## Smart Metadata and its role in statistical production

CIS-STAT has been collecting and summarizing data from the CIS NSOs for over 30 years and providing support in their statistical activities.

For these purposes, we have created a data and metadata hub which creates a number of advantages for users. First of all, it provides a single point of access to the statistical data we have accumulated for the CIS countries, as well as to the main international statistical resources. At the same time, the data hub provides them not only with data, but also with rich metadata formed on the basis of semantic models in accordance with the FAIR principles. An important feature of the hub is its multilingualism (English, Russian and also the official languages of a number of the CIS countries.

Structurally, the data hub consists of two subsystems: an analytical system and a knowledge management system in the field of statistical methodology and classification and related open data.

Analytical system of Statistical committee of CIS is a main source of CIS statistics and world statistics. It collects a set of statistical indicators of CIS countries and publishes the in
Statistical Portal in a form of highly-interactive dashboards. The system as well collects statistical data from all main international statistical sources such as World Bank, IMF, FAO,
ILO, Eurostat etc.

The system is based on a BI platform which consists of Data Warehouse, BI portal, OLAP server, Localizer and Open Data Catalog. All the users can pivot, filter, drill down, and the system calculates total, subtotal and time variances and displays it as rich visualizations. All the data is translated into Russian, English and Kazakh languages, so users can read the data on they own languages. In Open Data catalog users can briefly find a dataset using filters by all attributes, and download it in

various formats – CSV, JSON-LD and SDMX, all data is going with code list.

Based on the best practices and important international initiatives, such as the UNECE High-Level Group on the Modernization of Official Statistics in the sphere of ensuring interoperability, CIS-Stat is creating, as part of the Data hub development, a Knowledge Management System and Linked Open Data (KMS-LOD), which provides an interpretation environment open to external users. KMS-LOD, implemented in accordance with the FAIR principles (Findability, Accessibility, Interoperability, Reuse), forms rich semantics for the KMS, which allows generating “smart” metadata.

The system also supports the collaboration of a large team of statisticians and IT specialists to create a single, consistent terminology and methodological base necessary for the development of rich semantic models and for their use in the preparation and dissemination of linked data and “smart” metadata.

Our solutions provide:

- knowledge management in the field of statistics, harmonization of terminology and classifications in the statistics of the CIS countries;
- preparation of statistical indicators descriptions and corresponding linked datasets via user-friendly visualization tools;
- development (generation) and cataloging of semantic models and sets of knowledge management systems with the possibility of their distribution, including in machine-readable and machine-interpretable form.

The system was piloted on the "Labor Statistics" domain. Currently, the semantically-oriented knowledge management system of the Data hub already covers 9 statistical domains. Its electronic library contains

more than 250 documents, about 100 of which are structured and published on XWiki pages using hypertext markup linking them to more than 50 glossary terms and relevant documents. The catalog of knowledge management systems contains 35 developed knowledge management systems for 18 indicators of the "Labor Statistics" section.

By the end of December 2024, the system will be launched into trial operation, and the number of knowledge management systems for 9 statistical domains will be more than 100.

CIS-STAT is interested in expanding international cooperation in the creation of smart metadata for the purpose of integrating information resources and improving the technology of integration and dissemination of data and metadata.