

Цели

Целью создания **СУЗ-СОД** является формирование:

- **библиотеки** структурированных методологических материалов;
- **гlossария** терминов статистики;
- **наборов связанных открытых статистических данных (СОСД)** по показателям СНГ, обогащенных широкой нормативно-справочной информацией, для обеспечения однозначной содержательной интерпретации данных, в том числе в информационных системах потребителей.

Достижение поставленных целей обеспечивается путем:

- Создания системы общедоступных семантических моделей в области статистики СНГ на базе национальных и международных глоссариев, классификаторов, стандартов, а также формирования на ее основе среды содержательной интерпретации СОСД.
- Построения системы управления знаниями на основе общедоступных семантических моделей, обеспечения выравнивания и гармонизации статистической методологии стран СНГ.
- Создания семантических информационных ресурсов, отработкой технологий распространения и применения СОСД.
- Улучшения организации, интеграции и предоставления данных в сети Интернет, развития возможностей обнаружения, поиска, навигации и сопоставления статистических данных, полученных из различных источников с использованием семантических методов.
- Повышения осведомленности потребителей и поставщиков данных о преимуществах СОСД и возможности их применения в цифровых экосистемах и сервисах.

Связанные открытые статистические данные

В качестве наиболее эффективного способа предоставления данных в Интернете Консорциум [W3C](#) рекомендует использовать связанные данные, разработанные в соответствии с правилами [Semantic Web](#) (глобальной семантической сети) – набором основных правил представления данных в Интернет в виде, позволяющем адекватно интерпретировать их в информационных системах с учётом смыслового значения.

Связанные открытые статистические данные (СОСД) – это связанные данные в области статистики, которые публикуются с использованием открытой лицензии [CC BY 4.0](#), что способствует их повторному бесплатному использованию. Интероперабельность обеспечивается за счет создания, обмена и использования СОСД так, что их значение и контекст сохраняются независимо от системы или набора систем.

Понимание и использование данных потребителями, не являющимися экспертами в данной области, а также их интерпретация информационными системами зачастую затруднены из-за отсутствия формализованных знаний о предметной области и машиночитаемых данных, дополненных семантикой для понимания смысла набора данных. Бедность семантики приводит к тому, что связанные данные трудно обнаружить в Интернете по их описаниям и однозначно связать с понятиями предметной области.

[Группа высокого уровня по модернизации статистики \(HLG MOS\) ЕЭК ООН](#) на международном уровне решает задачи достижения интероперабельности данных в Национальных статистических управлениях (НСУ) путем разработки собственных и внедрения существующих методов, моделей (в том числе семантических, таких как онтологии) и стандартов в рамках реализации специализированных программ. HLG MOS осуществляет проект «Структура управления данными для статистической интероперабельности». В 2023 году HLG MOS выпустила документ [Data Governance Framework for Statistical Interoperability \(DAFI\)](#), в котором представлена справочная структура (reference framework), содержащая основные элементы для внедрения программы управления, направленной на достижение интероперабельности статистических данных. Как результат, внедрение этих подходов позволят статистическим организациям улучшить управление (мета)данными и обеспечит возможность создания, обмена и использования данных там, где их значение и контекст сохраняются независимо от системы или набора систем.

Ещё одним новым направлением развития, продвигаемым HLG MOS в сфере модернизации статистических процессов, является формирование богатых или [«умных» метаданных](#), которые не только стандартны (т.е. их можно понять и использовать везде) и активны (позволяют реализовать управление статистическими процессами), но и делают данные находимыми (F), доступными (A), интероперабельными (I) и пригодными для повторного использования (R). Ключевым способом реализации [принципов FAIR](#) считается применение семантических технологий не только для распространения данных, но и для формализации знаний в виде семантических моделей – стандартных словарей, глоссариев, тезаурусов, онтологий.

Применение СОСД обеспечит развитие возможностей обнаружения и поиска статистических данных СНГ в сети Интернет, сопоставления и совместной обработки данных из различных источников с использованием технологий [Semantic Web](#), в том числе методов семантической интеграции.

Применяемые стандарты и технологии

1. [FOAF \(Friend Of A Friend\)](#) – словарь именованных свойств и классов, построенных с использованием технологии RDF от W3C и языка OWL.
2. [Vcard \(The Electronic Business Card\)](#) – формат данных для передачи контактной информации в стандартизированном виде. Например, на телефон или в почтовый агент.
3. [OWL \(Web Ontology Language\)](#) – язык онтологий для Semantic Web (SW), описывающий классы и отношения между ними, присущие документам и приложениям Semantic Web.
4. [Дублинское ядро \(Dublin Core™\)](#) – набор элементов метаданных «Дублинское ядро», определённый Дублинской основной инициативой по метаданным, включающий свойства, схемы кодирования словаря, схемы кодирования синтаксиса и классы.
5. [RDF Concepts \(RDF 1.1 Concepts and Abstract Syntax\)](#) – модель представления знаний (данных и метаданных). RDF-схема для терминов словаря RDF в пространстве имен RDF, определённая в RDF 1.1 Concepts.
6. [RDFS \(RDF Schema 1.1\)](#) – набор классов и свойств для модели представления знаний RDF, составляющий основу для описания онтологий с использованием расширенного RDF-словаря для структуры RDF-ресурсов.
7. [RDF Data Cube Vocabulary](#) – словарь RDF Data Cube позволяет публиковать многомерные данные, такие как статистика с помощью стандарта W3C RDF. Модель, лежащая в основе словаря Data Cube, совместима с моделью куба, лежащей в основе SDMX.
8. [SDMX \(Statistical Data and Metadata Exchange\)](#) – международный стандарт обмена статистическими данными и метаданными.
9. [SKOS \(Simple Knowledge Organization Model\)](#) – базовая модель организации знаний для Semantics Web, облегчающая взаимодействие различных информационных систем за счёт

стандартизации тезаурусов, систем классификации, таксономий, фолксономий и других видов нормализации лексики.

10. [SKOS-XL \(SKOS eXtension\)](#) - словарь RDF, расширяющий SKOS, предоставляя дополнительную поддержку для описания и связывания лексических единиц.
11. [XKOS \(SKOS extension for representing statistical classifications\)](#) – специализированное расширение SKOS для управления статистическими классификациями.