



РАЗРАБОТКА КОНЦЕПЦИИ ПО СОЗДАНИЮ И РАСПРОСТРАНЕНИЮ СВЯЗАННЫХ ОТКРЫТЫХ СТАТИСТИЧЕСКИХ ДАННЫХ РОССТАТА

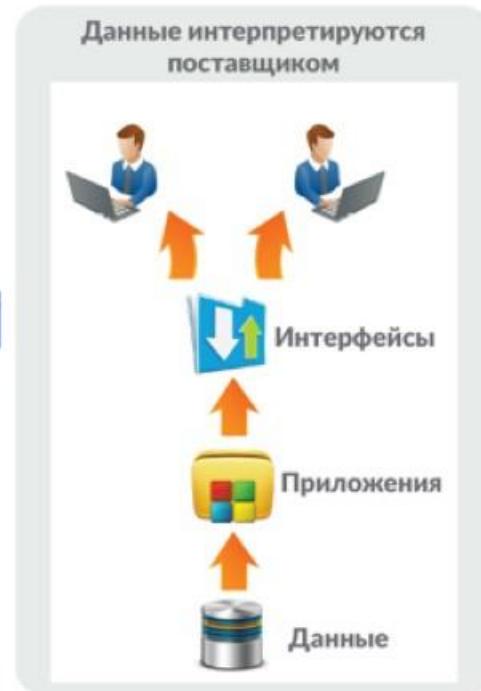
Юрий Михайлович Акаткин, к.э.н,
заведующий научной лаборатории
семантического анализа и интеграции
РЭУ им. Плеханова



Роль связанных данных для цифровизации



Переход к датацентричности



- **Цель создания** Цифровой аналитической платформы предоставления статистических данных:
Обеспечение перехода к новой структурной и функциональной модели производства и распространения статистических данных, реализующей принципы единого информационного пространства данных, принципиально нового качества данных, прослеживаемости и взаимосвязи понятий и объектов статистического учета, однократного предоставления первичных статистических данных и их многократного повторного использования, интеграции этой новой модели в цифровую экономику
- Один из **принципов** создания платформы:
связанность, сопоставимость, однозначная интерпретируемость сведений, полученных из разных источников официальной статистики

Связанные открытые данные

- ❖ Связанные открытые данные (СОД) – наборы открытых данных, разработанные в соответствии с принципами Semantic Web.
- ❖ Публикация СОД облегчает поиск и интеграцию данных, а технологии Semantic Web обеспечивают среду, в которой приложения могут запрашивать данные и управлять ими, формировать интерфейсы и делать выводы с учетом семантических связей.
- ❖ В 2017 г. Консорциум всемирной паутины (W3C) рекомендовал связанные данные в качестве наиболее эффективного способа открытия данных в Интернете.

Принципы

Используйте URI в качестве имен объектов

Используйте http URI таким образом, чтобы эти имена можно было найти.

Когда кто-то обращается по URI, предоставьте полезную информацию с помощью стандартов (RDF, SPARQL)

Включайте ссылки на другие URI так, чтобы можно было обнаружить больше объектов

Стандарты

Пользовательский интерфейс и приложения

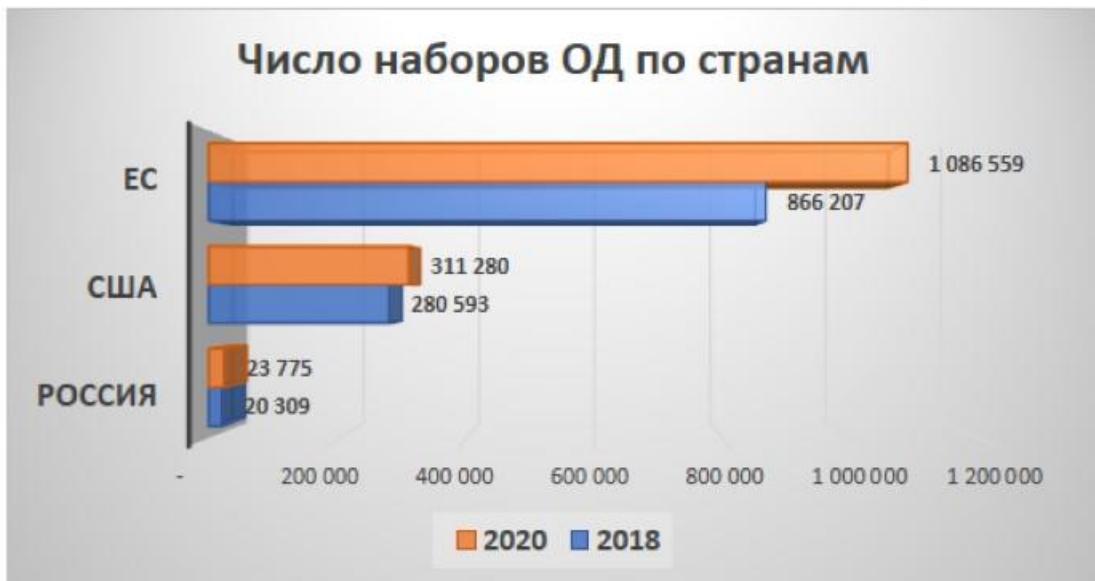
Запросы:	SPARQL	Онтологии: OWL	Правила: RIF/SWRL
Обмен данными:	RDF	Таксономии:	RDFS
Синтаксис:			XML
Идентификаторы:			URI
Кодировка:			UNICODE

Качество

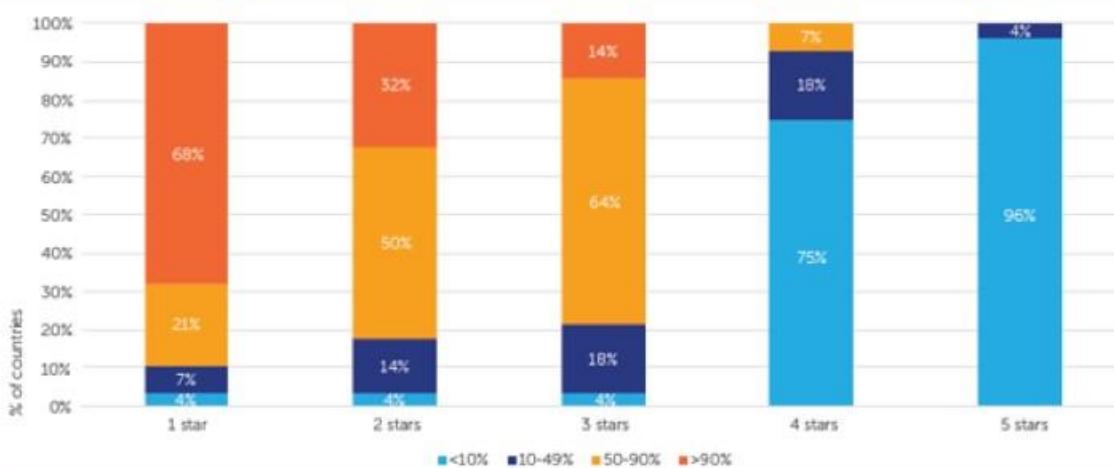


★	Сделайте доступными свои материалы в Интернете (вне зависимости от формата) под открытой лицензией
★★	Сделайте их доступными в виде структурированных данных (например, Excel вместо сканированного рисунка или таблицы)
★★★	Используйте незапатентованные (непроприетарные) форматы (например, CSV, JSON или XML)
★★★★	Используйте URI для обозначения сущностей, чтобы люди могли ссылаться на них
★★★★★	Связывайте свои данные с другими для обеспечения контекста

Текущее состояние СОСД в России и в мире



ОТКРЫТЫЕ ДАННЫЕ ЕС В СООТВЕТСТВИИ С 5* МОДЕЛЬЮ



ФОРМАТЫ ОД В РОССИИ



Работы первого этапа НИР

1.

Исследование международного опыта разработки нормативных правовых и нормативно-технических документов, а также научно-методических материалов по подготовке, хранению и распространению связанных открытых статистических данных и по применению для этих целей семантических методов.

2.

Исследование международных и российских технологических подходов и технических решений по подготовке связанных статистических данных с использованием семантических активов (тезаурусов, онтологий, моделей предметных областей и других семантических моделей), а также к их хранению и распространению.

3.

Разработка предложений к проекту концепции по созданию и распространению связанных открытых статистических данных Росстата.

4.

Разработка предложений к проекту дорожной карты по созданию и распространению связанных открытых статистических данных Росстата.



Результаты исследования нормативных и научно-методических подходов к подготовке, хранению и распространению связанных открытых статистических данных

Рассмотрены научно-методические подходы к подготовке связанных открытых данных с учетом применения семантических методов

Изучены международные проекты и инициативы в области связанных открытых статистических данных и применения семантических активов для их подготовки

Исследованы стратегические, концептуальные, методические и нормативные документы, направленные на развитие связанных открытых данных в статистике



Цифровизация мировой статистики

The screenshot shows the 'Modernisation Groups' section of the HLG-MOS website. It includes a sidebar with links to 'ModemData Models', 'Modernisation Topics', and 'Collaboration Spaces'. The main content area discusses the 'Supporting Standards' group, which aims to find ways to develop, enhance, integrate, promote, support and facilitate implementation of the range of standards needed for statistical modernisation. It also mentions the 'Sharing Tools' group, which continues the work of the HLG CSIA implementation Standards and the MC Production and Methods. There are sections for 'Task teams' and icons representing different groups like 'Statistical Data' and 'Geospatial Business'.

The screenshot shows the 'Search and Share Development Data' interface on the World Bank datahub. It features a search bar, a 'Featured Data' section with cards for 'COVID-19 Data', 'Data on Statistical Capacity', and 'World Bank Open Data', and a detailed view of the 'World Bank Linked Data' dataset, which provides statistical observations and data from World Bank Finances, World Bank Projects and Operations, and World Bank Climate Change data.



Цифровизация национальных статистических служб



e-Stat Statistical LOD of Japan

Outline Guidelines Providing data RDF Schema SPARQL endpoint



Linked Open Data

Statistical data of Japan is presented as LOD(Linked Open Data) that is the best rank of open data (5-star open data).



Istituto Nazionale
di Statistica

POPULATION & HOUSEHOLDS INSTITUTIONS & SOCIETY EDUCATION & LABOUR ENVIRONMENT & TERRITORY



HOME > METHODS AND TOOLS > ONTOLOGIES

ONTOLOGIES



ANALYSIS AND PRODUCTS

METHODS AND TOOLS

GLOSSARY

CLASSIFICATIONS

ONTOLOGIES

METHODS AND TOOLS

Design

Collect

Process

Analyse

ON LINE SYSTEMS

TOOLS FOR DATA QUALITY

References

The term ontology means a formal representation of the real world that "translates" the natural language (ambiguous by its nature) into a formal (unambiguous) language that can be used for the creation of integrated information systems.

Therefore, a **computational ontology** – for a domain of interest – is a representation:

- **formal**, expressed in a symbolic non ambiguous language that can be processed by computers;
- **shared**, resulting from the approval of a plurality, the widest possible, of domain experts;
- **explicit**, every assumption is made as explicit.

The main advantages of the representation of the domains of interest through

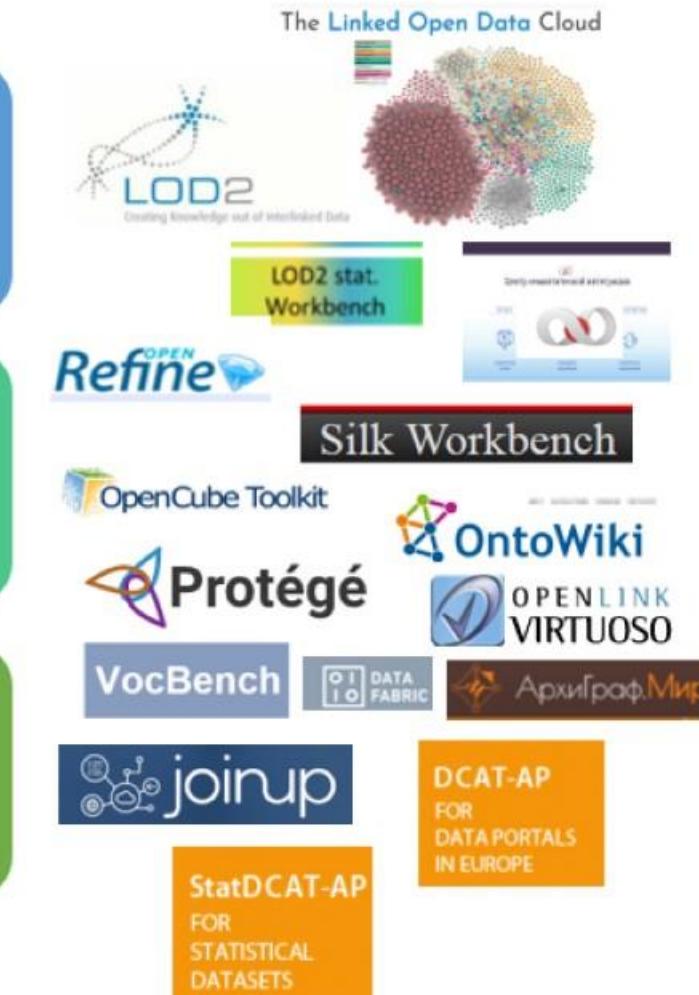
2.

Результаты исследования технологических подходов к подготовке связанных статистических данных с использованием семантических активов, а также к их хранению и распространению

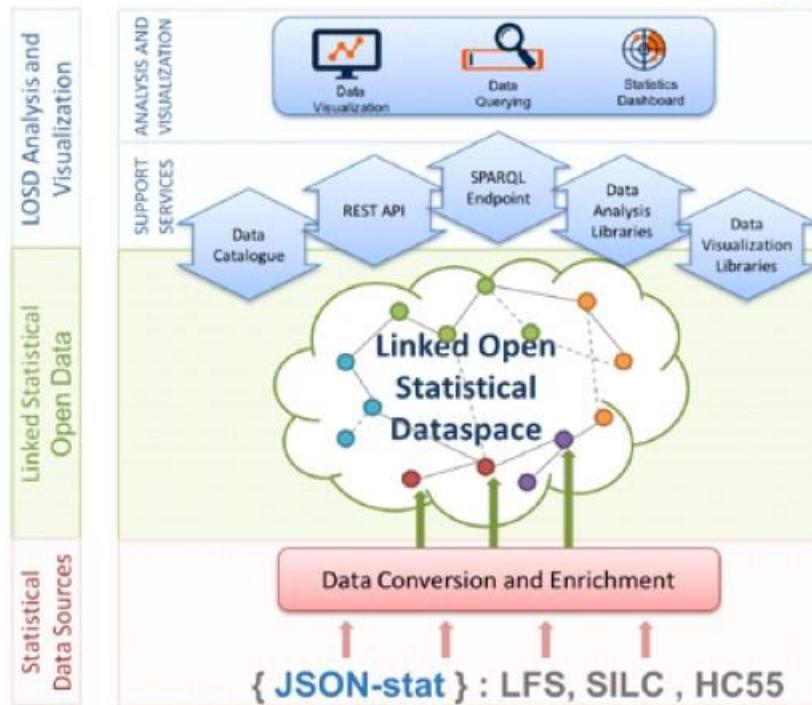
Представлен обзор зарубежных и российских платформ подготовки связанных данных с использованием семантических активов

Рассмотрены зарубежные и российские инструменты создания и распространения семантических активов, необходимых для подготовки связанных данных

Представлены технологии хранения и распространения связанных открытых данных и семантических активов, используемых для их подготовки



Инструменты цифровой статистики



Российский опыт подготовки связанных статистических данных

Центр семантической интеграции

- ПРОЕКТ
- ЭКСПЕРТЫ
- БИБЛИОТЕКА
- МЕРОПРИЯТИЯ

Семантические
данные

Связывание и
интеграция

Семантическое
моделирование

Virtuoso Universal Server

Данные EuroSTAT

Данные по демографии

Измерения

FREQ - периодичность сбора показателя

GEO - Страна

TIME - Время сбора показателя

SEX - Пол человека

AGE - Возраст человека

INDIC_DE - Название показателя

Факты

Observation value - значение показателя

Доступные наборы

Population aged 15-34 by **educational**

Population aged 15-34 by **sex and age** (1 OC)

Population aged 30-34 with tertiary **educational**

Population and **employment**

Population as a percentage of EU28 populati

Population by age group

Population by sex and age

Population by sex and year of birth

GEO	2006	2007	2008	2009	2010	2011
Austria	1.7	1.7	1.7	1.7	1.7	1.7
Belgium	2.1	2.1	2.1	2.1	2.2	2.2
Bulgaria	1.5	1.5	1.5	1.5	1.5	1.5
Croatia	0.9	0.9	0.9	0.9	0.9	0.9
Cyprus	0.1	0.2	0.2	0.2	0.2	0.2
Czech Republic	2.1	2.1	2.1	2.1	2.1	2.1
Denmark	1.1	1.1	1.1	1.1	1.1	1.1
Estonia	0.3	0.3	0.3	0.3	0.3	0.3
European Union (current composition)	100	100	100	100	100	100
Finland	1.1	1.1	1.1	1.1	1.1	1.1

Visual dimensions
Select the two dimensions that define the table of the browser:
Column Headings: TIME
Rows (values in first column): GEO

Fixed dimension
Change
FREQ: Ann.

INDIC_DE: Population as a percentage of EU-28 population

Пользователь Экстра Помощь

Поиск по ресурсам

Data Selection

Select a part of the Dataset

Population as a...

Select Unit and Measurement

No attribute...
Attributes available for selection

OBS_VALUE

Configure the Dimensions

TIME

GEO

INDIC_DE

Population as a percentage of EU28 population

Jump to Legend Show MetaData

Austria
Belgium
Cyrus
Czech Republic
Denmark
Estonia
Finland
Greece
Spain

0 1 2 3 4 5 6 7 8 9 10

2013 2014 2015 2016 2017

Show Information About Selected Configuration Retrieved Data Export as

3.

Цели создания и распространения связанных открытых статистических данных с использованием семантических методов

**Повышение эффективности повторного
использования статистических данных**

**Расширение возможностей интеграции СОСД
и их совместной обработки**

**Обеспечение однозначной содержательной
интерпретации данных за счет их
формального описания**

3.

Приоритеты создания и распространения СОСД Росстата

Повышение осведомленности потребителей и поставщиков данных, улучшение возможностей обнаружения, поиска, навигации и сопоставления статистических данных, полученных из различных источников

Улучшение организации, интеграции и предоставления данных в сети Интернет путем структуризации ландшафта статистических данных с использованием семантических методов

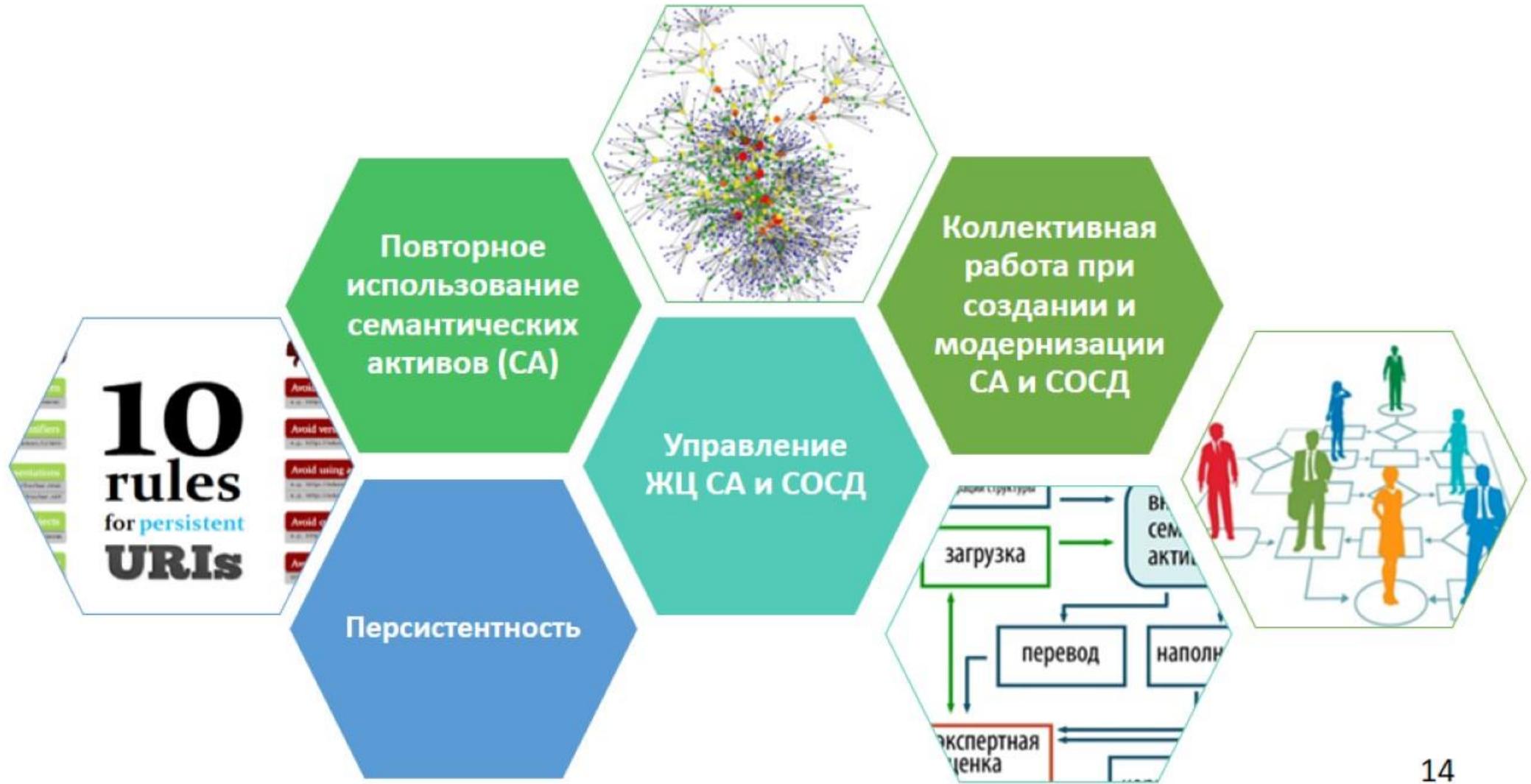
Унификация взаимодействия государственных органов и организаций государственного сектора, обеспечение совместного использования данных, повышение уровня семантической интероперабельности государственных информационных систем

Сокращение сроков и снижение затрат на проведение анализа статистических данных за счет уменьшения объема ручного труда при подготовке данных для обработки и семантической интеграции

Создание базовых семантических информационных ресурсов, отработка методов и технологий для распространения культуры применения связанных данных на все области государственного управления

3

Принципы создания и распространения связанных открытых статистических данных



3.

Система мероприятий по реализации Концепции

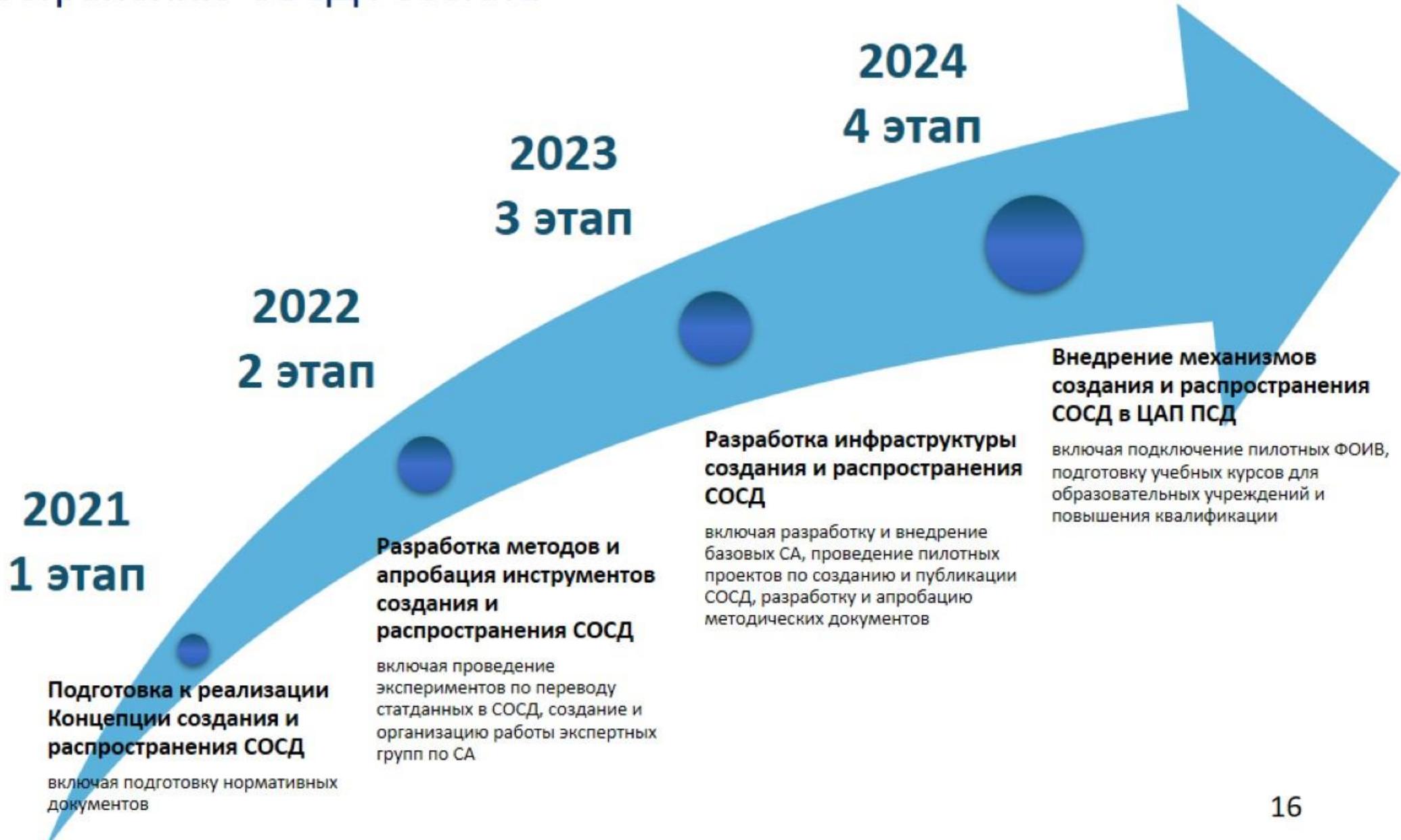
- ❖ Методики
- ❖ Правила
- ❖ Рекомендации
- ❖ Открытые стандарты
- ❖ Экспертное сообщество
- ❖ Международное сотрудничество

- ❖ Образовательные программы
- ❖ Повышение квалификации
- ❖ Мультимедийные ресурсы
- ❖ Популяризация



- ❖ Платформа коллективной работы подготовки и распространения СОСД
- ❖ Портал публикации и распространения СОСД
- ❖ Инструменты и сервисы взаимодействия с НСУД и ЦАП PSD

Предложения по основным этапам дорожной карты по созданию и распространению СОСД Росстата



Работы второго этапа НИР

Разработка предложений по нормативному правовому и нормативно-техническому регулированию подходов к подготовке, хранению и распространению связанных открытых статистических данных Росстата с учетом применения семантических методов

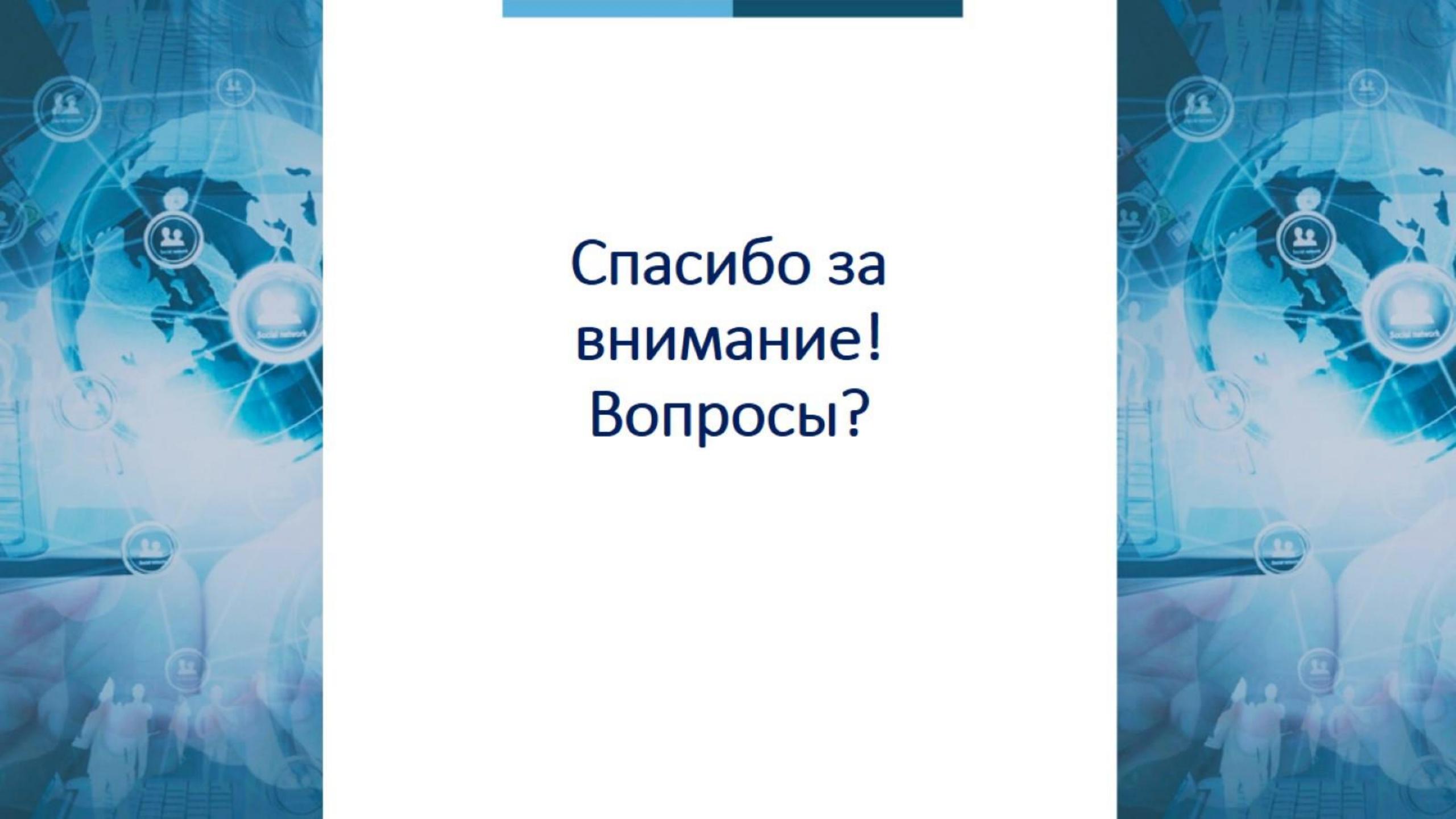
Разработка предложений по научно-методическим подходам к формированию связанных открытых статистических данных Росстата с учетом применения семантических методов

Определение технологических подходов к подготовке связанных статистических данных Росстата с использованием семантических активов (тезаурусов, онтологий, моделей предметных областей и других семантических моделей), а также к их хранению и распространению.

Выбор и обоснование технологий для реализации предлагаемой концепции

Доработка проекта концепции по созданию и распространению связанных открытых статистических данных Росстата

Доработка дорожной карты по созданию и распространению связанных открытых статистических данных Росстата

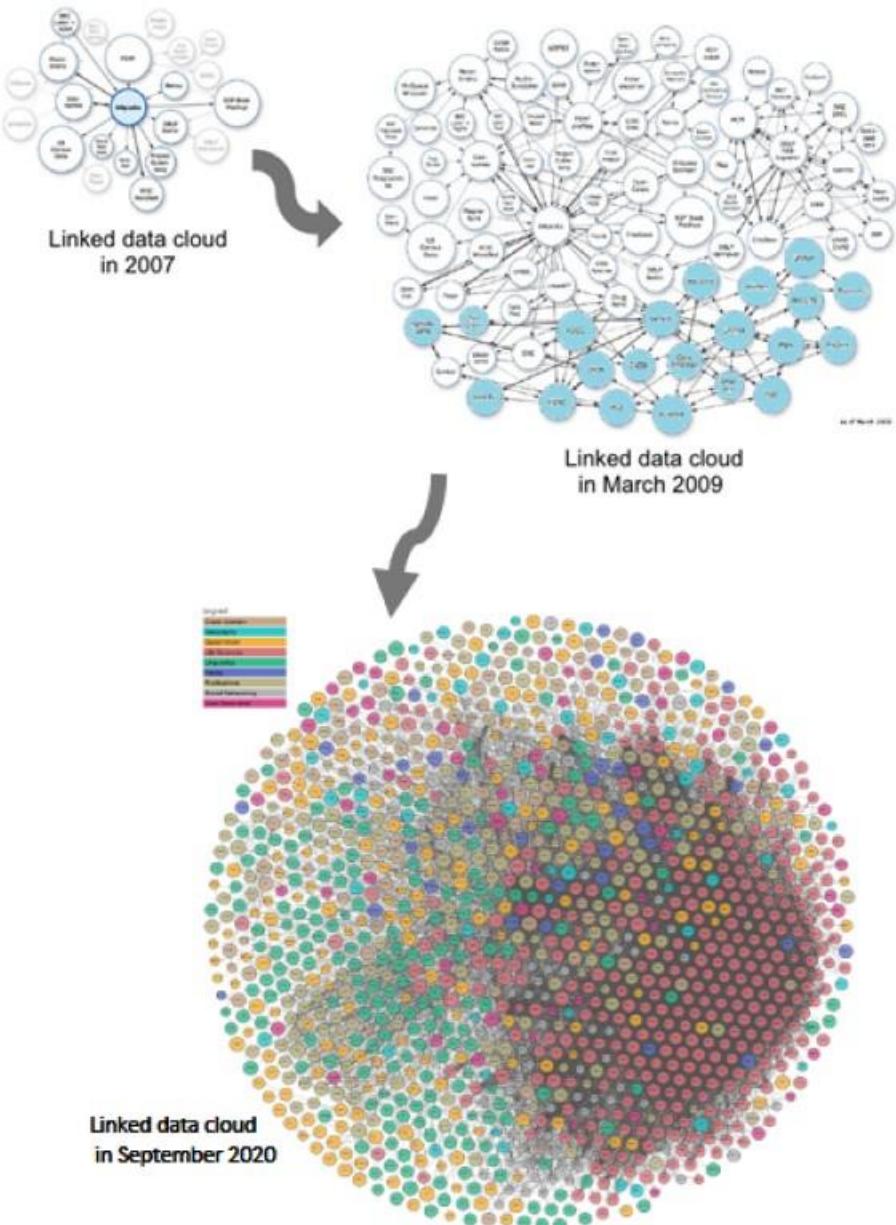


Спасибо за
внимание!
Вопросы?

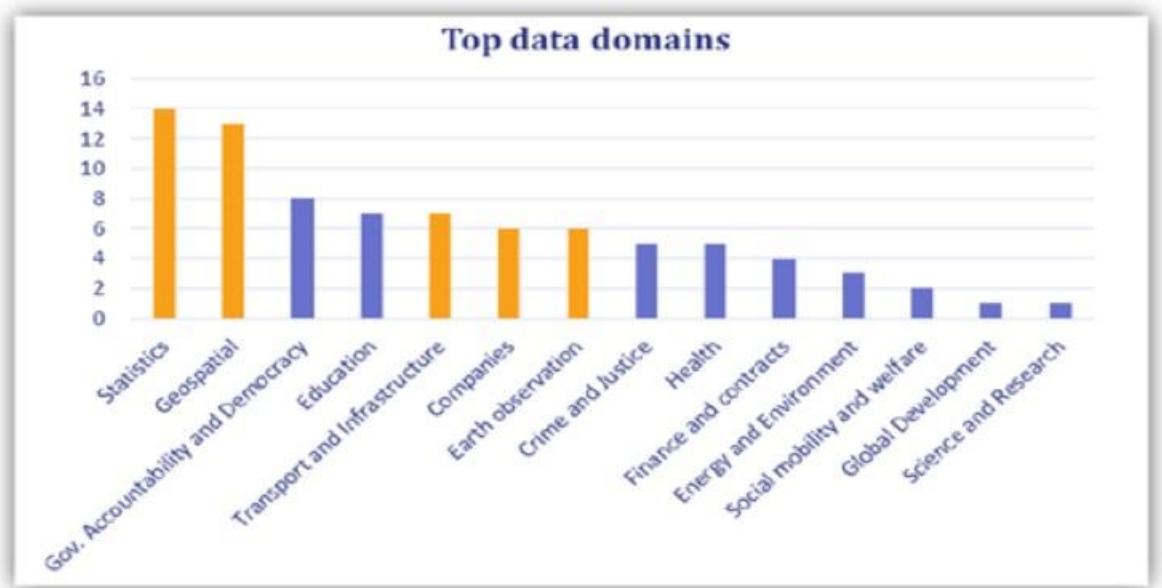
Дополнительные материалы



Развитие связанных данных



Top data domains consulted by the EU28+ countries on European Data Portal



Статистика – ключевой домен
государственных данных

Ожидаемые результаты реализации Концепции

Нормативный уровень

Разработаны и приняты нормативные и правовые акты, регулирующие создание, распространение и применение СОСД.

Организационный уровень

Определены функции и полномочия структурных подразделений в Росстата по подготовке и публикации СОСД. Сформирован Центр компетенций, в рамках которого обеспечивается коллективная работа ИТ-специалистов и экспертов по созданию и распространению СОСД, а также их публикации, визуализации и анализу для продвижения принципов СОД, включая привлечение заинтересованных потребителей СОСД. Организовано экспертное сообщество, внедрена методика и практика коллективной работы экспертов, распространение лучших практик.

Методический уровень

Разработаны и приняты методические документы, регламентирующие процессы разработки и подготовки СОСД. Разработаны базовые обучающие курсы для студентов и аспирантов профильных ВУЗов в рамках программ высшего образования, программы повышения квалификации специалистов в области статистики. Разработаны мультимедийные ресурсы для популяризации идей СОСД и обучения пользователей.

Семантический уровень

Исследованы, загружены, локализованы и/или разработаны семантические активы, включая международные, которые используются при создании, распространении и публикации СОСД. Разработаны и внедрены базовые СА. Проведены pilotные проекты по созданию и публикации СОСД.

Технический уровень

Разработаны, адаптированы и внедрены инструменты создания и распространения связанных открытых статистических данных, создана и внедрены платформы подготовки СА и СОСД, а также необходимые инструменты и сервисы. Механизмы создания и распространения СОСД интегрированы в ЦАП ПСД. Подключены pilotные ФОИВ.

4. Подготовка к реализации Концепции создания и распространения СОСД

Нормативное регулирование

1. Утверждение нормативных документов, регулирующих деятельность по созданию и развитию СОСД Росстата.

Организационно-методическое обеспечение

1. Утверждение Положения и создание Центра Компетенций СОСД.
2. Утверждение принципов и правил формирования экспертного сообщества.
3. Разработка и утверждение базовых методических документов:
 - Методика создания и каталогизации СА;
 - Методика каталогизации и связывания наборов данных;
 - Правила персистентности.
4. Разработка и согласование технических заданий на проведение работ 3 этапа.

Кадры

1. Формирование рабочей группы (или подразделения Росстата).
2. Разработка базовых обучающих материалов.
3. Проведение обучения участников рабочей группы.
4. Формирование экспертного сообщества, привлечение внешних экспертов, создание и организация работы экспертных групп по СА.
5. Развитие международного сотрудничества.

Разработка методов и апробация инструментов создания и распространения СОСД

Нормативное регулирование

1. Разработка и утверждение концепций и положений о платформах, инструментах и сервисах инфраструктуры создания и распространения СОСД.
2. Разработка и утверждение регламентирующих документов о пилотных проектах и их участниках.

Организационно-методическое обеспечение

1. Разработка и утверждение прикладных методик работы со связанными данными:
 - Методика управления ЖЦ СА и управления изменениями;
 - Методика коллективной работы экспертов и ИТ-специалистов.
2. Разработка и согласование технических заданий на проведение работ 4 этапа.

Кадры

1. Разработка обучающих материалов по базовым методикам.
2. Разработка мультимедийных ресурсов для популяризации идей СОСД, обучения пользователей, а также привлечения заинтересованных потребителей СОСД.
3. Организация обучающих курсов и привлечение пилотных групп обучающихся (экспертов, поставщиков СОСД, заинтересованных потребителей СОСД).

Разработка методов и аprobация инструментов создания и распространения СОСД

2 этап

Формирование исследовательских компетенций и технологических заделов

1. Исследование существующих методов подготовки и распространения СОСД
2. Исследование, выбор и аprobация инструментов

Информационная инфраструктура

1. Разворачивание тестовой инфраструктуры для исследования и аprobации инструментов и методов
2. Разработка и согласование требований к платформам, инструментам и сервисам.
3. Разработка и согласование технических заданий.
4. Проведение экспериментов по переводу статданных в СОСД и разработка демонстрационных примеров.

3 этап

Нормативное регулирование

1. Разработка и утверждение концепций и положений о платформах, инструментах и сервисах инфраструктуры создания и распространения СОСД.
2. Разработка и утверждение регламентирующих документов о пилотных проектах и их участниках.

Организационно-методическое обеспечение

1. Разработка и утверждение прикладных методик работы со связанными данными:
 - Методика управления ЖЦ СА и управления изменениями;
 - Методика коллективной работы экспертов и ИТ-специалистов.
2. Разработка и согласование технических заданий на проведение работ 4 этапа.

Кадры

1. Разработка обучающих материалов по базовым методикам.
2. Разработка мультимедийных ресурсов для популяризации идей СОСД, обучения пользователей, а также привлечения заинтересованных потребителей СОСД.
3. Организация обучающих курсов и привлечение пилотных групп обучающихся (экспертов, поставщиков СОСД, заинтересованных потребителей СОСД).

Разработка инфраструктуры создания и распространения СОСД

Формирование исследовательских компетенций и технологических заделов

1. Разработка и внедрение базовых СА.
2. Проведение пилотных проектов по созданию и публикации СОСД.
3. Апробация инструментов визуализации и анализа для демонстрации возможностей СОСД.

Информационная инфраструктура

1. Эскизное и техническое проектирование.
2. Проведение опытной эксплуатации.
3. Разворачивание пилотных проектов и демонстрационных примеров.

Внедрение механизмов создания и распространения СОСД в Цифровую аналитическую платформу предоставления статистических данных

Нормативное регулирование

1. Разработка и утверждение документов, регламентирующих СОСД, в т.ч. Разработка изменений в Методические рекомендации по публикации открытых данных ГО и ОМСУ от 29 мая 2014 г. № 4 в части связывания статистических данных.
2. Внесение изменений в действующие документы НСУД и ЦАП, регулирующих взаимодействие с инфраструктурой СОСД.
3. Внесение изменений в действующие документы, регулирующие сбор и публикацию статистических данных ФОИВ (для пилотных ФОИВ).

Организационно-методическое обеспечение

1. Формирование постоянного пула экспертов Центра компетенций.
2. Организация рабочих групп, ведущих коллективную работу по подготовке и распространению СОСД в областях деятельности пилотных ФОИВ.
3. Доработка методических документов по итогам разворачивания пилотных проектов и их утверждение.
4. Формирование пула заинтересованных потребителей СОСД, проведение просветительских мероприятий – семинаров, обучающих программ, продвигающих возможности применения СОСД в различных приложениях и сервисах.

Кадры

1. Разработка обучающих материалов по прикладным методикам.
2. Подготовка образовательных программ среднего специального, высшего и дополнительного образования статистических специалистов, ИТ-специалистов и административных работников.
3. Разработка базовых обучающих курсов (учебных пособий) для студентов и аспирантов профильных ВУЗов в рамках программ высшего образования, программы повышения квалификации специалистов в области статистики.
4. Разработка мультимедийных ресурсов для популяризации идей СОСД, обучения пользователей, а также привлечения заинтересованных потребителей СОСД.

Внедрение механизмов создания и распространения СОСД в Цифровую аналитическую платформу предоставления статистических данных

4 этап

Формирование исследовательских компетенций и технологических заделов

1. Исследование перспективных методов и инструментов создания и распространения СОСД. Подготовка предложений по развитию СОСД.
2. Разработка различных вариантов визуализации и анализа для демонстрации возможностей СОСД, в том числе для привлечения заинтересованных потребителей СОСД.

Информационная инфраструктура

1. Разработка программных платформ и сервисов, запуск в опытную эксплуатацию
2. Разработка и внедрение семантических активов в ходе работы экспертного сообщества и специалистов по подготовке СОСД.
3. Внедрение механизмов создания и распространения связанных открытых статистических данных в ЦАП ПСД.
4. Подключение пилотных ФОИВ.