**Тезисы**

**доклада на заседании секции статистики ЦДУ РАН 06 апреля 2023 г.**

**по теме: «SDMX в современной статистической практике»**

* ***Что такое SDMX, этапы создания и интеграции стандарта***

SDMX – Международный стандарт обмена данными и метаданными, которому соответствует утвержденный стандарт ISO (17369:2013), призванный обеспечить эффективный обмен статистическими данными и метаданными и способствующий их распространению среди национальных и международных статистических организаций. Стандартизация в рамках формата достигается за счет использования утвержденных списков кодов и описаний структур статистических данных и метаданных.

* ***Функциональные особенности при работе со стандартом***

Стандарт предусматривает передачу как данных, так и метаданных. При этом выделяются два вида метаданных: структурные и справочные. Структурные метаданные выступают идентификаторами и дескрипторами (например, переменные, кодовые списки, наборы данных) и необходимы для определения данных (наименование динамического ряда, временной интервал и единица измерения).

Справочные метаданные включают:

* «концептуальные» (используемые понятия и их практическая реализация;
* «методологические» (описывают методы, используемые для производства статистических данных);
* «качественные» (различные измерения качества полученных статистических данных).

Описание структуры данных (DSD) включает три основных компонента:

Измерения (dimensions), описывающие статистические данные и формирующие их идентификатор (ключ) к соответствующим данным (например, время, область, пол);

Значения (measures), которые отражают значение статистической величины за период времени;

Атрибуты (attributes), которые отражают дополнительную информацию о какой-либо части рассматриваемого набора данных и могут характеризовать набор данных, наблюдение или группу измерений.

* ***Общий механизм организации обмена данными в соответствии с SDMX***

1. Определение данных для передачи (в рамках стандарта осуществляется распространение данных по основным тематическим областям статистики (statistical domains): демографическая и социальная статистика, экономическая статистика, экологическая и многоотраслевая статистика);

2. Определение концептуальной схемы передачи данных;

3. Создание списка кодов;

4. Создание специальных структур данных (DSD) и метаданных (MSD);

5. Формирование потоков данных и метаданных;

6. Создание связей между национальными данными и DSD;

7. Создание набора данных;

8. Настройка процесса *маппирования*;

9. Формирование данных в формате SDMX.

* ***Инструментарий, позволяющий обеспечить передачу данных в соответствии со стандартом***
* **DSD Constructor** Приложение, способное преобразовывать/редактировать часто используемые форматы метаданных в форматы SDMX-ML. Включает в себя интерфейс, который позволяет создавать DSD и редактировать данные в соответствии с требованиями.
* **Fusion Registry** Полностью интегрированная система управления данными и метаданными SDMX, обеспечивающая создание структурных метаданных, справочных метаданных, сбор данных, проверку данных, преобразование данных и распространение данных.
* **SDMX Converter** Конвертер SDMX — это инструмент, который преобразует наборы статистических данных в файл, соответствующий требованиям SDMX
* **.Stat Suite** Платформа с открытым исходным кодом, обеспечивающая производство и распространение высококачественных статистических данных. Продукт ориентируется на Типовую модель производства статистической информации (GSBPM) и SDMX.
* ***Передача данных ЦУР статистическому отделу ООН***

Росстат является членом международной МЭГ-ЦУР по SDMX для показателей ЦУР и принимает регулярное участие в заседаниях Рабочей группы.

В сентябре 2019 года Россия приняла участие в апробации SDG DSD v.1.0. Росстатом был подготовлен пакет данных в формате SDMX-ML по 53 показателям ЦУР и представлен в Статистический отдел ООН. В декабре 2020 года выпущено новое описание структур данных SDG DSD v 1.3, в соответствии с которым оперативно подготовлен обновленный пакет данных по 59 показателям и представлен в Статистический отдел ООН. К 2023 году Росстат на постоянной основе направляет пакет данных в формате SDMX-ML по 77 показателям ЦУР

***Перспективы развития стандарта SDMX в международной статистической практике*** В 2019 году Европейских центральный банк начал проект по созданию платформы, которая должна заменить существующую уже более 26 лет информационную систему макроэкономической статистики. Разрабатываемая система ориентирована на обработку больших данных, а технологический стек программных продуктов Hadoop был выбран как основа для создания платформы.

Возможность интеграции методологических подходов SDMX в Big Data архитектуру была отмечена разработчиками еще на стадии создания проекта. SDMX наблюдения можно идентифицировать по его значениями, заданной контрольной дате и прикрепленным атрибутам, таким образом данный метод описания информации подходит для базы данных столбцового хранения, предназначенной для хранения денормализованных таблиц.

Также ведется работа направленная на:

• упрощение и усовершенствование модели справочных метаданных;

• поддержку микроданных;

• поддержку геопространственных данных;

• поддержку расширения списков кодов и дизъюнктного объединения списков кодов;

• усовершенствование картирования структуры;

• усовершенствование иерархий кода для обнаружения данных;

• усовершенствование ограничений;

• усовершенствование версионирования артефактов структурных метаданных;

• усовершенствование интерфейса программирования приложений веб-служб REST;

• усовершенствование и упрощение форматов обмена данными и метаданными.