IT solution to enhance KAS reference metadata system

Bekim Canolli, Kosovo Agency of Statistics, bekim.canolli@rks-gov.net

Burim Limolli, Kosovo Agency of Statistic, burim.limolli@rks-gov.net

Abstract

The Kosovo Agency of Statistics (KAS) has started producing and using standardized reference and quality related metadata ten years ago. In the past decades a lot of this information has been collected, stored and to a certain extent also disseminated, making it accessible to various (internal and external) users. The quality reports were written mainly in MS Word and were published in KAS webpage and sent to Eurostat via Metadata Handler. To overcome this problem, we started to develop a new application that would enable easier and more effective usage of reference metadata produced through the statistical process and would support the evaluation phase of our statistical business model. The application was developed and came to its production life in 2023. The main functionalities of the application are:

- *All the collected reference and quality related metadata are stored in one, central database and are ready to be used for different purposes.*
- *The application enables automated creation of Quality Reports based on the information stored in the central database.*
- *The application generates html and xml files that can be used for sending metadata via Metadata Handler.*

The paper describes the integrated architecture of the application, some further details of its functionalities, and points out main challenges that we met during its development.

Keywords: reference metadata, quality indicators, quality reporting, metadata, standardization.

1. Introduction

Commitment to quality is one of the principles of the European Statistics Code of Practice (ESCoP), stating that statistical authorities are committed to quality and systematically and regularly identify strengths and weaknesses to continuously improve process and product quality. Kosovo Agency of Statistics (KAS) have declared to apply the ESCoP in their daily practice, and implement a quality framework and a quality management system.

The overall objective is to provide high quality products and services to meet user needs. Quality is however not only needed and valued for products and services, it also relates to the institutions as a whole as well as to the institutional environment, hence, requiring actions for the overall management, organization and governance. Strong emphasis on quality remains a pre-condition for maintaining public trust in official statistics because the notion of quality is a key asset and comparative advantage in a world experiencing a growing trend of instant information, which often lacks the necessary proof of quality.

The Generic Statistical Business Process Model (GSBPM) is indeed a crucial framework that aims to standardize and improve statistical processes across various domains. By defining and

outlining the key steps and stages involved in statistical activities, the GSBPM provides a common language and structure for statistical organizations to follow.

One of the key benefits of implementing the GSBPM is the harmonization of metadata. Metadata, which refers to data that provides information about other data, plays a crucial role in ensuring the quality and consistency of statistical outputs. By standardizing metadata elements and definitions according to the GSBPM framework, statistical agencies can improve data comparability, facilitate data exchange and integration, and enhance data quality within the European Statistical System (ESS) and beyond.

Overall, the adoption of the GSBPM and the harmonization of metadata contribute to streamlining statistical processes, promoting interoperability among different statistical domains, and ultimately improving the overall quality and reliability of statistical outputs disseminated by statistical organizations.

Over the past few years, the Kosovo Agency of Statistics (KAS) has dedicated significant resources to modernize its statistical business processes. The primary objective has been to develop generic solutions that enhance the standardization of business processes and boost the efficiency of statistical processing. To overcome this problem, we started to develop a new application that would enable easier and more effective usage of reference metadata produced through the statistical process and would support the evaluation phase of our statistical business model.

2.   Methodological documentation at KAS

During the time when printed materials were widespread, printed publications often included additional methodological content. This extra information aimed to assist users in better understanding and interpreting the statistics presented.

The dissemination of the statistical data online raised the need for the standardization of the published metadata. A lot of effort has been invested at KAS in improving the structure of these methodological reports, as well as in improving the clarity and usability of the content. The structure and the content of the methodological explanations are now standardized for all statistical areas.

3.   Standardized quality reporting at KAS

The preparation of quality reports started in 2013. The first national template and guidance for quality reporting were prepared as a part of IPA projects.

The quality reports were being regularly prepared for all relevant domains, published in Albanian, Serbian and English versions. The reports were prepared based on a standard

template in Word format, most of the quality indicators were still calculated with the ad-hoc procedures, designed and implemented only for the particular domain. The word document was placed in webpage of KAS in dedicated place for metadata.

4.   IT solution to enhance KAS reference metadata system

After identification of problems it was agreed that a new, generic IT solution (SIMS application), which would improve the management of reference metadata, would be needed. The concrete activities to build up such an application started in 2020. In 2023 the application was launched as a part of KAS webpage.

Through this system, management and dissemination of metadata will be done through the KAS website and the same can be generated in SDMX format for transmission to Eurostat through the "Metadata Handler platform".

The conceptual premises that were the main drivers of the development process could be summarised through the following demands:

•   Reference metadata for all statistical surveys; these data are predominantly of qualitative nature and should be inserted into the database by the survey managers.

•   User-friendly graphical interfaces, which will enable easy and quick management of the centrally stored metadata by the end users, should be developed.

•   The application generates html and xml files that can be used for sending metadata via Metadata Handler.

The development of the entire application took about three years and during the development we were faced with many challenges due to the lack of IT staff.

Nonetheless, the internal IT specialists managed to carry out activities that are related to application maintenance, analysis of some security improvements, and the creative users to access the application, to train how to use the application and to be familiar with the concepts of the system.

It has been regulated access to the application through the use of credentials from the KAS domain console, which in this case facilitates access and provides greater security during the use of the application.

The hierarchy scheme is based on the organogram of the KAS. Starting with, the ordinary users who are managed by their supervisors, in this case, the heads of divisions, then the directors of the departments who manage the work of the users within the department. The next higher level comes the responsibility for the preparation of the material and documents

related to the topic they belong to and for controlling all users and downloading various reports as well as the highest level of access is the database owner.

The database is designed to communicate easily with the programming code. It is flexible to modification of SIMS because it is sufficient to add new concepts in the table. In the future, it will be more completed regarding the information on metadata flows and Metadata Structure Definitions.

Figure 1: The entity-relationship diagram
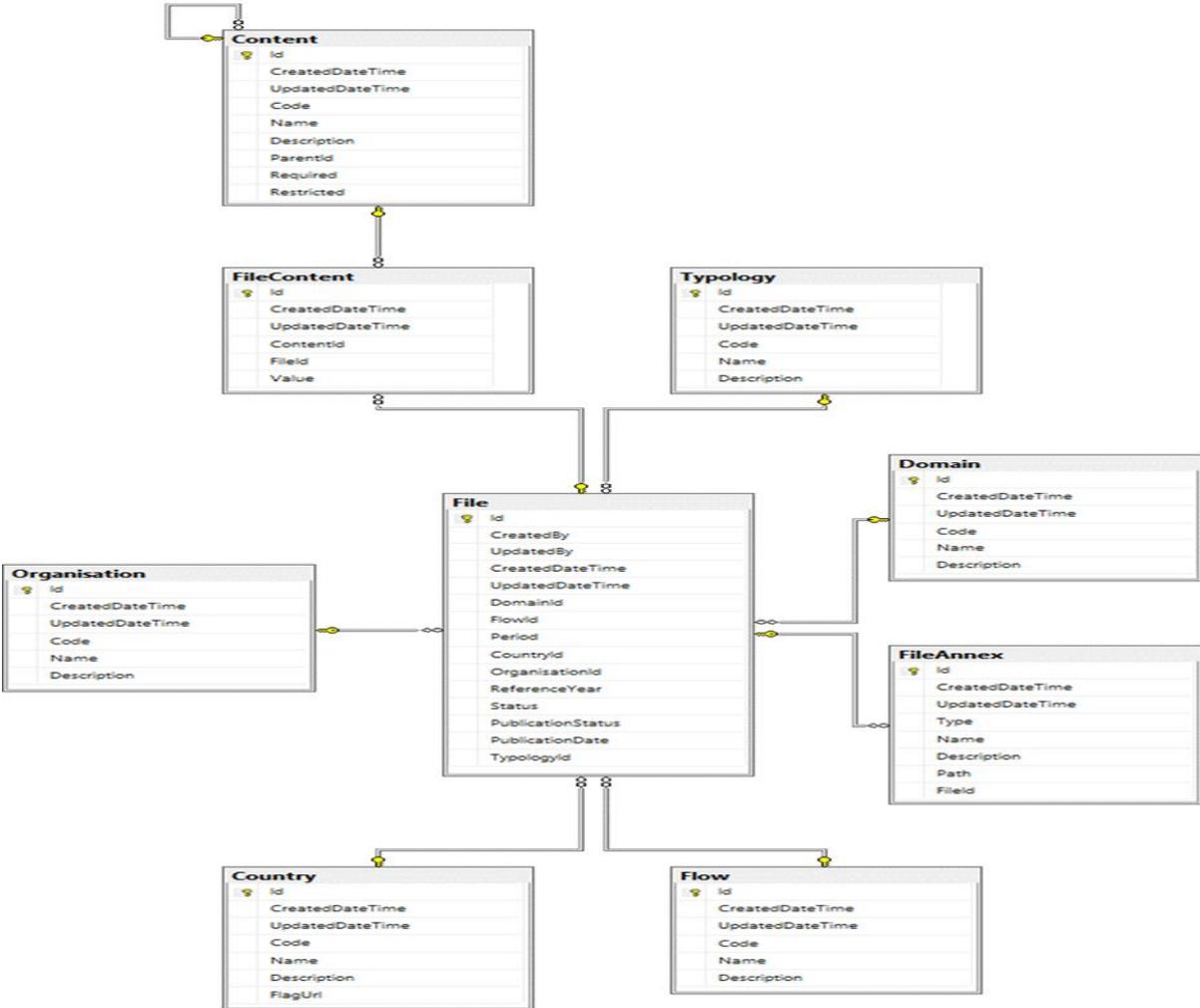
Upgrading the system is an explicit goal in the last version of the Agency's Roadmap, where is spelt out that "the SIMS (Single Integrated Metadata Structure) metadata system which is under development according to Eurostat standards will be upgraded. Through this system, the management and distribution of metadata will be done through the KAS website and the same can be generated in SDMX format for transmission to Eurostat through their Metadata Handler platform".

In the mid-long-term run, the system will produce quality reports according to the SIMS v2.0. Moreover, it will be possible to import domain-specific MSDs and other artefacts from the SDMX Registries available.

The metadata reports could be facilitated by features such as embedded guidelines, translation functionalities and concept pre-filling. The system will allow the automatic dissemination of quality reports on the KAS webpage. Moreover, it would be such as to enable KAS to send its national metadata to Eurostat via the ESS-MH, generating XML metadata files compliant with SDMX to enable the data transmission.

5. Main benefits of the new system

The implementation of the new metadata management system at KAS marks a significant advancement in our ongoing efforts to modernize our production system. Centralization and standardization serve as the foundational pillars of the system's conceptual architecture. While these principles undoubtedly offer numerous advantages to the statistical process, they also introduce certain drawbacks. Here, we provide a brief overview of the most notable benefits that have been identified during the system's development and initial implementation phase. The most important benefits detected are:

- The system of single entry for two purposes. Metadata is collected only once and inserted into the central database, from which it can be extracted and utilized for sending metadata to Eurostat via Metadata Handler. This approach significantly reduces the burden on subject matter specialists, as they only need to prepare metadata once.

- Total standardisation of the output methodological documents (methodological explanations and quality reports) is an important step forward to the better communication with the users of the statistical products.

- The application could be used from other National Statistical Authorities that are part of Statistical System.

5. Conclusions

Analyzing and documenting the quality of statistical data is a complex and often overlooked task, yet it is crucial and highly valued by those who rely on it. Systematically monitoring the quality of statistical processes and products is a key advantage of official statistics that must be preserved, particularly in times when the value of such data is questioned, as has happened in the past.

We understand that quality reports and reference metadata may not be the most popular reads, and they are unlikely to ever become popular. However, we recognize their significance as essential safeguards for the transparency of official statistics.

The technical and methodological efforts at KAS have yielded a solution that enables standardized and centralized management of reference metadata, with a particular focus on quality-related metadata. Our objective for the next phase of development is to enhance the utility of this information by integrating it more systematically into the assessment and improvement cycle. Therefore, one of our important tasks is to raise awareness that this information is not merely a requirement but also a valuable resource for utilization.

6. References

Eurostat, European Statistics Code of Practice, https://ec.europa.eu/eurostat/web/products-catalogues/-/ks-02-18-142

Programme of Official Statistics, Kosovo Agency of Statistics, https://ask.rks-gov.net/AboutUs/plans

Seljak R, Steenvoorden T. (2022), https://lsvdv-my.sharepoint.com/:w:/g/personal/ingabal_stat_gov_lt/Ea9JxOGXIHxFspOEjkr4XRYBg6is9KXn25ji_zlkwY-bHQ?e=GvBpcL