SMART DATA AND SEMANTICALLY RICH METADATA: EXPERIENCE IN PREPARATION AND DISSEMINATION IN THE CIS STATCOMMITTEE DATAHUB

III-й МЕЖДУНАРОДНЫЙ СТАТИСТИЧЕСКИЙ ФОРУМ СНГ

IIIrd CIS INTERNATIONAL STATISTICAL FORUM

E-PROJECTING

Yury Akatkin
Head of Research

Elena Yasinovskaya
Chief Analyst

SI PLATFORM

SEMANTIC KNOWLEDGE MANAGEMENT AND LOSD IN CISSTAT DATAHUB

UNIFIED INFORMATIONAL ANALYTICAL SYSTEM

SEMANTIC KNOWLEDGE MANGEMENT SYSTEM

SKMS

LOSD PREPARATION AND DISSEMINATION SYSTEM

LOSD PDS

Google Dataset Search

COPMETENCE CENTRE. EXPERT COMMUNITY

GOOGLE DATASET SEARCH CENTER

WEB & SEMANTIC WEB PUBLICATION
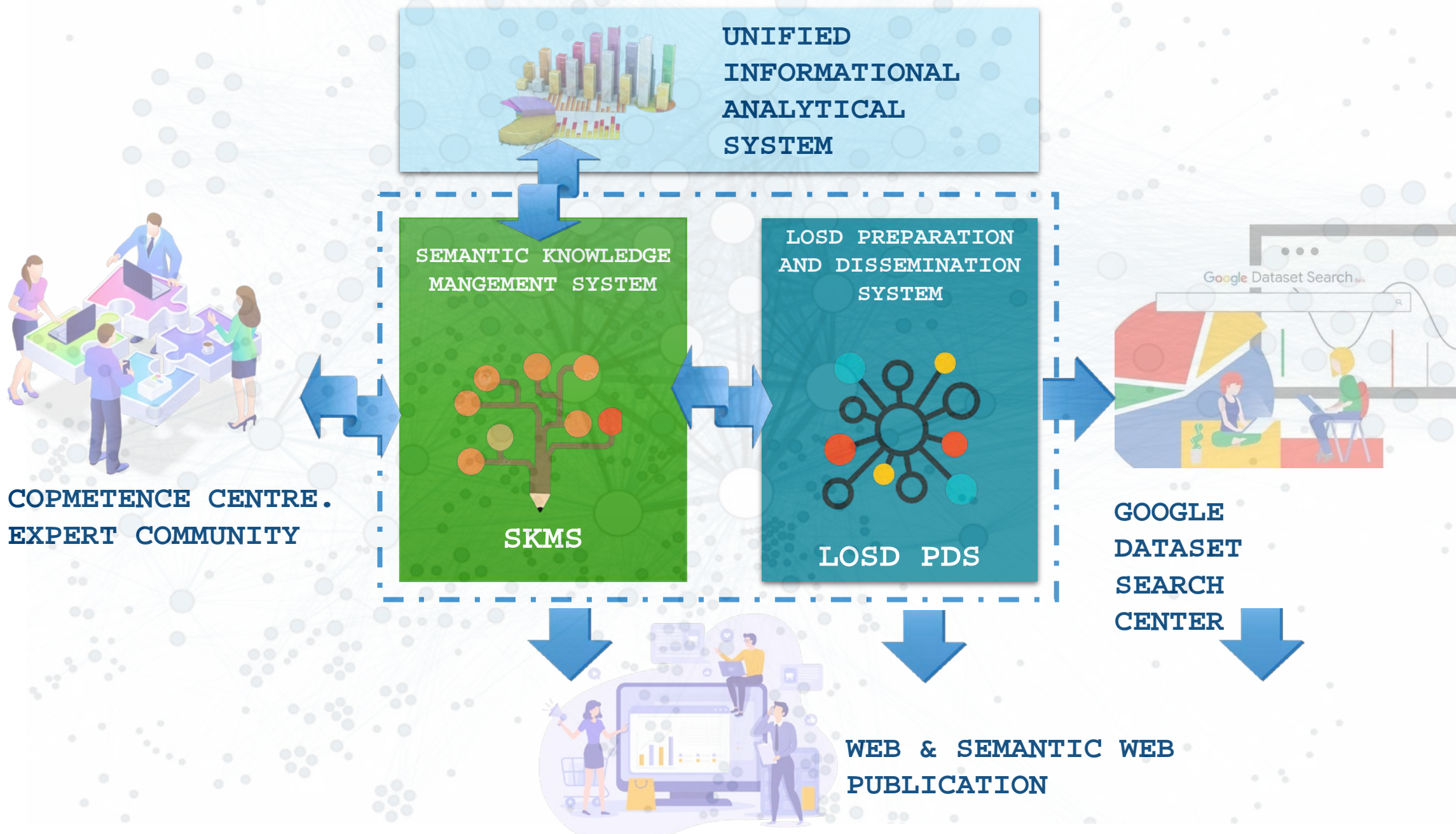
FORMATION AND DISSEMINATION OF SMART DATA AND SEMANTICALLY RICH METADATA
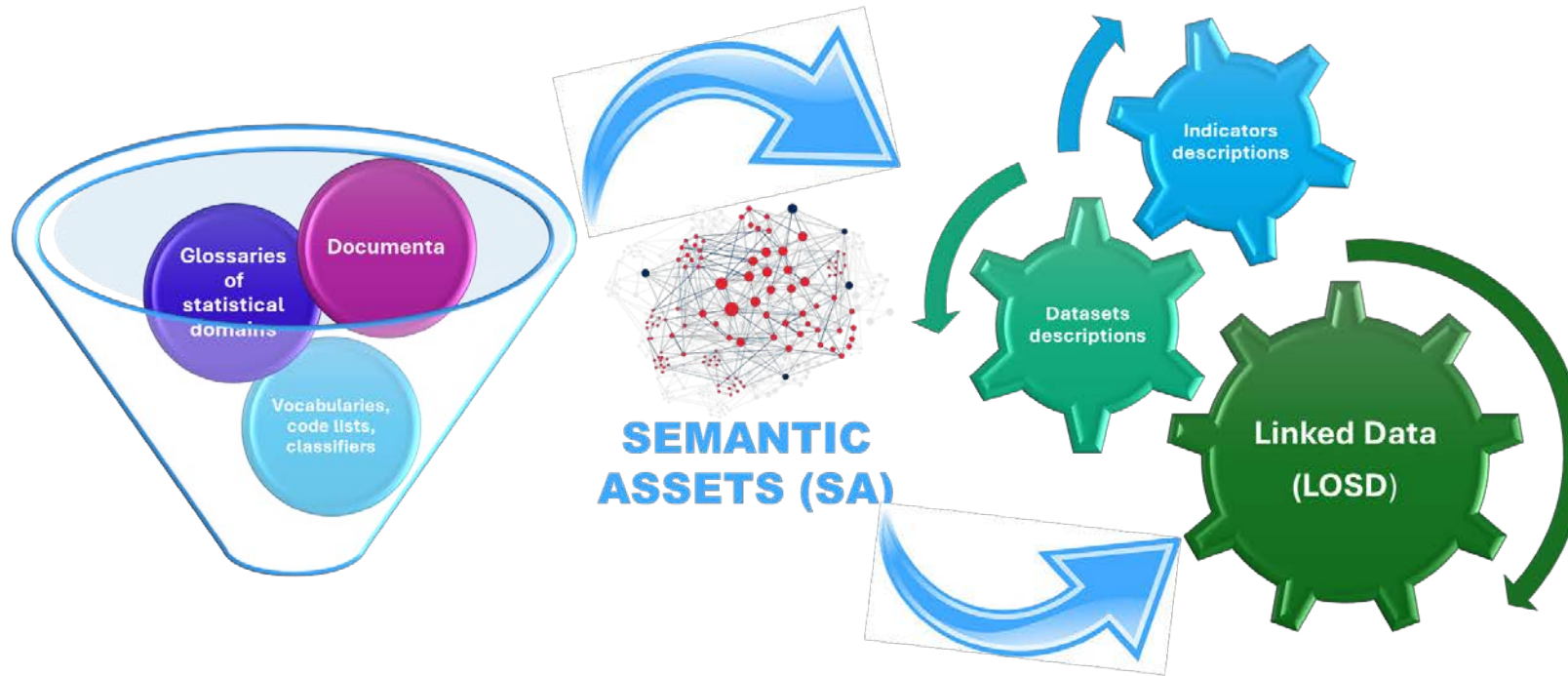
LOSD and semantic assets are published in the open catalogs of the LOSD PDS system

LOSD are formed using Semantic Web technologies as datasets containing an extended set of metadata

CISSTAT SEMANTIC KNOWLEDGE MANAGEMENT SYSTEM

Система управления знаниями Статкомитета СНГ

About the system | For experts | About the section | Indicators | Directories | Glossary | Methodology | Harmonization | Library

Sections

National accounts
Industry
Agriculture
Construction and investment in fixed assets
Prices
Finance
Population
Labor market
Standard and conditions of living
Tourism
General

Library
Edited by Andrey Shilin 2024/09/13 10:07

Methodological documents
International organizations
- ILO 2023 Resolution I. Resolution concerning statistics of the informal economy adopted by the 21st ICLS
- ILO 2023 Resolution II. Resolution concerning statistics of work, employment and labour underutilization (amending Resolution I of the 19th ICLS concerning statistics of work, employment and labour underutilization), adopted by the 21st ICLS
- ILO 2023 Resolution III amending Resolution 16th ICLS concerning the measurement of employment-related income, adopted by the 21st ICLS
- ILO 2020 Public Employment Services and the Future of Work. A Manual for the Regional Cooperation Network on Promoting Youth Employment in the CIS Countries

Commonwealth of Nations

CIS
Azerbaijan
Armenia
Belarus
Kazakhstan
Kyrgyzstan
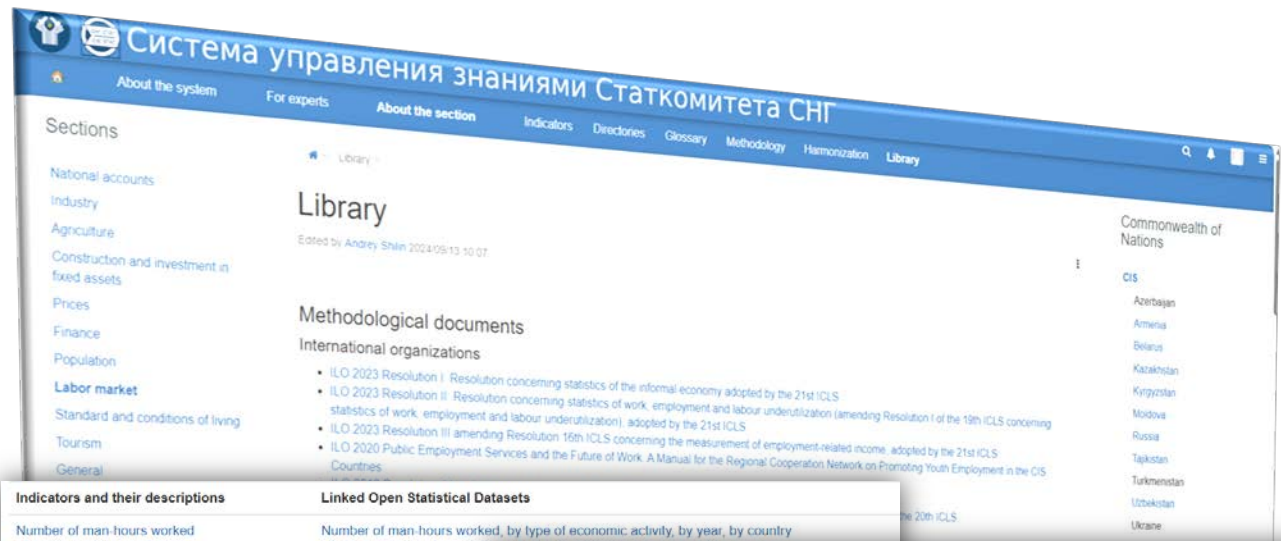Moldova
Russia
Tajikistan
Turkmenistan
Uzbekistan
Ukraine

the 20th ICLS

Indicators and their descriptions	Linked Open Statistical Datasets
Number of man-hours worked	Number of man-hours worked, by type of economic activity, by year, by country / Number of man-hours work...
Unemployment rate	Unemployment rate, total, b... / Unemployment rate, by age... / Unemployment rate, by gen...
Employment rate of the population	Employment rate of the pop... / Employment rate of the pop... / Employment rate of the pop...
Labor force participation rate/economic activity rate	Labor force participation/ec... / Labor force participation/ec... / Labor force participation/ec...
Number of employed people in the informal sector	Number of employed people...
Number of employed population	Number of employed popula... / Number of employed popula... population, by age, by year / Number of employed popula... / Number of employed popula... / Number of employed popula...
Number of young people who are not working (unemployed or not in employment) and not studying aged 15-24	Number of young people wh... years, total, by year
Labor force/economically active population	Labor force/economically ac... / Labor force/economically ac... / Labor force/economically ac...

DESCRIPTION
The module contains reference books prepared by experts based on the analysis of the applied classifiers and data collecti...

Organizational structures
Organizational structure of the CIS Statistical Committee

SDMX Code Lists
SDMX Codes
List of Codes "Unit of Measure"
List of Codes "Value per Period"
List of Codes "Multiplier"
List of Codes "Time Series Transformation"
List of Codes "Frequency"

Section directories
Economic activities (NACE Rev. 1)
Economic activities (NACE Rev. 2)
Units of measurement
Period value
Multiplier
Reporting period
Frequency of collection/dissemination
Gender
CIS countries
of time series transformation

Methodology
Edited by Andrey Shilin 2024/12/28 17:48

The section contains methodological materials "Standard and conditions of life" in a structured...

Methodological documents
International organizations
- ILO.2003.Report II. Household Income and Expenditure Statistics. 17th International Co...
- ILO.2003.Resolution concerning statistics of household income and expenditure adopted...
- CIS.2023. Methodological explanations for filling out questionnaire No. 15 CIS on house...
- CIS.2023. Methodological explanations for filling out questionnaire No. 18 of the CIS on...
- CIS.2021.Development of a system of poverty indicators for monitoring the achievement...
- UN.2010.Sample Designs for Household Surveys: Practical Guidelines (Excerpts)
- UNECE.2017.Guidelines on Measuring Poverty (extracts)

Belarus
- RB.2022. Methodological provisions on statistics of household income and consumption...
- RB.2022. Methodological provisions on housing stock statistics
- RB.2014. Methodology for calculating statistical indicators of the standard of living of the...
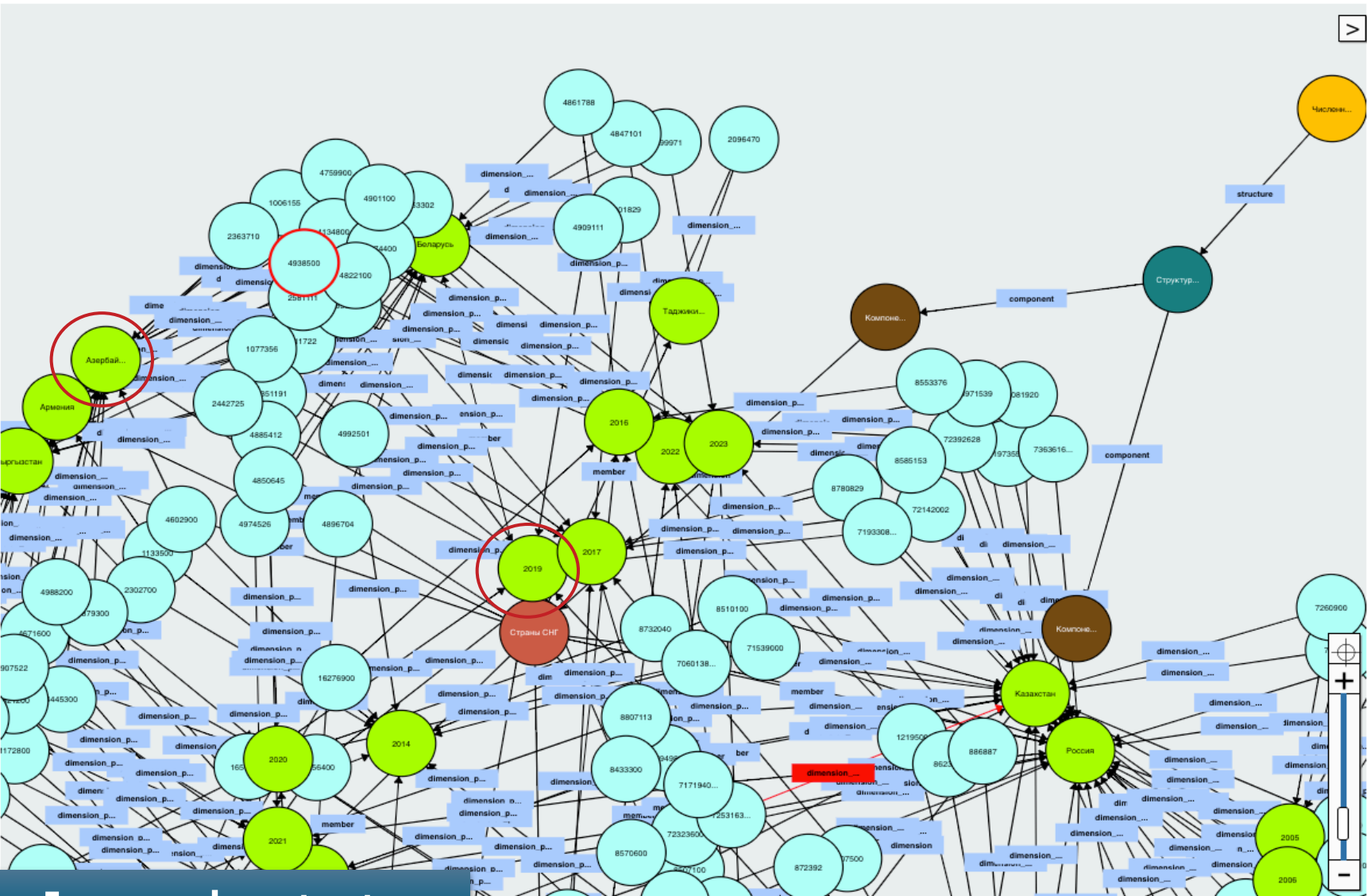
Russia
- RF.2019. Methodological provisions for the formation of aggregated indicators of househ... income and participation in social programs, approved by the order of Rosstat dated Ma... July 30, 2021 No. 455
- RF.2017. Methodological provisions for the formation of aggregate indicators of income,... household budgets
- RF.2011.Housing stock and public utilities (methodological provisions)

Glossary
Edited by Mikhail Beach 2025/06/15 15:34

IRI
http://purl.cisstat.org/StandardsOfLving/Glossary

DESCRIPTION
Glossary of the section "Standard and conditions of life"

The glossary is a dictionary of highly specialized terms of the section and is formed on the basis of the analysis of methodological documents, explar... encyclopedias in the field of statistics of the level and conditions of life. The glossary is implemented as a set of concepts that have an interpretation... with support for multilingualism, comments and examples, as well as semantic relations between concepts.

EXPLANATION
In developing the glossary, the results of the analysis of documents and materials, external sources, dictionaries and encyclopedias on the section, a... of the harmonization of concepts were used.

SOURCES
The glossary is based on the following sources:
- Harmonization
- Brief glossary of statistical terms
- Methodological provisions on housing stock statistics
- Methodological explanations for filling out questionnaire No. 15 of the CIS on household statistics
- Methodological explanations for filling out questionnaire No. 18 of the CIS on statistics of housing conditions of the population
- Resolution concerning stat...
- Resolution concerning stat...
- Handbook of Poverty Meas...
- Encyclopedia of statistical...

Glossary in RDF format   Downl...

Glossary contents:

A B C D E F G H...

Harmonization
Edited by Administrator 2024/08/14 20:53

Harmonization is mutual agreement, systemization, unification, coordination, ordering, ensuring mutual compliance.

Harmonization of terms
Harmonization of the concept of "Balance of Labor Resources"
Harmonization of the concept of "Unemployment"
Harmonization of the concept of "Employed Persons" Harmonization of the concept of "Persons not included in the labor force"
Harmonization of the concept of "Median Wage" Harmonization of the concept of "Minimum Wage"
Harmonization of the concept of "Part-Time Employee " Harmonization of the concept of "Labor Force" Harmonization of the concept of "Real Wages"
Harmonization of the concept of "Vacancies"
Harmonization of the concept of "Wage Rate Earnings"
Harmonization of the concept of "Working Age"
Harmonization of the concept of "Actually Worked Time"
Harmonization of the concept of "Economically Active Population"

Harmonization of classifiers and reference books
Harmonization of the reference book "Duration of job search"
Harmonization of the reference book "Methods of job search for the unemployed"

suz.cisstat.org

The SKMS provides interpretation of LOSD for people, while semantic assets (SA) enable it for information systems

SEMANTICALLY RICH INTERPRETATION ENVIRONMENT
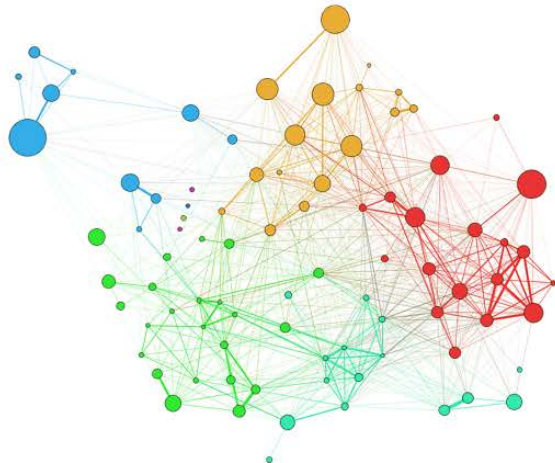
IMPROVE THE QUALITY OF STATISTICAL DATA AND METADATA

HARMONIZE STATISTICAL TERMINOLOGY AND ALIGHN METHODOLOGY

COMPLY WITH FAIR PRINCIPLES

PROVIDE SEMANTIC INTEROPERABILITY

FACILITATE (META)DATA RELEVANT INTERPRETATION

SEMANTIC MODELS

SMART METADATA

SEMANTICALLY RICH LOSD

VISUALISATION FOR EXPERT VALIDATION

>300 terms

>100 indicators

>500 docs in the library

>300 structured and marked up

2. Generation of semantic assets (SA): glossaries, indicators descriptions

3. Cataloging of SA: glossaries, indicators descriptions

>100 semantic models, incl. SDMX

1. Collection and systematization of documents, structuring, HTML-markup and publication on xWiki pages

4. Development and cataloging of SA: controlled vocabularies, code lists, statistics domain ontologies

SKMS

- xWiKi
- Templates for documents, glossary terms and indicator descriptions
- SA generators
- "Smart" metadata constructor

PDS LOSD

- OpenLink Virtuoso Database
- Liferay Portal
- SA & LOSD Catalogs
- JSON-LD generator

5. Loading datasets from UIAS, their semantic enrichment (RDF Data Cube) and cataloging

>350 LOSD sets

7. Construction of "smart" metadata, transfer to UIAS

6. Visualization and validation of semantic models (SA) and LOSD

8

HLG-MOS. INTEROPERABILITY IN STATISTICS

modernstats

DATA GOVERNANCE FRAMEWORK FOR STATISTICAL INTEROPERABILITY (DAFI)

[HLG2023 DAFI Final_0.pdf (unece.org)](HLG2023 DAFI Final_0.pdf)

INTEROPERABILITY

The ability of two or more information systems to exchange information and to use information obtained as a result of the exchange

RECOMENDATI ONS

Publish statistical data as LOD

- This will make it easier for machines to understand and use the data, and to link it to data from other sources
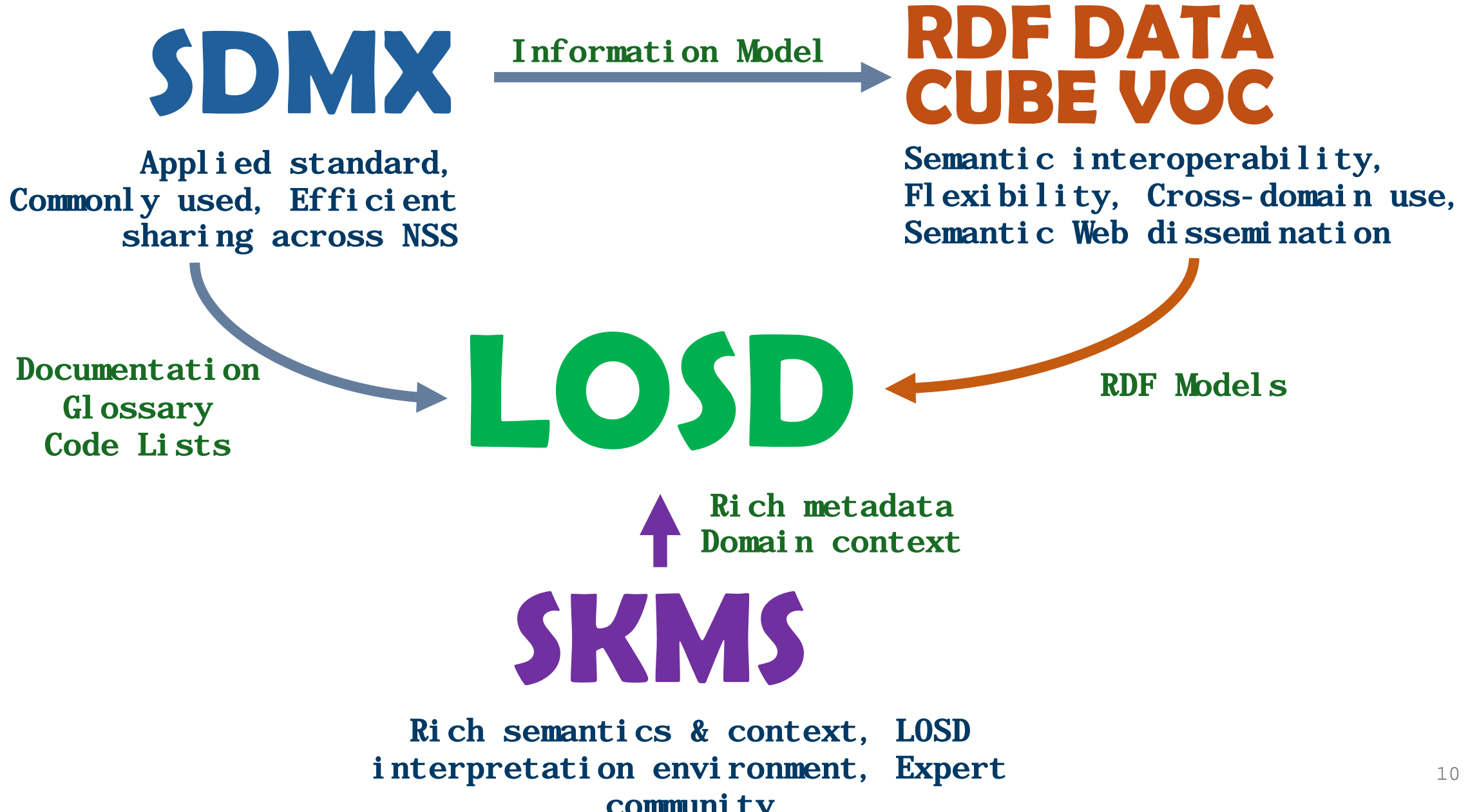
Use LOD to create a central repository for statistical metadata

- This will make it easier for users to find and understand the data that is available.

Develop applications that use LOD to automatically discover and use statistical data

- This will make it easier for users to access and use the data, and to create new and innovative statistical products and services

SDMX

Information Model →

RDF DATA CUBE VOC

Applied standard,
Commonly used, Efficient
sharing across NSS

Semantic interoperability,
Flexibility, Cross-domain use,
Semantic Web dissemination

Documentation
Glossary
Code Lists

LOSD

RDF Models

Rich metadata
Domain context

↑

SKMS

Rich semantics & context, LOSD
interpretation environment, Expert
community

INTEROPERABILITY BASIS

Open, non-profit initiative aimed at overcoming technological and organizational barriers that hinder the effective exchange and dissemination of Linked Data

GOAL

To integrate existing data exchange standards, classifications, and reference systems into the Semantic Web environment to achieve **sustainable semantic interoperability** across a wide range of user scenarios

IoBP

Interoperability Basis Platform (IoBP) supports **semantic alignment**, **enrichment**, and publication of existing standards using a knowledge management system, modeling tools, **namespace control, and persistent URI infrastructure**

⬡ INTEROPERABILITY BASIS PLATFORM   `basis.semanticip`   Log-in   🔍

🏠   SDMX   🌐 Agencies   🗄 Code Lists   🔀 Glossary   📄 Documents   📔 Library   `.org` 🏳 English   For Developers

Number of employed population

INTERNATIONAL NAME
Number of employed

SUBSECTIONS
Labor resources

DESCRIPTION OF THE INDICATOR
The indicator characterizes the number of employed people (aged 15 years and older , established for measuring the lab...

According to the document Methodological Explanations for Completing Questionnaire No. 14 , the employed populatio...
all sectors of the economy. Thus, the number of employed should include persons working in state enterprises and organi...
(peasant) households, as well as those engaged in individual labor activity, in personal subsidiary farming and for individu...

LINKS TO REGULATORY DOCUMENTS
- Resolution I of the 19th ICLS on statistics of work, employment and labour underutilization
- Methodological explanations for filling out questionnaire No. 14

DATA SOURCES
Form Table 14.1b. Distribution of the employed population by type of economic activity on average per year (persons)
Form Table 14.10. Economic activity of the population/labor force (people)

DATA COLLECTION METHODOLOGY
The labour force survey is conducted according to the methodology of the International Labour Organization (ILO) in ac...

The data are collected using forms Table 14.10 and Table 14.1b , which are completed based on data from the Labor Fc...

Recommendations for filling out are presented in the document Methodological explanations for filling out questionnaire N

VALUES FOR THE PERIOD
Average for the period

UNITS OF MEASUREMENT
Human

PERIODICITY (FREQUENCY) OF DISTRIBUTION
Annually

FREQUENCY OF COLLECTION
Annually

THE SYSTEM OF CLASSIFIERS
1. Reporting period
2. Types of economic activities (NACE Rev. 2)
3. Floor
4. CIS countries
5. Level of education
6. Age group
7. Employment status

CIS countries

Description

URI
http://purl.cisstat.org/Common/vocabs/country/1.0

SOURCE
The CIS Country Directory (hereinafter the Directory) is prepared on the basis of the Classifier of Countries of the World (hereinafter, the CIS CCM). The Directory is linked to the elements of the SDG reference area code list (SDMX) , using the Interoperability Basis Platform . Links have been established with the elements of DBPedia , EU Vocabularies , GeoNames , OASIS GeoLang TC .

PURPOSE
The reference book is used in statistical data sets that contain data across the CIS countries.

STRUCTURE
The directory structurally consists of a list of elements, each of which includes blocks: identification, description and URI.

Identification includes a three-digit numeric code, two-letter and three-letter ISO codes, and the URI of the element. Description includes the names of the countries that are part of the CIS. URI includes related external URIs.

The Handbook also shows the composition of the CIS.

The list of countries presented in the Directory includes countries (territories, regions) from Table 1 of the CIS CSM.

DESIGNATION
C

Directory in RDF fo

COMMONWEALTH OF INDEPENDENT STATES

URI : http://purl.cisstat.org/Common/vocabs/country/1.0#CIS

Digital code : 172

Equivalents :

https://dbpedia.org/page/Commonwealth_of_Independent_States ,

https://purl.semanticip.org/linked-data/sdmx/code/area-172 ,
https://purl.semanticip.org/linked-data/sdmx/code/area-R14

AI IN STATISTICS

Artificial intelligence in statistics today is a key area of development, unlocking new opportunities for data analysis, improving the quality of statistical (meta)data, and driving the digital transformation of workflows

SDMX Global Conference
September 2025

November 2025
Expert Meeting on Dissemination and Communication of Statistics

AI APPLICATION SCENARIOS

FOR SKMS EXPERTS

- Creation of document annotations
- Working with glossary terms
- Finding out international equivalents

FOR STATISTICIANS

- Preparation of methodological guidelines for questionnaires in accordance with international standards
- Translation and mapping of international classifications

ANY QUESTIONS?
ASK R&D TEAM!

elena@semanticpro.org

www.e-projecting.ru

E-PROJECTING