DSRI
Data
Research
Brief

vol. 2026-3

ISSN 2765-1002(Online)



국가통계 개념의 체계적 관리를 위한 GSIM 활용 방안

김능경
경제동향통계심의관
물가동향과

배미경
국가데이터연구원
데이터방법연구실

- **(연구 목적)** 통계 간 개념의 공통점과 차이를 식별·비교할 수 있는 일관된 통계 정보 제공을 위해 UNECE 승인 통계정보 참조모델인 **GSIM(Generic Statistical Information Model) 2.0**과 캐나다·노르웨이의 적용 사례를 분석하여 **국가통계 개념정보 관리체계 방안**을 제시하고자 한다.
- **(핵심 결과)** GSIM은 **설계와 구현을 분리**함으로써 재사용성과 일관성을 확보한다. 캐나다는 **표준 정책 기반**의 중앙 관리, 노르웨이는 **기능별 시스템** 연계 방식이며, 양국 모두 개념변수에 판정기준, 도출방법 등 구체적인 사항을 포함하고 있다.
- **(제안 사항)** **Unit Type**(단위유형), **Variable**(변수), **Code List**(코드목록)를 주요 관리대상으로 설정하고 표준 변수와 통계 프로그램별 변수로 **변수 관리를 이원화**하되 통계 프로그램별 정보 구축 후 공통성이 높은 정보부터 표준화하는 단계적 접근을 제안한다.
- **(향후 활용)** 향후 다양한 통계로 **실증 범위를 확대**하여 실무 적용을 위한 최적안을 마련하고 표준의 제정·개정·승인 등 **표준 관리 절차를 확립**함으로써 통계 간 비교 가능성과 재사용성을 획기적으로 향상시키는 기반이 되기를 기대한다.

연구배경 및 목적

❍ **(연구 배경)** 같은 명칭의 개념이라도 조사대상, 측정 단위·방법, 범주 구성에 따라 의미가 다르고, 유사한 개념이 서로 다른 명칭으로 제공되기도 함(예: 노동시장 규모 관련 통계표의 '취업자', '종사자', '일자리')

❍ **(문제 인식)** 통계설명자료서비스의 용어 조회는 상세 작성 기준 없이 기관별로 작성되어 통계 간 비교에 한계가 있으며, 통계용어사전은 단위유형 등 정보의 구조화와 관련 통계와의 연계가 부족함

❍ **(연구 필요성)** 이용자의 통계 선별·해석, 중복 생산 방지와 통계 간 연계는 물론 생성형 AI의 기계 가독성 측면에서도 체계적인 통계 개념 정보가 필요함

❍ **(연구 목표)** GSIM과 이를 도입한 해외 사례를 분석하여 국가통계 전반에 적용할 수 있는 개념 정의·기술 체계의 공통 기준을 제시하고자 함

☞ **(연구 질문) 국가통계 개념정보를 어떤 대상과 구조로, 어떤 절차로 관리해야 통계 간 비교가능성과 재사용성을 높일 수 있는가?**

데이터 및 방법

❍ **(국제표준 검토)** GSIM 2.0에 대한 다양한 문서(클래스 정의, 그룹 다이어그램, User Guide, Linking GSBPM and GSIM, DDI-CDI Specification, ISO/IEC 11179-1 등

❍ **(해외 사례)** 캐나다 통계청과 노르웨이 통계청의 표준관리 정책, 메타데이터 조회 시스템을 통한 개념정보 관리체계와 정보 클래스별 관리 현황 비교

❍ **(국내 시범 적용)** 통계정책관리시스템, 통계설명자료서비스, 국가통계포털 자료로 현황을 진단하고, 노동 분야를 중심으로 GSIM 정보 클래스를 시범 적용

주요 결과

❍ **GSIM의 구조: 5개 그룹과 변수 연쇄(Variable Cascade)**

- GSIM은 통계 생산 전 과정을 포괄하며, 역할에 따라 정보 클래스를 5개 그룹으로 구분하여 기초 그룹은 기술적 정보, 비즈니스 그룹은 업무 프로세스, 교환 그룹은 수집·배포, 구조 그룹은 데이터·메타데이터 구조화, 개념 그룹은 데이터 의미·표현에 대한 정보를 담고 있음

<표1> GSIM 그룹별 정보 클래스

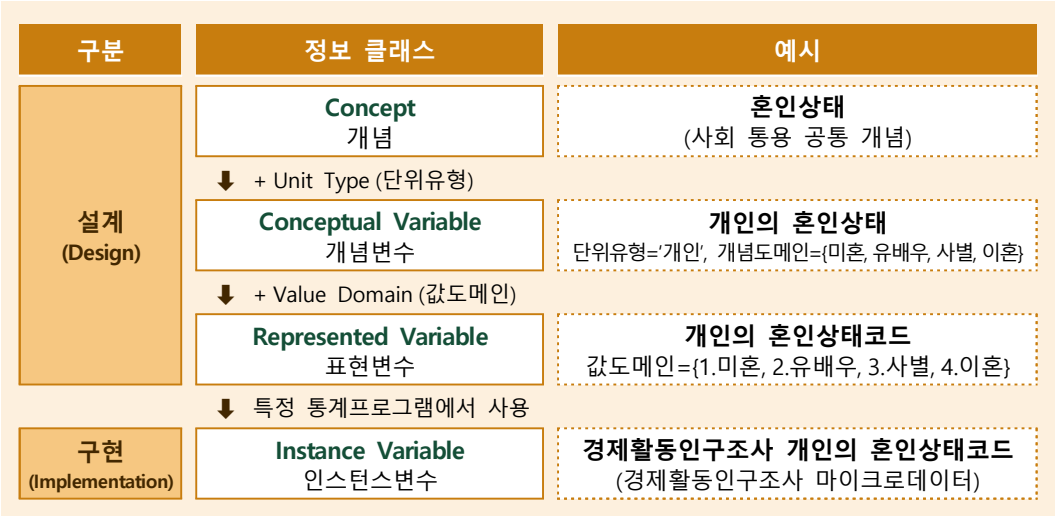
그룹	역할	주요 정보 클래스	개수*
기초(Base)	식별, 버전 관리 등	주체, 역할, 이력, 식별가능 아티팩트 등	12개
비즈니스(Business)	통계 프로그램 설계, 수행	통계 프로그램, 업무 프로세스 등	32개
교환(Exchange)	데이터 수집 및 배포	조사표, 산출물, 정보 이용자 등	21개
구조(Structures)	데이터 구성 방법	데이터세트 구조, 구성요소 등	23개
개념(Concepts)	데이터의 의미 및 표현	개념, 변수, 단위유형, 분류체계 등	39개

* Sentinel Value Domain 등 User Guide에서 설명하지 않는 일부 정보 클래스를 모두 포함한 개수임
출처: UNECE, GSIM Version v2.0 (December 2023), https://unece.github.io/GSIM-2.0

- 개념 그룹의 Concept는 ❶ 측정할 특성(개념변수), ❷ 대상이 되는 집합(단위유형·유니버스·모집단), ❸ 외연적 정의를 위한 범주(통계분류·코드목록)의 세 가지 역할을 수행함

- 추상적 개념이 개념변수(개념+단위유형), 표현변수(개념변수+값도메인)를 거쳐 개별 인스턴스변수로 구체화되는 변수 연쇄(Variable Cascade) 구조를 통해 설계와 구현을 분리함으로써 공통 정보의 재사용성과 조사별 차별성을 동시에 확보함

[그림1] 변수 연쇄(Variable Cascade) 구조와 적용 예시



출처: UNECE(2024), Generic Statistical Information Model (GSIM): Version 2.0 User Guide

주요 결과 (계속)

❍ 해외 사례: 중앙 표준 관리(캐나다)와 기능별 시스템(노르웨이)

- 두 기관 모두 GSIM을 활용하여 개념, 통계단위, 변수, 분류 및 코드목록을 체계적으로 관리하나, 캐나다는 표준의 제정과 재사용에, 노르웨이는 기능별 시스템 구현 및 연계에 중점을 둠

- (표준 의무 수준) 캐나다는 표준을 '부처 표준', '권고 표준', '프로그램별 표준'으로 구분하여 표준화의 강제력과 유연성을 함께 확보함

- (코드목록 제공 방식) 노르웨이는 전체 코드 구조와 버전·대응표·변형을 구조적으로 제공하지만 범주 정의는 없고, 캐나다는 범주 정의와 용법을 제공하지만 전체 구조 조회에는 제약이 있어 상호 보완적 고려가 필요함

- (GSIM 현지화) 두 기관 모두 변수의 개념도메인(Conceptual Domain) 정보는 제공하지 않으며, 캐나다는 하나의 표준 변수에 여러 분류를 연계하는 등 GSIM을 실제 통계작성 환경에 맞게 변형하여 적용하고 있음

<표2> 캐나다·노르웨이 통계 개념정보 관리체계 비교

구분	캐나다 통계청	노르웨이 통계청
관계 체계	표준정책 및 중앙 메타데이터 저장소 중심	기능별 메타데이터 시스템 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
Unit Type	Statistical Unit 33개 - 맥락이 부여된 Universe까지 포괄	Unit Type 31개 - GSIM의 Unit Type 개념을 충실히 적용
Variable	표준 변수 162개 - 표준 변수와 통계 프로그램별 변수 별도 관리 - Classification의 범주는 불포함하기도 함	Conceptual Variable 1,193개 - 표준 여부 미구분, 범주 여부 무관 - 통계표 조회 시스템과 연계
Node Set (통계분류/ 코드목록)	Classification 226개 - 통계분류와 코드목록 미구분 - 범주 정의와 사용법 등을 제공	Statistical Classification 151개, Codelist 596개 - 통계분류와 코드목록 엄격히 구분 - 코드구조·버전·변경이력·대응표·변형 등 제공

❍ 해외 사례: 변수 정의에 구체적 내용까지 포함

- (취업자) GSIM상 개념변수는 추상성을 유지할 것으로 예상했으나, 양국 모두 '취업자' 변수에 수입 목적 근로, 무급가족종사, 질병·휴가 등에 따른 일시적 결근 등 판정기준을 구체적으로 명시함

* 캐나다는 정의(Definition)와 도출방법(Derivation)을 분리하여 세부 기준을 기술하고, 노르웨이는 '주 1시간 이상 소득창출활동', '3개월 이내 일시해고' 같은 수치 기준을 정의문에 직접 포함함

- (혼인상태) 캐나다는 사실혼 포함 여부에 따라 변수를 둘로 나누어 각각 표준으로 승인한 반면, 노르웨이는 법률혼 기준의 단일 변수에 2009년 동성혼 법제화 이전의 관계 유형 범주를 그대로 유지함

* 같은 상위 개념이라도 변수와 범주 구성은 이용 목적과 제도적 맥락에 따라 달라질 수 있음

주요 결과 (계속)

❍ 국내 시범 적용: 노동 분야

- SIM 정보 클래스를 국내 노동 분야에 시범 적용한 결과는 <표3>와 같음
- 개념은 추상적으로 정의되나, 개념변수는 단위유형을 특정하고 개념도메인 1개를 참조하여 구체화되며, GSIM 최상위 정보 클래스인 Identifiable Artefact의 Description 속성을 활용하여 조작적 정의에 해당하는 구체적인 설명을 추가하였음
- 표현변수는 개념도메인에서 표현방식이 추가된 값도메인을 참조하며, 유니버스는 정보 클래스의 자기참조 관계를 이용하여 조사 및 변수 수준으로 정의할 수 있음

<표3> 노동 분야 GSIM 정보 클래스 시범 적용 결과 (일부)

정보 클래스		사례1	사례2	사례3
주제		노동		
개념	명칭	Employment		
	정의	임금 또는 수익을 위해 재화를 생산하거나 서비스를 제공하는 활동에 참여		
개념변수	단위유형	개인	사업체	
	명칭	개인의 취업여부	사업체의 종사자 수	
	개념도메인	{취업, 미취업}	인원수	
	설명	수입목적1시간이상근로 또는 18시간이상 무급가족일 또는 일시휴직	소속외근로자(파견, 용역), 3개월 이상 장기 결근자 제외	자영업자, 무급가족종사자, 소속외근로자(파견, 용역) 제외
표현변수	값도메인	{1 취업, 2 미취업}	{0 이상의 정수}	{0 이상의 정수}
	측정단위	-	명	명
유니버스		만 15세이상 인구	모든 사업체	종사자 1인 이상을 고용하고 있는 모든 사업체
통계프로그램		경제활동인구조사, 지역별고용조사	전국사업체조사	사업체노동력조사

❍ 국내 활용 방안: 정보 클래스별 관리

- 관리 실익과 재사용성을 고려하여 Unit Type, Code List, Variable을 주요 표준 관리 대상으로 선정하고, Universe는 작성 지침 중심으로 관리할 것을 제안함

* Conceptual Domain은 개념변수당 1개만 참조해야 하는 제약관계로 인해 재사용성이 제한되므로 표현변수의 Value Domain으로 관리 갈음

<표4> 주요 정보 클래스별 국내 현황 및 관리 방안 제안

정보 클래스	국내 현황	관리 방안 제안
Unit Type	조사 수준에서만 관리, '기타' 36%	변수 수준으로 확장, 표준값 집합 마련
Universe	'작성 대상 범위'로 조사 수준에서 기술	표준값 대신 작성 원칙·지침 제공
Statistical Classification	통계법에 따라 체계적으로 제정·관리	현행 방식 유지
Code List	별도 관리체계 부재, 같은 범주도 통계별 명칭 상이	표준 관리대상으로 설정, 범주 정의·위계·버전·이력·대응표·변형 관리
Variable	작성 틀 부재로 통계 프로그램별 편차	표준 변수와 통계 프로그램별 변수 이원화

주요 결과 (계속)

- (Code List) '혼인상태' 범주와 같이 인구주택총조사 '배우자있음', 경제활동인구조사 '유배우', 생활시간조사 '배우자있음(동거포함)'처럼 동일 개념이면서 범주 명칭이 다른 경우, 표준 코드목록을 정의하고 각 통계가 표준과의 대응표를 작성하면 비교가능성 제고

- (Variable) 표준 변수는 공통 개념과 정의를, 통계 프로그램별 변수는 작성 목적과 맥락에 따른 구체적 정의와 용법을 이원화하여 관리하고, 코드목록, 관련 통계표 등은 링크를 통해 재사용성 및 정보 클래스 간 연계성 제고

<표5> 표준 변수와 통계 프로그램 변수에 대한 작성 정보(안)

구분	표준 변수	통계 프로그램 변수
정보 클래스	• 주제영역 | 개념 | 개념 정의 • 변수명 | 단위 유형 | 변수 설명 • 코드목록(Link)	• 주제영역 • 변수명 | 단위 유형 | 변수 설명 • 코드목록 | 표준 코드목록 대응표 | 측정단위
추가 정보	• 관련 통계 프로그램 변수(Links) • 관련 통계표(Links)	• 통계 프로그램명 • 관련 통계표(Links)

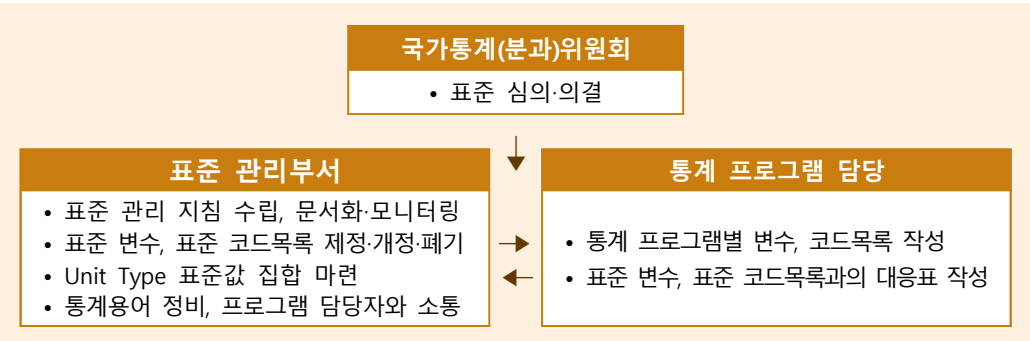
❍ 국내 활용 방안: 개념정보 관리 거버넌스

- (현황) 국내에는 국가통계 개념을 정의하는 틀이 일원화되어 있지 않고, 통계분류 외에는 개념을 공식적으로 표준화하고 관리·조정하는 절차가 마련되어 있지 않음

- (표준 관리 지침) 캐나다의 '표준에 관한 정책'을 벤치마킹하여 표준 관리 범위, 명칭, 의무 수준을 규정하고, 제정·개정·승인·폐기 절차와 관리 주체를 명확화 필요

- (역할 구분) 표준의 심의·의결은 국가통계위원회를 활용하고, 표준 관리부서는 공통 개념과 표준값을, 통계 프로그램 담당자는 개별 통계의 개념정보를 관리

[그림2] 국가통계 개념정보 관리 거버넌스(안)



❍ 실무 적용 고려사항

- (작성 범위) 조사표 문항 및 마이크로데이터 수준에서 나타나는 개념을 관리하는 것이 이상적이나, 관리 가능성을 고려하여 공표 통계표에 사용되는 개념과 변수를 우선 포함하되, 통계 프로그램 담당자에게 일정한 자율을 부여하는 방안 고려 가능

  * 교육정도 산출에 쓰이는 학력, 수학구분 코드처럼 공표 통계표에 드러나지 않지만 여러 통계에서 활용되는 개념이 있을 수 있음

- (개념 체계) 개념을 체계적으로 이해하기 위한 보조수단으로, Concept System 정보 클래스 활용하거나 현행 30개 주제 영역을 세분화하는 방안 고려 가능

- (개념의 역할 연계) '취업'은 개념변수, Universe, 범주로 모두 쓰일 수 있으므로, 개념 저장소 설계 시 개념의 역할 간 범용 활용성을 고려해야 함

시사점

❍ (선별 적용) GSIM 정보 클래스를 그대로 도입하기보다, 각 정보 클래스의 관리 실익과 국내 통계환경을 고려하여 선별·현지화하여 적용하는 것이 바람직함

❍ (벤치마킹) 캐나다의 표준 정책, 의무 수준 차등화와 범주 정의 제공, 노르웨이의 코드목록 구조 조회와 버전·대응표 관리를 상호 보완적으로 참조할 필요가 있음

❍ (단계적 접근) 통계 개념 표준 관리체계의 연착륙을 위해 통계 프로그램별 변수와 코드목록을 먼저 구조화하여 확보한 뒤, 공통성과 활용성이 높은 정보부터 단계적으로 표준을 지정·확대하는 접근 방식을 제안함

* 국가데이터처 작성 통계 → 중앙부처 작성 통계 → 지역통계 순으로 단계적 접근 고려 가능

❍ (기대 효과) 개별 통계에 분산된 개념 정보를 공통 구조로 통합 관리하고 표준 개념과 연계함으로써 통계 간 비교가능성과 재사용성을 획기적으로 향상하고, 나아가 통계·데이터 연계·융합을 촉진하는 핵심 기반이 되기를 기대함

한계 및 향후 연구

❍ (적용 범위) 본 연구는 캐나다·노르웨이 등 일부 국가의 사례를 검토하고 노동 분야를 중심으로 시범 적용했다는 한계가 있으므로, 다양한 통계 분야로 실증 범위를 넓히고 통계작성기관과의 협의를 통해 실효성을 검증할 필요

❍ (매뉴얼) GSIM 정보 클래스에 대응하는 국문 통계용어를 정립하고, 통계작성기관 담당자가 이해할 수 있도록 정보 클래스별 상세 작성 기준을 명확히 해야 함

❍ (시스템 연계) 변수·코드목록 정보를 국가통계포털과 연계하여, 하나의 개념정보를 여러 시스템에서 일관되게 조회·활용할 수 있는 구조 마련 필요

참고 문헌

○ 국가데이터처(2026),「2026년 경제활동인구조사 지침서 외부용(2026년)」
○ 국가데이터처, 통계설명자료서비스, https://www.k-stat.go.kr/metasvc
○ 국가데이터처, 통계정책시스템, https://www.narastat.kr/pms/index.do
○ 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
○ Statistics Canada, Classifications, variables and statistical units, https://www.statcan.gc.ca/en/concepts/search
○ 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
○ UNECE(2024), Generic Statistical Information Model (GSIM): Version 2.0 User Guide
○ 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=PLe6uBaBVebZDQXgqM-G__clN-MAVhXQ08
○ UNECE Statswiki, Generic Statistical Information Model, https://statswiki.unece.org/spaces/gsim/pages/59703371/Generic+Statistical+Information+Model

⇨ **(연구보고서 원본)** 국가데이터연구원홈페이지 > 연구간행물 > 연구보고서 > **국가통계 개념의 체계적 관리를 위한 GSIM 활용 방안**

출처를 밝히지 않고 국가데이터연구원 데이터 리서치 브리프의 내용을 무단전재 또는 복제하는 것을 금합니다.
본 리서치 브리프의 내용은 필자들의 개인적 의견이며, 국가데이터연구원의 공식적인 의견이 아님을 밝힙니다.

발 행 일 2026년 10월 30일
발 행 처 국가데이터처 국가데이터연구원
발 행 인 김 서 영
주 소 35220 대전광역시 서구 한밭대로 713 통계센터 6~8층
전화번호 042-366-7118
홈페이지 mods.go.kr/dsri