INTERSTATE STATISTICAL COMMITTEE OF THE COMMONWEALTH OF INDEPENDENT STATES

WELCOME TO THE STATISTICS OF THE COMMONWEALTH OF INDEPENDENT STATES

FAIR Statistics: Semantically Rich Environment for LOSD Creation and Interpretation

Konstantin Laykam, the Chairman
Statistical Committee of the
Commonwealth of Independent States

Yury Akatkin, Elena Yasinovskaya
Plekhanov Russian University of Economics

MOTIVATION. SDMX CHALLENGES

CHALLENGES

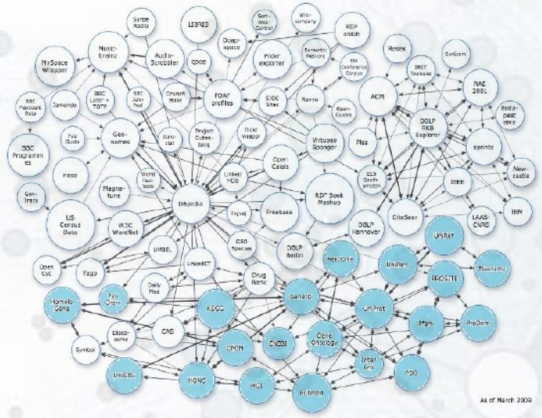
SDMX object information model **doesn't support the entire variety of relationships** between metadata entities

Extended metadata **(ESMS) is optional, quite complex to complete, and often not used at all**. Brief metadata (**Data Structure Definition**), describes **only the structure and encoding** of data sets

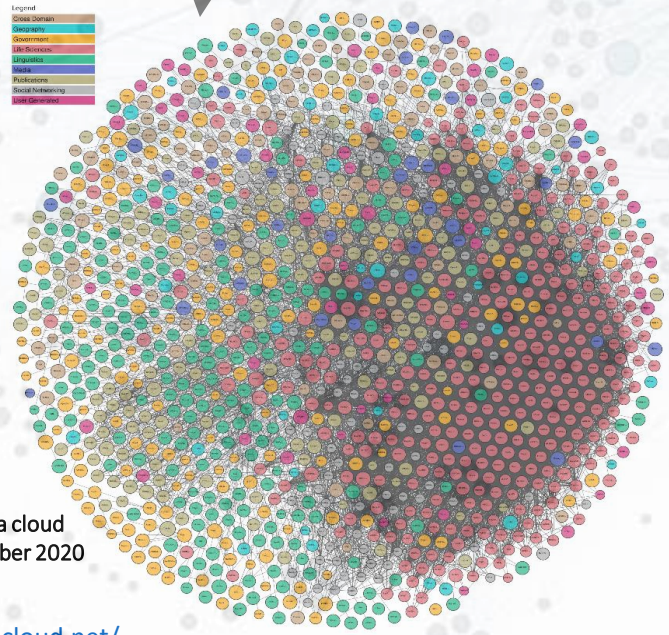
Metadata is usually distributed **separately** from data

MOTIVATION. LOSD DEVELOPMENT

Linked data cloud
in 2007



Linked data cloud
in March 2009



Linked data cloud
in September 2020

https://lod-cloud.net/

HLG-MOS PROPOSAL

The harmonization of statistical reference and national classifications in XKOS (Simple Knowledge Organization System Extension for Statistics)

Development and dissemination of data models for representing statistical classifications in the semantic web

Dissemination according to the LOD (Linked Open Data) principles in RDF

HLG-MOS Project Proposal for 2023

modernstats in HLG-MOS

MOTIVATION. POOR SEMANTICS

BASIC TOOLS AND MODELS

POOR SEMANTICS

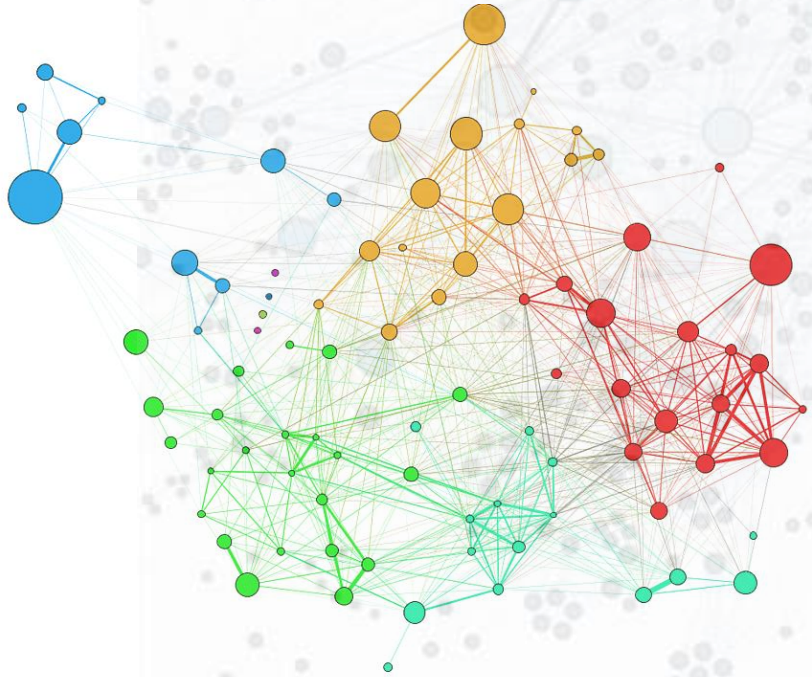
Lack of the context

No guarantee for unambiguous machine interpretation

No clear human understanding of data

CIC STAT. GOALS & BENEFITS

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**

CIS STAT. OBJECTIVES

To form a **methodological and terminological basis** for **constructing "smart" metadata and preparing LOSD** in CIS Statcommittee Data Hub

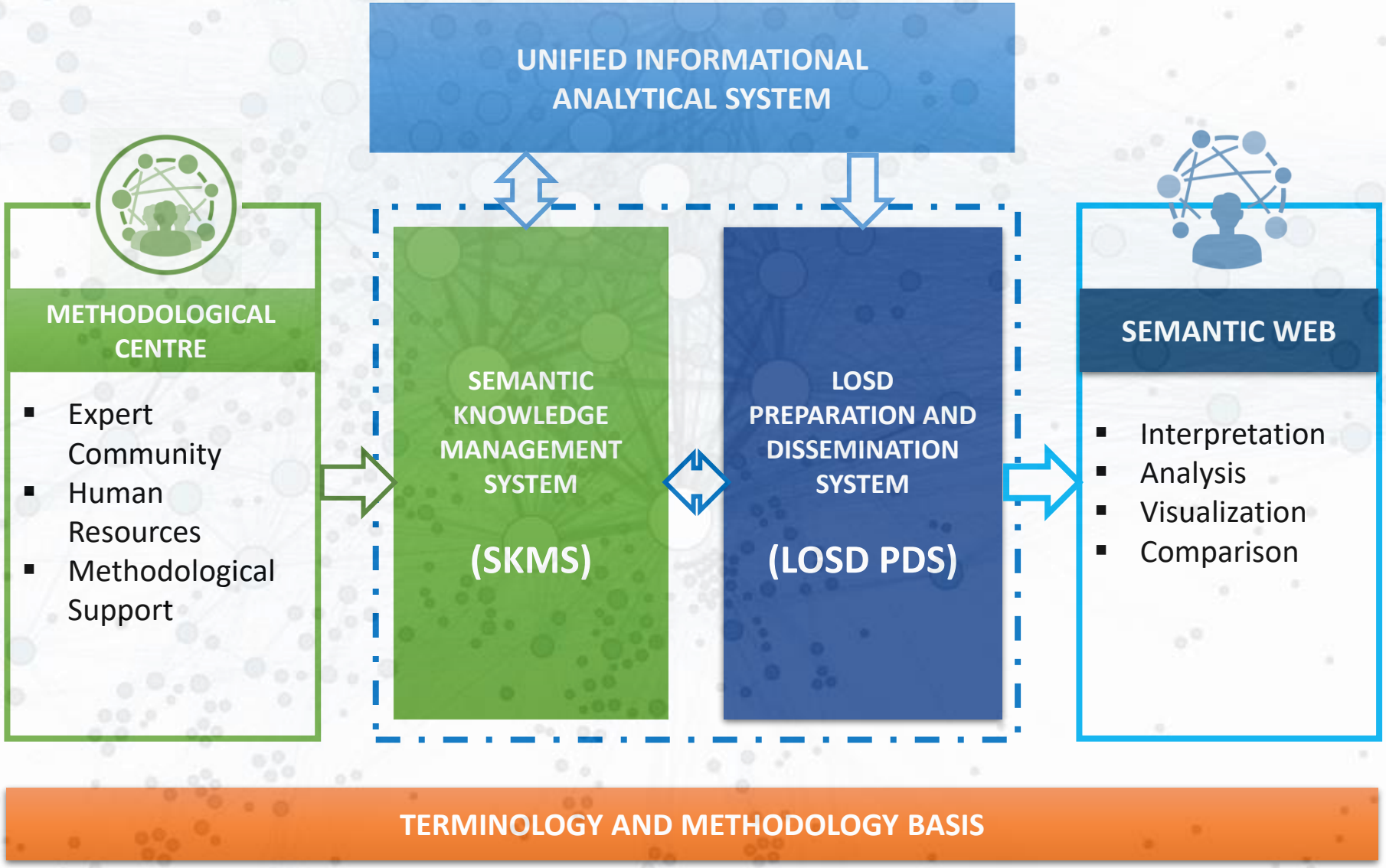
To provide **collaboration tools for expert statisticians and IT specialists** supporting **harmonization of statistical terminology and alignment of classifications**

To implement the **technology for describing statistical indicators**, to generate "smart" metadata based on this description and **semantically enrich the corresponding LOSD**

To ensure the **cataloging and dissemination LOSD and semantic assets** – core vocabularies, glossaries, controlled vocabularies

To support the **consistency validation** of developed "smart" metadata & LOSD and **semantically rich interpretation environment**

CIS STAT. DATA HUB



CIS STAT. SYSTEMS' OPERATION CYCLE

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

3.Cataloging of SA: glossaries, indicators descriptions

35 terms

10 indicators

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

>80 docs in the library

>40 structured marked up

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

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

>25 developed semantic assets

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

32 LOSD sets

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

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

FAIR COMPLIANCE

Principle	Requirements	Solution Compliance
Findability **F**	**F1. (Meta)data are assigned a globally unique and persistent identifier.** **F2. Data are described with rich metadata (defined by R1 below).** **F3. Metadata clearly and explicitly include the identifier of the data they describe.** **F4. (Meta)data are registered or indexed in a searchable resource.**	**Automatically generated and published on the catalog pages JSON-LD includes PURLs of data and "smart" metadata.** **"Smart" metadata are available via the provided PURLs using a standard protocol.** **(Meta)data is indexed by the Google Dataset Search Center using published JSON-LD.** **Generation mechanisms are described in our paper "The Challenges of Linked Open Data Semantic Enrichment, Discovery, and Dissemination".**
Accessibility **A**	**A1. (Meta)data are retrievable by their identifier using a standardized communication protocol.** **A2. Metadata are accessible, even when the data are no longer available.**	
Interoperability **I**	**I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.** **I2. (Meta)data use vocabularies that follow FAIR principles.** **I3. (Meta)data include qualified references to other (meta)data.**	**The semantically-oriented knowledge management system ensures the compliance with I1 – I3 due to semantic data enrichment and the construction of "smart" metadata.**
Reusability **R**	**R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.**	**Rich interpretation environment based on semantic models supports the reuse of "smart" metadata and LOSD.**

CONCLUSION

INTERPRETATION ENVIRONMENT BASED ON SEMANTIC MODELS

Knowledge management in statistics, harmonization of terminology and alignment of classification in statistics of the CIS countries

Human friendly means of visualization, machine-readable descriptions, supported by semantic models for indicators and corresponding LOSD sets

Development (generation), cataloging, and dissemination of semantic models and LOSD sets

GENERATION OF **SMART METADATA**

LOSD WITH **RICH SEMANTICS**

FAIR COMPLIANCE

CIS STATISTICAL COMMITTEE

SEMANTIC KNOWLEDGE MANAGEMENT SYSTEM & SYSTEM FOR LOSD PREPARATION AND DISSEMINATION

COMMUNICATION & COOPERATION

ANY QUESTIONS?

ASK R&D TEAM!

elena@semanticpro.org