit Type, Conceptual Variable, Code List

연구진

- 김능경(국가데이터처 국가데이터연구원 데이터방법연구실 사무관)
- 배미경(국가데이터처 국가데이터연구원 데이터방법연구실 주무관)

* 연구진의 소속 및 직급은 연구과제 완료 시 기준임을 알려드립니다.

연구보고서 2026-01

국가통계 개념의 체계적 관리를 위한 GSIM 활용 방안 연구

인 쇄	2026년 9월
발 행	2026년 9월
발 행 인	백지선
발 행 처	국가데이터처 국가데이터연구원 35220 대전광역시 서구 한밭대로 713 Tel.(042)366-7100 Fax.(042)366-7123
홈페이지	https://mods.go.kr/dsri/
ISSN(Online)	2733-4120



국가데이터처
국가데이터연구원