

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ

по публикации открытых данных государственными органами,
органами местного самоуправления и подведомственными им организациями

версия 4.0

Основные понятия и глоссарий

Виджет (widget) — это небольшой графический элемент, вставляемый на страницу сайта для отображения какой-либо периодически обновляемой информации.

Метаданные, метаинформация (Meta) — структурированные данные, представляющие собой характеристики описываемых сущностей для целей их идентификации, поиска, оценки, управления этими сущностями.

Набор открытых данных (ОД), набор данных (Dataset) — систематизированная совокупность однотипных машиночитаемых данных, представленных в форме открытых данных, состоящая из отдельных элементов, характеризующихся набором атрибутов, и позволяющая автоматизированным системам без участия человека идентифицировать, интерпретировать и обрабатывать такие элементы.

Открытые данные (Opendata) — информация, размещенная в сети «Интернет» в виде систематизированных данных, организованных в формате, обеспечивающем ее автоматическую обработку без предварительного изменения человеком, в целях неоднократного, свободного и бесплатного использования.

Паспорт набора открытых данных (Passport) - совокупность сведений о наборе открытых данных, необходимых для установления факта принадлежности набора открытых данных в соответствии с выбранными ключевыми словами его потенциальной пригодности для решения задач потребителя, а также установления адреса размещения, способа загрузки и последующей автоматической обработки набора открытых данных.

Портал открытых данных Российской Федерации (далее – Портал ОД) – комплексная государственная информационная система, которая реализует функцию предоставления неограниченному кругу лиц доступа к реестру наборов ОД органов государственной власти Российской Федерации, органов местного самоуправления и организаций, действующих на территории Российской Федерации, а также иные функции, обеспечивающие реализацию политики Российской Федерации в области публикации открытых данных.

Реестр наборов ОД (Reestr) — систематизированный перечень наборов ОД, позволяющий в автоматическом режиме осуществлять навигацию по наборам открытых данных и получать ключевые параметры этих наборов, включая гиперссылки, необходимые для доступа к наборам открытых данных.

Сводный реестр наборов ОД – систематизированный перечень наборов ОД, публикуемых органами государственной власти Российской Федерации, органами местного самоуправления и организациями, действующими на территории Российской Федерации, содержащийся на Портале ОД.

Формат данных (Format) — структура данных (спецификация) файла, предназначенная для представления данных и пригодная для машинной обработки.

CSV — текстовый формат, предназначенный для представления табличных данных. Каждая строка файла — это одна строка таблицы. Значения отдельных колонок отделяются друг от друга установленным в системе разделительным символом.

JSON — текстовый формат обмена данными, основанный на JavaScript и обычно используемый именно с этим языком. JSON легко читается и обрабатывается.

RDF — модель, разработанная консорциумом W3C для представления данных, в особенности — метаданных. RDF представляет сведения о ресурсах в пригодном виде.

XML — язык разметки, рекомендованный Консорциумом Всемирной паутины (W3C). Спецификация XML описывает XML-документы и частично описывает поведение XML-процессоров (программ, читающих XML-документы и обеспечивающих доступ к их содержимому).

Оглавление

1 Общие положения	6
2 Подготовка наборов ОД	10
3 Требования к набору, паспорту, структуре	11
3.1 Форматы и их особенности	13
3.1.1 XML формат	13
3.1.1.1 Признаки для выбора формата	14
3.1.1.2 Типовые ошибки при работе с форматом	14
3.1.1.3 Основные правила при составлении XML вручную	16
3.1.1.4 Основные правила составления структуры в формате XSD	17
3.1.1.5 Чек-лист проверки набора в формате XML	18
3.1.2 CSV формат	19
3.1.2.1 Признаки для выбора формата	20
3.1.2.2 Типовые ошибки при работе с форматами	20
3.1.2.3 Основные правила при составлении CSV вручную	21
3.1.2.4 Основные правила в составлении структуры	23
3.1.2.5 Чек-лист проверки набора в формате CSV	23
3.1.3 JSON формат	24
3.1.3.1 Признаки для выбора формата	25
3.1.3.2 Типовые ошибки при работе с форматом	25
3.1.3.3 Основные правила при составлении JSON вручную	26
3.1.3.4 Основные правила составления структуры	27
3.1.3.5 Чек-лист проверки набора в формате JSON	28
3.1.4 RDF формат	28
3.1.4.1 Признаки выбора для формата	29
3.1.4.2 Типовые ошибки при работе с форматом	29
3.1.4.3 Основные правила при составлении RDFa вручную	29
3.1.4.4 Основные правила составления структуры	31
3.1.4.5 Чек-лист проверки набора в формате RDFa	31
4 Работа с наборами ОД	33
4.1 Функциональные возможности работы с наборами на Портале ОД	33
4.2 Уровень подготовки пользователей	33
4.3 Перечень документов к предварительному ознакомлению	33
4.4 Рабочее место и дистрибутивы	34
5 Вспомогательные материалы	36
5.1 Эталонные примеры наборов ОД и структуры	36
5.2 Эталонный пример паспорта набора ОД	38
6 Публикация наборов ОД на Портале ОД	40

- 6.1 Определение и границы роли «Поставщик»40
 - 6.1.1 Регистрация/авторизация/верификация Поставщика40
 - 6.1.2 Карточка Поставщика открытых данных42
- 6.2 Этапы публикации46
 - 6.2.1 Паспорт набора ОД47
 - 6.2.2 Набор ОД и его структура48
 - 6.2.3 Форматно-логический контроль52
 - 6.2.4 Статусы набора ОД53
 - 6.2.5 Модерация набора ОД54
 - 6.2.6 Статистика и показатели по набору ОД56
 - 6.2.7 Изменение/обновление набора ОД57
 - 6.2.8 Снятие с публикации57
- 6.3 Варианты добавления наборов ОД58
 - 6.3.1 Ручной вариант добавления наборов ОД58
 - 6.3.1.1 Сценарий добавления набора ОД через Портал ОД58
 - 6.3.1.2 Отчёт о валидации набора ОД59
 - 6.3.2 Автоматизированный (HTML-парсер)59
 - 6.3.2.1 Правила ведения HTML страницы открытых данных в ГИС Поставщика59
 - 6.3.2.2 Сценарий добавления набора через HTML-парсер;62
 - 6.3.2.3 Отчёт об итогах парсинга64
 - 6.3.3 Автоматизированный64
 - 6.3.3.1 Структура запроса данных в формате json;64
 - 6.3.3.1 Процедура получения доступа к интеграции66

1 Общие положения

1.1. Методические рекомендации по публикации открытых данных государственными органами, органами местного самоуправления и подведомственными им организациями (далее соответственно – Методические рекомендации, органы публичной власти и подведомственные им организации) предназначены для ответственных сотрудников органов публичной власти и подведомственных им организаций при подготовке указанными сотрудниками наборов ОД и их последующей публикации в информационно-телекоммуникационной сети «Интернет» (далее – сеть Интернет), включая сопровождение наборов ОД после публикации.

1.2. Публикация информации в форме открытых данных обеспечивает их хранение и использование в удобном формате, допускающем их автоматизированную обработку и анализ, актуальность, достоверность и доступность.

1.3. Набор ОД представляет собой файл в машиночитаемом формате, предлагаемом в соответствии с Методическими рекомендациями, содержащий информацию о деятельности органов публичной власти и подведомственных им организаций. Наборы ОД представляют собой структурированную информацию, представленную в открытом формате и доступны для свободного использования, анализа и распространения. Вместе с набором ОД в соответствии с разделом 2 Методических рекомендаций подготавливается файл со структурой набора ОД и паспорт набора ОД.

1.4. Для подготовки наборов ОД используется общедоступная информация о деятельности органов публичной власти и подведомственных им организаций, которая была отнесена к таковой в соответствии с законодательством Российской Федерации и соответствующими решениями органов публичной власти и подведомственных им организаций. Включение в наборы ОД информации ограниченного доступа не допускается.

1.5. Отнесение информации о деятельности органов публичной власти и подведомственных им организаций к общедоступной информации осуществляется в соответствии с:

- Федеральным законом от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации»;
- Федеральным законом от 9 февраля 2009 г. № 8-ФЗ «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления»;
- Постановлением Правительства Российской Федерации от 24 ноября 2009 г. № 953 «Об обеспечении доступа к информации о деятельности Правительства Российской Федерации и федеральных органов исполнительной власти»;
- Постановлением Правительства Российской Федерации от 10 июля 2013 г. № 583 «Об обеспечении доступа к общедоступной информации о деятельности государственных

органов и органов местного самоуправления в информационно-телекоммуникационной сети «Интернет» в форме открытых данных»;

– Распоряжением Правительства Российской Федерации от 10 июля 2013 г. № 1187-р (О Перечнях информации о деятельности государственных органов, органов местного самоуправления, размещаемой в сети «Интернет» в форме открытых данных);

– Приказом Минэкономразвития России от 23 апреля 2024 г. № 247 «Об утверждении методических указаний по отнесению информации к общедоступной информации, размещаемой государственными органами и органами местного самоуправления на их официальных сайтах в информационно-телекоммуникационной сети «Интернет» в форме открытых данных» (далее – Приказ Минэкономразвития России № 247).

– иными нормативными правовыми актами, в соответствии с которыми информация о деятельности органов публичной власти и подведомственных им организаций признается общедоступной информацией и подлежащей размещению в сети «Интернет» в форме открытых данных.

1.6. Подготовка наборов ОД осуществляется с использованием информации о деятельности органов публичной власти и подведомственных им организаций, в отношении которой установлена ее общественная потребность. Установление общественной потребности в информации о деятельности органов публичной власти и подведомственных им организаций осуществляется в порядке, определенном Методическими указаниями по отнесению информации к общедоступной информации, размещаемой государственными органами и органами местного самоуправления на их официальных сайтах в информационно – телекоммуникационной сети «Интернет» в форме открытых данных, утвержденных Приказом Минэкономразвития России от 23 апреля 2024 г. № 247.

1.7. Основные характеристики открытых данных:

- Доступность: Данные доступны для всех без исключения, что позволяет любому заинтересованному лицу получить доступ к информации.
- Машиночитаемость: Данные представлены в формате, который может быть обработан компьютерами без необходимости ручного ввода.
- Открытые лицензии: Данные доступны для использования, модификации и распространения без ограничений, что стимулирует инновации и развитие.
- Актуальность: Данные регулярно обновляются, чтобы обеспечивать их релевантность и точность.

1.8. Публикация наборов ОД в сети Интернет осуществляется органами публичной власти и подведомственными им организациями следующими рекомендованными способами:

- в специальном разделе для публикации наборов ОД официальных сайтов в сети Интернет (далее – официальные сайты);
- на Портале ОД Российской Федерации.
- на порталах открытых данных, а также в государственных и муниципальных информационных системах, в информационных системах органов публичной власти и подведомственных им организаций (иные ресурсы для публикации наборов ОД).

При публикации наборов ОД на официальных сайтах и иных ресурсах по публикации наборов ОД они должны быть продублированы на Портале ОД. Дублирование наборов ОД осуществляется посредством функционала Портала ОД в автоматическом режиме при условии строго соблюдения установленных Методическими рекомендациями требований к набору ОД, в том числе технических требований к форматам, установленных в разделе 3.

На официальном сайте и иных ресурсах для публикации наборов ОД может быть размещена ссылка на страницу Портала ОД, где опубликован соответствующий набор ОД.

1.9. Органы публичной власти и подведомственные им организации определяют ответственных сотрудников за подготовку и публикацию наборов ОД в сети Интернет (далее – ответственные сотрудники).

Информация об ответственных сотрудниках является общедоступной и подлежит ежегодной актуализации органами публичной власти и подведомственными им организациями.

Информация об ответственных сотрудниках включает: ФИО, должность, структурное подразделение, контактный телефон и адрес электронной почты.

Информация об ответственном сотруднике для каждого набора ОД включается в паспорт набора ОД (заполняется для каждого набора ОД перед его публикацией в сети Интернет).

1.10. Доступ к наборам ОД предоставляется неограниченному кругу лиц.

Ответственные сотрудники в отношении каждого публикуемого в сети Интернет набора ОД определяют целесообразность идентификации пользователей, которые получают доступ к наборам ОД. Идентификация пользователей наборов ОД может осуществляться в целях проведения статистического учета востребованности наборов ОД. Для идентификации пользователей наборов ОД рекомендуется использовать функционал Федеральной государственной информационной системы «Единая система идентификации аутентификации в инфраструктуре, обеспечивающей информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных и муниципальных услуг в электронной форме» (ЕСИА).

1.11. Ответственные сотрудники обеспечивают актуальность наборов ОД в соответствии со сроками их актуализации, указанными в паспортах наборов ОД, а также осуществляют взаимодействие с оператором Портала ОД, а также пользователями официальных сайтов, Портала

ОД и иных ресурсов для публикации наборов ОД при уточнении, актуализации, внесении изменений, приостановке доступности наборов ОД.

1.12. Органы публичной власти и подведомственные им организации ведут реестры наборов ОД, которые в машиночитаемом формате файла размещаются в специальных разделах с открытыми данными официальных сайтов и иных ресурсов для публикации наборов ОД.

В реестры наборов ОД включается информация о всех наборах ОД, опубликованных в сети «Интернет» органами публичной власти и подведомственными им организациями.

1.13. Сведения, включенные в реестры наборов ОД, должны передаваться в сводный реестр наборов ОД, который ведется оператором Портала ОД.

2. Подготовка наборов ОД

2.1. Подготовка набора ОД включает в себя подготовку:

2.1.1. файла с набором ОД в машиночитаемом формате, содержащего общедоступную информацию о деятельности органа публичной власти или подведомственной им организации;

2.1.2. файла со структурой набора ОД;

2.1.3. паспорта набора ОД.

2.2. Подготовка наборов ОД может осуществляться в автоматическом режиме посредством загрузки необходимой информации о деятельности органов публичной власти и подведомственных им организаций в файл с набором ОД, а также в ручном режиме, посредством создания файла с набором ОД.

2.3. Файлы с наборами ОД могут быть подготовлены в одном из машиночитаемых форматов из раздела 3.1 Методических рекомендаций.

2.4. Выбор машиночитаемого формата файла осуществляется ответственными сотрудниками самостоятельно.

2.4.1. Подготовка наборов ОД в машиночитаемом формате XML описана в разделе 3.1.1 Методических рекомендаций.

2.4.2. Подготовка наборов ОД в машиночитаемом формате CSV описана в разделе 3.1.2 Методических рекомендаций.

2.4.3. Подготовка наборов ОД в машиночитаемом формате JSON описана в разделе 3.1.3 Методических рекомендаций.

2.5. Подготовка наборов ОД в машиночитаемом формате RDFa описана в разделе 3.1.4 Методических рекомендаций.

2.6. Паспорт ОД - совокупность сведений о наборе ОД. Паспорт набора ОД имеет требования, установленные в разделе 3 Методических рекомендаций.

2.7. Публикация набора ОД может быть осуществлена на Портале ОД ручным, автоматическим или автоматизированным способами (согласно разделу 6.3 Методических рекомендаций).

3 Требования к набору, паспорту, структуре

Каждый набор ОД сопровождается метаинформацией, которая описывает его структуру, источники, методы сбора и другие характеристики. Метаинформация помогает пользователям лучше понять контекст данных и правильно интерпретировать их при анализе.

Опубликование информации в форме открытых данных необходимо осуществлять поэтапно, учитывая при этом следующие факторы:

1. востребованность соответствующих наборов ОД потенциальными потребителями информации, оцениваемую в соответствии с *методикой мониторинга и оценки востребованности открытых данных*;
2. степень готовности наборов ОД, характеризующуюся наличием необходимых данных в электронном виде, а также готовностью организационных, технических, технологических и иных средств, необходимых для опубликования наборов ОД в порядке и формате, определяемых Методическими рекомендациями;
3. затраты на публикацию (финансовые, временные, трудовые), необходимые для опубликования наборов ОД и поддержания их в актуальном состоянии.

Сведения, обладающие высокой востребованностью и степенью готовности, требующие при этом минимальных затрат, должны публиковаться в форме открытых данных государственными органами и органами местного самоуправления в первоочередном порядке.

При формировании как реестра наборов ОД, так и самих наборов ОД, следует учитывать:

1. публикации подлежат общедоступная информация;
2. для каждого набора данных должно быть определено должностное лицо, ответственное по вопросам публикации открытых данных;
3. для каждого набора данных должна быть установлена периодичность обновления с учетом потребностей потребителей открытых данных.

Содержание файла с набором ОД, файла со структурой набора ОД и паспорт набора ОД должны соответствовать минимальным требованиям (в зависимости от формата файла, в котором они были сделаны), установленными в разделе 3 Методических рекомендаций.

Атрибутивный состав паспорта набора ОД должен включать:

- Идентификационный номер (ID набора ОД)
- Наименование набора ОД
- Описание набора ОД
- Владелец набора ОД
- Ответственное лицо
- Телефон ответственного лица

- Адрес электронной почты ответственного лица
- Гиперссылка (URL) на открытые данные
- Формат набора ОД
- Описание структуры набора ОД
- Дата первой публикации набора ОД
- Дата последнего внесения изменений
- Содержание последнего изменения
- Дата актуальности набора данных
- Ключевые слова, соответствующие содержанию набора данных
- Гиперссылки (URL) на версии ОД
- Гиперссылки (URL) на версии структуры набора

Сформированный реестр ОД утверждается государственным органом или органом местного самоуправления и подлежит публикации на официальном сайте органа в сети Интернет.

Атрибутивный состав информации в реестрах наборов ОД, утвержденных государственным органом или органом местного самоуправления, должен включать:

- наименование органа публичной власти и подведомственной им организации;
- категорию владельца набора ОД (государственный орган, подведомственная организация государственного органа, муниципальный орган, подведомственная организация муниципального органа);
- статус набора ОД (действующий, переведен в архив);
- наименование органа публичной власти (указывается для подведомственных организаций);
- порядковые номера наборов ОД, опубликованных в сети Интернет;
- ID набора ОД;
- указание на принадлежность наборов ОД к определенным категориям открытых данных;
- наименования наборов ОД;
- дату первичной публикации наборов ОД;
- источник первичной публикации наборов ОД (официальный сайт, Портал ОД, иной ресурс для публикации наборов ОД (указывается официальное наименование источника));
- ссылку на первоисточник опубликования наборов ОД;
- периодичность актуализации для наборов ОД;
- дату последней актуализации набора ОД;
- статус размещения наборов ОД на Портале ОД (да, нет);
- ссылку на наборы ОД, размещенные на Портале ОД;

- ФИО, должность, структурное подразделение, контактный телефон и адрес электронной почты ответственного сотрудника за подготовку и размещение наборов ОД в сети Интернет;
- количество скачиваний наборов ОД;

3.1 Форматы и их особенности

Машиночитаемые данные - данные, представленные в описанном формате, позволяющем без необходимости ручного ввода идентифицировать, обрабатывать, преобразовывать такие данные и их составные части (элементы), а также обеспечивать доступ к ним.

Машиночитаемое представление открытых данных должно быть реализовано в одном из следующих форматов:

- XML;
- CSV;
- JSON;
- RDFa;

3.1.1 XML формат

XML (eXtensible Markup Language – расширяемый язык разметки) — язык разметки, который используется для хранения и передачи данных. Его основное назначение — структурирование информации в текстовом виде, чтобы она была легко читаемой как человеком, так и машиной. XML позволяет описывать данные через набор тегов (могут быть заданы разработчиком), которые могут быть вложены друг в друга и могут содержать атрибуты, что делает структуру данных гибкой и расширяемой. Это универсальный формат, поддерживаемый множеством программ и библиотек, что делает его удобным для обмена данными между разными системами.

В отличие от языков программирования, XML не может выполнять вычислительные операции сам по себе. Вместо этого для управления структурированными данными используется любой язык программирования или программное обеспечение.

Формат XML часто используется для передачи данных по API, причем является единственно возможным форматом данных в SOAP API.

Текстовый документ с расширением .xml может быть открыт с помощью следующих инструментов:

- Microsoft Excel;
- Microsoft Word;
- блокнота Windows;
- любого интернет-браузера;
- редакторов XML (например, XML Editor).

3.1.1.1 Признаки для выбора формата

Формат XML хорошо подходит для структурированных данных, если важна иерархическая структура и метаданные.

1. Сложные иерархические данные — если данные имеют несколько уровней вложенности (например, данные о компаниях, содержащие информацию о филиалах, сотрудниках, их ролях и т.д.).

2. Метаданные и схемы — если данные включают много метаданных, таких как описания, единицы измерения, ссылки на источники и другие атрибуты. В формате XML легко определяются схемы (XSD), которые обеспечат строгость и валидацию данных, что важно для гарантии их качества.

3. Совместимость и интеграция — если набор данных должен быть интегрирован с другими системами, которые ожидают данные в формате XML (например, веб-сервисы, SOAP).

4. Малый объем данных — для небольших наборов данных, где производительность не является критическим фактором, XML может быть удобен благодаря своей читабельности и доступности для обработки различными инструментами. Однако для больших объемов данных (big data) XML может быть менее эффективным из-за увеличенного размера файла по сравнению с другими форматами, такими как CSV или JSON.

Например, если необходимо опубликовать набор данных о демографической статистике с детализированной информацией по регионам, возрастным группам, профессиям, уровням образования и т.д., где каждый уровень требует своей метки и вложенности, XML будет хорошим выбором, чтобы сохранить иерархию и метаданные.

В общем случае XML лучше всего подходит для структурированных, иерархически организованных данных, требующих метаданных и строгой валидации. Для простых числовых данных или big data лучше выбирать другие форматы, такие как CSV или JSON, которые обеспечивают большую компактность и производительность.

3.1.1.2 Типовые ошибки при работе с форматом

Существует несколько правил «well formed» XML, то есть его синтаксической корректности:

1. Все XML документы должны иметь один корневой элемент. Отсутствие корневого элемента делает документ недействительным (в данном примере элемент <credential> является корневым и содержит данные об учетных записях).

Ошибка	Верный вариант
<pre> <test> <user>Username</user> <pass>123</pass> </test> <dev> <user>Антон</user> <pass>123</pass> </dev> </pre>	<pre> <credential> <test> <user>Username</user> <pass>123</pass> </test> <dev> <user>Антон</user> <pass>123</pass> </dev> </credential> </pre>

2. У каждого элемента должен быть закрывающий тег. Кроме того, имена тегов не должны содержать пробелов, специальных символов, и должны начинаться с буквы или подчеркивания.

Ошибка	Верный вариант
<User name> Тест	<Username> Тест </Username>

3. Учет регистрозависимости тегов.

Ошибка	Верный вариант
<pre> <Name>Тест</name> <NAME>Иван</name> <NAME>Тест</name> </pre>	<pre> <name>Тест</name> </pre>

4. Ошибки в структуре — неправильная вложенность тегов, где открытый тег внутри вложенного элемента не закрыт до окончания этого вложенного элемента.

Ошибка	Верный вариант
<pre> <fio>Иванов <name>Иван</fio> Иванович </name> </pre>	<pre> <fio>Иванов Иван Иванович</fio> <name>Иван</name> </pre>

5. Оформление атрибутов в кавычках, даже если они представляют из себя числовое значение.

Ошибка	Верный вариант
<pre> <query attr1=123>Виктор Иван</query> </pre>	<pre> <query attr1="123">Виктор Иван</query> </pre>

Также дополнительно в качестве одной из ошибок при работе с XML можно выделить несоответствие кодировки — например, в декларации указано UTF-8, а фактически используется другая кодировка.

Для проверки XML на синтаксис возможно использование любого XML Validator.

3.1.1.3 Основные правила при составлении XML вручную

Рассмотрим простой пример XML документа:

```
<?xml version="1.0" encoding="UTF-8"?>
<catalog>
  <book id="bk101">
    <author>John Snow</author>
    <title>Learning XML</title>
    <genre>Computer</genre>
    <price>39.95</price>
  </book>
</catalog>
```

1. Строка `<?xml version="1.0" encoding="UTF-8"?>` называется прологом и означает, что используется XML версии 1.0 и кодировка UTF-8. Пролог является первой строкой XML документа.

2. В XML каждый элемент заключен в теги. Тег — некий текст, заключенный в угловые скобки. Текст внутри тега — его название. Существует два типа тегов: открывающий `<catalog>` и закрывающий `</catalog>`.

3. В каждом XML документе должен быть только один корневой элемент, который включает все другие элементы (иными словами, это тег, с которого документ начинается и которым заканчивается). В приведенном примере корневым элементом является тег `<catalog>`.

4. Значение элемента хранится между открывающим и закрывающим тегами. Может быть числом, строкой, или вложенным тегом. В примере `<author>Alexander Pushkin</author>` элемент `<author>` указывает на автора книги, здесь это «Alexander Pushkin».

5. У элементов могут быть атрибуты — один или несколько. Они указываются внутри открывающегося тега после названия тега через пробел. В примере `<book id="bk101">` элемент `book` представляет одну книгу в каталоге, у него есть атрибут `id`, который обозначает уникальный идентификатор книги (в данном случае `bk101`).

6. Все теги в XML являются регистро-чувствительными. Если тег `<catalog>` закрыт как `</Catalog>`, документ будет оформлен некорректно.

7. В XML контролируется вложенность тегов, поэтому необходимо следить за порядком открывающих и закрывающих тегов.

8. Пробелы и переносы строк внутри текста тегов учитываются как часть данных, на это следует обращать внимание при форматировании.

Пространства имен

Так как в XML имена элементов создаются разработчиком, это может стать причиной конфликта наименований при попытке одновременного использования нескольких XML документов. Пространство имен — метод избегания таких конфликтов.

Согласно спецификации W3C пространство имен XML - это идентифицируемая с помощью ссылки URI ((англ. Uniform Resource Identifier) - унифицированный (единообразный) идентификатор ресурса. URI - последовательность символов, идентифицирующая абстрактный или физический ресурс) коллекция имен, используемых в XML документах для обозначения типов элементов и именования атрибутов.

При задании пространства имен в XML документе необходимо иметь в виду следующее:

1. Элементы и атрибуты связываются с пространством имен с помощью квалифицированных имен;
2. В одном XML документе может быть задано несколько пространств имен;
3. Область действия пространства имен может распространяться:
 - на весь документ (когда оно задано в корневом элементе XML документа);
 - контейнер (когда оно задано в элементе верхнего уровня контейнера);
 - отдельный элемент или атрибут (когда оно задано непосредственно для данного элемента и атрибута).

Для объявления пространства имен в XML документе используется ключевое слово `xmlns`. Синтаксис объявления имеет следующий формат: `xmlns[:prefix]="uri"`, где:

- `prefix` – это квалифицированное имя, посредством которого элементы и атрибуты XML документа связываются с пространством имен;
- `uri` – это URI, задающий пространство имен, которое будет использоваться в XML документе.

В том случае, если параметр `prefix` не задан, то объявленное пространство имен будет использоваться как пространство имен по умолчанию.

3.1.1.4 Основные правила составления структуры в формате XSD

1. Основы XSD схемы:

Для определения структуры данных (Data Schema) используется XSD (XML Data Definition). XSD — это стандартный язык, предназначенный для описания структуры XML документов. В XSD

определяется, какие элементы и атрибуты могут присутствовать в документе, их типы, допустимые значения, порядок и количество повторений.

2. Определение основного элемента:

Основной элемент схемы задается с использованием тега `<xs:element>`. Внутри него указывается имя элемента и тип, который определяет структуру и свойства данного элемента.

3. Использование сложных типов (Complex Types):

Сложные типы (`<xs:complexType>`) используются для определения элементов, которые могут содержать другие элементы или атрибуты.

Внутри `<xs:complexType>` можно использовать `<xs:sequence>`, `<xs:choice>` или `<xs:all>`, чтобы задать порядок или набор возможных подэлементов.

4. Определение простых типов (Simple Types):

Простые типы (`<xs:simpleType>`) используются для определения ограничений на значения элементов, например, строки, числа, перечисления возможных значений (`<xs:enumeration>`).

5. Аннотации (Annotations):

Аннотации (`<xs:annotation>`) используются для добавления описаний и документации к элементам и типам в XSD. Это помогает разработчикам и пользователям схемы лучше понимать ее структуру и назначение.

6. Определение атрибутов:

Атрибуты (`<xs:attribute>`) определяются внутри элементов и задают дополнительные свойства, такие как уникальные идентификаторы или другие данные, связанные с элементом.

3.1.1.5 Чек-лист проверки набора в формате XML

1. Общие проверки

Корректность синтаксиса:

- Все открывающие теги имеют соответствующие закрывающие теги.
- Все элементы правильно вложены (не допускается перекрытие тегов).
- Все атрибуты заключены в кавычки (двойные или одинарные).

Использование корневого элемента:

- Документ содержит один корневой элемент, который включает все остальные элементы.

2. Проверка структуры документа

Правильность структуры:

- Структура документа соответствует схеме (XSD)

Проверка обязательных элементов:

- Наличие всех обязательных элементов, которые указаны в схеме.

Порядок элементов:

- Элементы находятся в правильном порядке, если схема требует определенной последовательности (xs:sequence).

3. Валидация данных

Проверка типов данных:

- Значения элементов и атрибутов соответствуют их типам данных, определенным в схеме (например, xs:integer, xs:date, xs:string).

Ограничения значений:

- Данные удовлетворяют ограничениям, определенным в схеме (например, длина строки, диапазон чисел, регулярные выражения).

4. Проверка атрибутов

Наличие обязательных атрибутов:

- Все обязательные атрибуты присутствуют в соответствующих элементах.

Правильность значений атрибутов:

- Значения атрибутов соответствуют требованиям (например, тип данных, допустимые значения).

5. Проверка комментариев

Корректность синтаксиса комментариев:

- Все комментарии корректно закрыты (<!-- comment -->) и не содержат запрещенных символов (-- внутри комментария).

6. Проверка на корректность содержания

Корректность данных:

- Все данные в документе актуальны и корректны (например, даты, числовые значения).

Смысловая проверка:

- Содержимое документа логично и согласовано (например, отсутствие противоречий между связанными элементами).

3.1.2 CSV формат

CSV (от англ. Comma-Separated Values — значения, разделённые запятыми) — текстовый формат хранения данных, использующийся для обмена и преобразования информации различными программами работы с электронными таблицами, также применяется для импорта и экспорта данных.

Стандартный формат файла CSV определяется данными строк и столбцов, где новая строка завершает каждую строку, а каждый столбец внутри строки разделяется запятой.

Файлы данного формата имеют расширение «.csv».

CSV легко читается человеком и может быть отредактирован с помощью любого текстового редактора.

3.1.2.1 Признаки для выбора формата

Формат CSV оптимально использовать для данных, которые представляют собой табличную структуру, и при этом все строки таблицы содержат одинаковое количество полей.

Формат отлично подходит для публикации данных больших объемов, так как CSV примерно вдвое меньше форматов XML и JSON, вследствие чего возможно уменьшить объем хранимых файлов и увеличить скорость их передачи по сети.

Формат не подходит для представления сложных таблиц, в которых имеются многоуровневые шапки, объединение ячеек. Такие данные не могут быть корректно представлены с помощью CSV без потери сведений либо нарушения структуры файла самого CSV.

Также при помощи формата CSV не получится опубликовать вложенные объекты или массивы.

3.1.2.2 Типовые ошибки при работе с форматами

В данном разделе перечислены часто встречающиеся ошибки при создании файла набора.

1. Некорректная кодировка файла

Если файл набора сохранен в кодировке отличной от UTF-8, возможно отображение части данных в нечитаемом виде. Пример такой ошибки показан на рисунке ниже.

```
11 P"РмРгРsРiСѓРe;30652;1911;1940
12 P"Р·РмСѓР¶РёРsРсѓРeРёРm;57700;1921;1981
13 P"РjРёС,СѓРsРi;64307;1154;1154
14 P"РsР»РiРsРiСѓРсѓРгРsРс<Рm;119957;1931;1957
15 P"РsРjРsРгРмРгРsРiРs;155421;1900;1947
16 P"СѓРмР·РsР°;12246;1897;1940
17 P"СѓР±РsР°;74193;1956;1956
18 Р·РiРsСѓСѓРмРiСѓРe;71322;1462;1778
19 Р-СѓРeРsРiСѓРeРёРm;110507;1947;1947
20 Р-Р°СѓР°РmСѓРe;20419;1146;1778
21 Р-РiРмРsРёРiРsСѓРsРг;37154;1152;1781
22 РРiР°РsС,РмРмРiРeР°;83725;1564;1938
23 РсѓС,СѓР°;36085;1589;1781
24 РљР°СёРёСѓР°;45231;1356;1356
25 РљР»Р»РёРs;88525;1317;1781
26 РљРsР»РsРjРsР°;133019;1177;1177
27 РљРsСѓРsР»С`Рi;226936;1938;1938
28 РљРsС,РмР»СѓРsРёРeРё;68758;2004;2004
```

Чтобы избавиться от неё, следует изменить кодировку файла на UTF-8. Более подробно об этом описано в пункте 1.2. Методических рекомендаций.

После этого, файл станет полностью читаемым, как изображено ниже.

11	Дедовск;30652;1911;1940
12	Дзержинский;57700;1921;1981
13	Дмитров;64307;1154;1154
14	Долгопрудный;119957;1931;1957
15	Домодедово;155421;1900;1947
16	Дрезна;12246;1897;1940
17	Дубна;74193;1956;1956
18	Егорьевск;71322;1462;1778
19	Жуковский;110507;1947;1947
20	Зарайск;20419;1146;1778
21	Звенигород;37154;1152;1781
22	Ивантеевка;83725;1564;1938
23	Истра;36085;1589;1781
24	Кашира;45231;1356;1356
25	Клин;88525;1317;1781
26	Коломна;133019;1177;1177
27	Королёв;226936;1938;1938
28	Котельники;68758;2004;2004

2. Нарушение порядка атрибутов в структуре

Если допущена данная ошибка, то при дальнейшей обработке набора название, описания полей и формат могут быть перепутаны между собой. Чтобы этого не допустить, лучше опираться на эталонный пример структуры.

3. Несоответствие количества описанных полей в структуре и файле набора

Если количество описанных полей в структуре не будет равно количеству полей в файле набора, это приведёт либо к смещению описания полей относительно их расположения в файле, либо у некоторых полей может не оказаться описания и указания формата.

4. Неправильные заголовки атрибутов в структуре

Для корректного описания и анализа полей данных, атрибуты структуры должны иметь четко заданные заголовки и порядок, а именно:

- field name
- english description
- russian description
- format

Детально создание правильной структуры разобрано в разделе ниже.

3.1.2.3 Основные правила при составлении CSV вручную

CSV файл открытых данных может быть создан с помощью различных текстовых редакторов:

- «МойОфис Текст»

- Встроенного текстового редактора «Блокнот» в Windows или его аналогов в других операционных системах
- Свободно распространяемого текстового редактора с открытым исходным кодом «Notepad++»
- Иных текстовых редакторов

Также файл набора может быть создан сначала в табличном редакторе (в виде электронной таблицы), а после экспортирован уже в CSV. Для этого можно использовать следующие программные продукты:

- «МойОфис Таблицы»
- Р7-Офис
- Microsoft Excel
- OpenOffice Calc
- Иные табличные редакторы

При таком варианте необходимы дополнительные настройки при экспорте в части кодировки и разделителя.

В независимости от способа создания, для получения корректного файла формата csv необходимо соблюдать все нижеперечисленные правила:

1. Все строки должны содержать одинаковое количество полей

Поскольку формат файла предполагает его последующее табличное представление, каждая строка в документе олицетворяет собой строку таблицы.

Нарушение требования приведет к невозможности дальнейшего использования данных. Структурированность файла будет нарушена и по нему будет невозможно построить таблицу.

2. Последнее поле документа не должно завершаться запятой

В ином случае структурированность файла также может быть нарушена, так как будет ожидаться еще одна строка.

3. Ограничения по объему отсутствуют

Файл формата может содержать сколь угодно много записей, а каждое поле файла может содержать сколько угодно символов, включая пробелы.

4. Файл должен быть в кодировке UTF-8

Правильная кодировка обеспечивает корректное отображение всех символов. Неверная кодировка файла может привести к отображению вместо текстовых данных символов «неизвестного языка».

Кодировку файла может быть указана в программном продукте, который используется при создании файла.

Если файл уже создан, но его кодировка отлична от UTF-8, то её необходимо изменить с помощью любого доступного текстового редактора, который обладает соответствующими функциями.

5. Файл должен соответствовать стандарту Separated Values (CSV) Files»

Подробнее со стандартом можно ознакомиться по ссылке (<http://tools.ietf.org/html/rfc4180>)

3.1.2.4 Основные правила в составлении структуры

Основные правила формирования файла структуры, следующие:

- Структура набора ОД должна быть представлена отдельным файлом.
- Файл структуры набора ОД также должен быть сохранен в формате UTF-8.
- По горизонтали должны быть перечислены атрибуты полей набора ОД, по вертикали перечислены все поля набора ОД.

Первая строка файла структуры набора ОД всегда должна содержать 4 атрибута. Атрибуты должны располагаться именно в таком порядке, как представлены ниже:

1. Наименование поля набора ОД
 - Должен быть установлен обязательный заголовок атрибута: «field name».
 - Значением атрибута является наименование поля набора ОД.
2. Английское описание поля набора ОД
 - Должен быть установлен обязательный заголовок атрибута: «english description».
 - Значением атрибута является подробное описание поля набора ОД на английском языке.
3. Русское описание поля набора ОД
 - Должен быть установлен следующий обязательный заголовок атрибута: «russian description».
 - Значением атрибута является подробное описание поля набора ОД на русском языке.
4. Формат поля набора ОД
 - Должен быть установлен следующий обязательный заголовок атрибута: «format».
 - Значением атрибута является формат поля набора ОД на английском языке

С второй строки необходимо описать все поля набора ОД. Пустая строка между атрибутами и описанием не нужна.

Количество строк файла структуры должно быть на одну больше, чем количество полей в файле набора данных. Так, если в файле набора 4 поля, то файл структуры должен содержать 5 строк (первая строка атрибутов + 4 строки описания полей из набора).

3.1.2.5 Чек-лист проверки набора в формате CSV

- 1) Первая строка в файле отведена под заголовок

2) В файле отсутствуют пустые строки (которые нередко добавляют с целью улучшения его визуального восприятия)

3) Если поле не содержит данных, то оно обозначено просто разделительными запятыми, без пробелов между ними.

4) При записи дробных чисел в формате CSV, разделитель десятичной части является точкой, либо (если необходима именно запятая) экранируется.

5) Любые данные внутри ячейки окружены двойными кавычками. Если же кавычки применяются не ко всему полю

6) При наличии в тексте ячейки двойных кавычек, запятых и символов возврата каретки, содержимое всей ячейки заключено в двойные кавычки (если этого не сделать, то данные невозможно будет корректно обработать).

7) Когда двойные кавычки присутствуют внутри содержимого ячейки, они экранируются при помощи еще одной пары двойных кавычек. К примеру: «начало»«выделенной фразы»«конец».

8) Не используется символ «/» для экранирования кавычек, поскольку такое решение некорректно отображается в большинстве программ по созданию и просмотру CSV файлов.

3.1.3 JSON формат

JSON (JavaScript Object Notation, нотация объектов JavaScript) — это текстовый формат обмена данными. Он представлен наборами пар "ключ-значение", причем ключ — это всегда строка, а значение может задаваться одним из следующих типов:

- число;
- строка;
- логическое значение;
- массив;
- объект;
- нулевое значение null.

JSON основан на двух структурах данных:

- Коллекция пар ключ/значение. В разных языках, эта концепция реализована как объект, запись, структура, словарь, хэш, именованный список или ассоциативный массив.
- Упорядоченный список значений. В большинстве языков это реализовано как массив, вектор, список или последовательность. Структура JSON data type разработана на основе синтаксиса объектов JavaScript.

Процесс сериализации превращает объекты в строки JSON с помощью `JSON.stringify()`, а десериализация возвращает их обратно в объекты через `JSON.parse()`. Это позволяет легко передавать данные через сеть.

Формат JSON имеет ряд преимуществ:

- помогает легко воспринимать информацию как человеку, так и компьютеру;
- занимает мало памяти, потому что состоит из текста, что уменьшает нагрузку на сеть, и передача данных происходит быстрее;
- поддерживает множество разных языков программирования, что делает JSON универсальным инструментом обмена информацией разных систем.

3.1.3.1 Признаки для выбора формата

Основные характеристики для использования формата JSON:

1. Простота чтения и записи. JSON представляет собой текст в удобном для человека формате. Это делает его легким для чтения и создания.
2. Независимость от языка. Может быть использован с любым языком программирования, так как он представляет собой текстовый формат обмена данными.
3. Структурированный формат. JSON поддерживает массивы и объекты, что позволяет создавать сложные структуры данных. Данные могут быть вложенными и организованными в иерархическую структуру.
4. Легкая поддержка. У многих популярных языков программирования в составе есть функции для работы с JSON, а в JavaScript эта нотация встроена прямо в язык.
5. Не занимает много места и является компактным в написании и быстро компилируется.
6. Универсальность в типах данных. JSON можно использовать для хранения массивов, упорядоченных списков и коллекций пар «ключ — значение». Формат подходит для представления сложных структур данных, включая деревья и иерархические структуры.
7. Широкая поддержка в браузерах. Все современные браузеры поддерживают работу с JSON.
8. Самостоятельное документирование. В JSON можно включить метаданные, которые описывают представленные данные, упрощая работу с ними.

3.1.3.2 Типовые ошибки при работе с форматом

Типовые ошибки, которые допускаются в файлах формата JSON:

1. Ошибки синтаксиса. Отсутствие даже одного необходимого символа может привести к неработоспособности всего кода.
2. Неструктурированные данные. Если строка содержит типы данных, которые не соответствуют ожидаемой структуре, это вызовет ошибку в базе. Например, если ожидается массив, а в строке указан тип «объект», или наоборот.
3. Отсутствие запятых между элементами и отсутствие запятых после последнего элемента в массивах и объектах.
4. Отсутствие разделения ключей и значений с помощью двоеточий.

5. Некорректное использование скобок: фигурные {} для объектов и квадратные [] для массивов.

6. Несоблюдение соответствия логических значений: в JSON – это true и false.

Бывает, что файлы JSON содержат специальные символы или возникают проблемы с кодировкой:

```
try:
```

```
    with open('your_filename.json', 'r', encoding='utf-8') as json_file:
```

```
        parsed_data = json.load(json_file)
```

```
except UnicodeDecodeError as err:
```

```
    print(f"Произошла ошибка кодирования: {err}")
```

Правильная настройка параметров кодировки файла поможет избежать проблем с чтением данных.

3.1.3.3 Основные правила при составлении JSON вручную

Два основных типа структур данных в JSON — это объекты и массивы:

1. В JSON объекты являются беспорядочным набором сочетаний «ключ-значение». Между ключами (названиями) и значениями ставятся двоеточия, а пары разделяются запятой. Открывающая и закрывающая фигурные скобки обозначают начало и конец объекта:

```
{
    "название": "JSON для начинающих",
    "автор": "Иван Петров"
}
```

2. Массивы представляют собой упорядоченный список значений. Значения в массиве разделяются запятой. Массивы начинаются с открывающей квадратной скобки и заканчиваются закрывающей квадратной скобкой:

```
[
    "яблоко",
    "апельсин",
    "банан"
]
```

Эти две структуры могут комбинироваться для создания более сложных иерархий данных в JSON. Например, объекты могут содержать вложенные объекты или массивы, а массивы могут содержать объекты или другие массивы.

Одной из ключевых характеристик JSON является его способность представлять разнообразные типы данных, используя ограниченный набор примитивных базовых типов, таких

как строки, числа, логические значения, массивы и вложенные объекты. Эта гибкость делает его идеальным для передачи структурированных данных в приложениях, веб-сервисах и API.

Кроме того, для более формального описания структуры данных в формате JSON существует спецификация JSON Schema, которая позволяет определить правила и требования к данным в формате JSON, обеспечивая их валидацию:

```
{
  "structure": "object",
  "parameters": {
    "fullname": { "structure": "unicode" },
    "years": { "structure": "numeric" }
  },
  "compulsory": ["fullname", "years"]
}
```

В этом варианте JSON Schema указано, что данные в формате JSON должны быть объектом с необходимыми полями fullname и years, где fullname олицетворяет собой строку, а years — числовое значение.

3.1.3.4 Основные правила составления структуры

В основе структуры JSON лежат объекты и свойства. Объекты представляют собой неупорядоченную коллекцию пар "ключ-значение", где ключом может выступать строка, а значением может быть число, строка, логическое значение, массив или другой объект. Свойства объектов перечисляются через запятую и заключаются в фигурные скобки.

Массивы в JSON являются упорядоченными списками значений. Они могут содержать любые значения, включая числа, строки, логические значения, массивы или даже другие объекты. Массивы обозначаются с помощью квадратных скобок и элементы внутри массива перечисляются через запятую.

Значения в JSON могут быть различными типами данных, такими как числа, строки или логические значения. Числа записываются без кавычек. Строки должны быть заключены в двойные кавычки. Логические значения представлены, как true или false.

Форматирование JSON регламентируется определенными правилами. Каждый новый уровень вложенности — это отступ, обычно используются пробелы или табуляция.

Отличие JSON от XML и других форматов данных заключается в его простой и легко читаемой структуре. В отличие от XML, JSON не использует закрывающих тегов и имеет более компактный синтаксис. Благодаря этому, JSON стал предпочтительным форматом для передачи данных через сеть, особенно в веб-разработке.

3.1.3.5 Чек-лист проверки набора в формате JSON

1. Правильное тело файла
2. Различные параметры (обязательность, работа параметров)
3. Перестановка мест слагаемых (заголовки, тело файла)

Обратите внимание, что объект в формате JSON имеет несколько важных отличий от объектного литерала:

- Строки используют двойные кавычки. Никаких одинарных кавычек или обратных кавычек в JSON. Так `Ivan'` становится `"Ivan"`.
- Имена свойств объекта также заключаются в двойные кавычки. Это обязательно. Так `age:30` становится `"age":30`.
- Важное ограничение: не должно быть циклических ссылок.
- Имена свойств должны быть в двойных кавычках
- Имена ключей должны быть в одинарных кавычках
- Не допускается конструктор `"new"`, только значения
- Отсутствует запятая после последнего свойства
- Не должно быть запятой после последнего свойства

3.1.4 RDF формат

RDF (Resource Description Framework) — это стандарт, разработанный W3C для описания данных в сети Интернет. RDF предназначен для моделирования информации о веб-ресурсах (например, веб-страницах, элементах контента) и их взаимосвязях и позволяет структурировать и связывать данные таким образом, чтобы они могли быть поняты и обработаны как людьми, так и машинами.

В основе RDF лежит концепция триплетов (троек), где каждая запись представляет собой утверждение, состоящее из трёх частей:

- Субъект (Subject) — то, о чём идёт речь (например, конкретный веб-ресурс).
- Предикат (Predicate) — характеристика или свойство субъекта (например, «является автором»).
- Объект (Object) — значение свойства или другой ресурс, связанный с субъектом (например, имя автора).

Таким образом, каждое утверждение в RDF можно прочитать как "Субъект имеет свойство Предикат со значением Объект".

RDFa (Resource Description Framework in Attributes) — это технология, позволяющая внедрять семантические данные RDF в существующие документы HTML, XHTML и XML, добавляя дополнительные атрибуты к стандартным элементам разметки. RDFa используется

для предоставления структурированных данных и улучшения обработки контента, чтобы он был понятен как для пользователей, так и для поисковых систем и других автоматизированных агентов.

RDFa позволяет интегрировать триплеты RDF (субъект-предикат-объект) в HTML-документ. Эти триплеты представляют собой основные единицы семантических данных, описывающие отношения между ресурсами, такими как веб-страницы, текстовые блоки, изображения и другие элементы.

3.1.4.1 Признаки выбора для формата

RDFa целесообразно использовать в случаях:

- *Публикация контента в вебе*: формат подходит для публикации данных, которые уже содержатся на веб-страницах. Это удобный способ добавить дополнительные метаданные к существующему содержанию, не меняя структуры HTML-документа.
- *Повышение видимости данных*: использование RDFa помогает поисковым системам лучше понимать содержание страниц и делать их более доступными для поиска.
- *Интеграция с веб-страницами*: RDFa подходит для интеграции метаданных прямо в HTML-контент. Если вы создаете веб-страницы, на которых публикуются данные, и хотите, чтобы эти данные были структурированы и доступны для машинного чтения, RDFa является отличным выбором. Это позволяет использовать существующие HTML-документы для хранения и передачи метаданных, не создавая отдельные файлы.

3.1.4.2 Типовые ошибки при работе с форматом

1. Неправильное использование атрибутов: путаница между `property` и `typeof` может привести к некорректной интерпретации данных. Атрибут `property` указывает на свойство объекта, а `typeof` описывает тип объекта.
2. Отсутствие или неверное указание URI: поскольку RDFa работает с ресурсами, важно корректно указывать URI. Отсутствие уникальных идентификаторов может затруднить объединение и обработку данных.
3. Ошибки в онтологии: использование неподходящих или неверных терминов из онтологии (например, `Schema.org`) может привести к тому, что данные не будут корректно распознаны сторонними сервисами.

3.1.4.3 Основные правила при составлении RDFa вручную

Основная идея RDFa заключается в использовании атрибутов HTML для встраивания метаданных. Это включает добавление атрибутов `about`, `typeof`, `property`, `content`, `resource` и `datatype` в HTML-теги.

- **about** — указывает на ресурс, о котором идет речь (например, URL страницы или конкретного объекта).

- **property** — указывает свойство ресурса, которое описывается (например, название книги).
- **content** — позволяет определить значение свойства, если оно не совпадает с содержимым тега.
- **typeof** — указывает тип ресурса (например, "Книга").
- **resource** — используется для обозначения ресурсов, на которые ссылаются другие ресурсы.

Пример использования RDFa+HTML:

```
<!DOCTYPE html>
<html lang="en">
<head>
  <title>Example Page</title>
</head>
<body>
  <div about="http://example.org/book/book1" typeof="schema:Book">
    <span property="schema:name">Ruslan and Lyudmila</span> by
    <span property="schema:author">A.S. Pushkin</span>
    is a <span property="schema:genre">Fantasy</span> book.
  </div>
</body>
</html>
```

Проведем более подробный разбор примера:

1. `<div about="http://example.org/book/book1" typeof="schema:Book">`
 - `about="http://example.org/book/book1"`: Этот атрибут определяет, о каком конкретно ресурсе (объекте) идет речь. В данном случае, `http://example.org/book/book1` — это уникальный идентификатор книги, которая описывается в данном блоке.
 - `typeof="schema:Book"`: Атрибут `typeof` указывает на тип ресурса. Здесь указано, что данный ресурс является объектом типа "Book" (Книга), где `schema:Book` ссылается на понятие книги в онтологии Schema.org.
2. `<span property="schema:name">Ruslan and Lyudmila</span>`
 - `property="schema:author"`: Этот атрибут указывает на свойство "author" (автор) для текущего ресурса (книги). Значение свойства — "A.S. Pushkin".
 - `<span property="schema:author">A.S. Pushkin</span>`

- `property="schema:author"`: Этот атрибут указывает на свойство "author" (автор) для текущего ресурса (книги). Значение свойства — "A.S. Pushkin".

3. `<span property="schema:genre">Fantasy</span>`

- `property="schema:genre"`: Атрибут `property` указывает на жанр книги. В данном случае, жанр — "Fantasy".

Процесс составления файла RDFa:

1. Определение целей семантической разметки:

- Определите, какие данные нужно разметить (например, информацию о продукте, авторе статьи и т.д.).

- Выберите подходящий словарь (например, Schema.org, Dublin Core).

2. Создание HTML-шаблона:

- Начните с базовой HTML-структуры (например, `<html>`, `<head>`, `<body>`).

- Определите элементы, которые будут содержать семантические данные.

3. Добавление атрибутов RDFa:

- Используйте атрибуты `vocab`, `typeof`, `property`, `about`, `resource`, `content` для добавления семантики:

- `vocab` определяет базовый словарь (например, `http://schema.org/`).

- `typeof` указывает тип сущности.

- `property` определяет свойство (например, имя, адрес).

- `about` и `resource` указывают на идентификаторы сущности (например, URL).

- `content` используется для определения значения свойства.

4. Проверка соответствия структуры:

- Проверьте файл на корректность разметки с помощью валидаторов RDFa (например, W3C RDFa Validator).

- Убедитесь, что структура файла соответствует выбранному словарю (аналогично проверке XSD для XML).

3.1.4.4 Основные правила составления структуры

Структура открытых данных должна иметь машиночитаемое представление в формате XSD для открытых данных, представленных в формате RDF, раздел 3.1.1.4.

3.1.4.5 Чек-лист проверки набора в формате RDFa

При создании файла в формате RDFa необходимо проверить корректность по следующим параметрам:

1. Используются уникальные URI для всех ресурсов, что позволит корректно их идентифицировать в глобальной сети.

2. Выбраны стандарты словарей и онтологий при использовании предикатов (например, Dublin Core, FOAF, Schema.org), для обеспечения совместимости данных.

3. Минимизирована избыточность данных, использовано минимально необходимое количество предикатов при описании каждого ресурса, избегается дублирование информации.

4. Соблюдается логическая структура и последовательность, организовывая документ таким образом, чтобы он был читаемым и логически структурированным, для упрощения его последующего использования и обработки.

5. Документ проверен с помощью RDF-валидаторов (например, W3C RDF Validator), для проверки корректности структуры и данных.

4 Работа с наборами ОД

4.1 Функциональные возможности работы с наборами на Портале ОД

Портал предоставляет пользователям широкий спектр функциональных возможностей для работы с наборами ОД. Эти возможности обеспечивают доступ к данным, их анализ и взаимодействие с сообществом пользователей.

Общие функциональные возможности для всех пользователей:

1. Пользователи могут просматривать паспорт набора ОД, который содержит ключевую информацию о наборе, такую как название, описание, дата последнего обновления, источник данных, ответственный за публикацию и формат данных. Это позволяет пользователям оценить релевантность и актуальность данных перед их использованием.
2. Пользователи имеют возможность просматривать непосредственно записи набора ОД. Это позволяет ознакомиться с содержанием данных и провести предварительный анализ.
3. Пользователи могут просматривать комментарии других пользователей, оставленные под набором данных, и оставлять собственные комментарии (только авторизованные пользователи). Это способствует обмену мнениями, обсуждению данных и выявлению возможных улучшений или корректировок.
4. Пользователи могут скачивать файлы набора данных и паспорт набора ОД в поддерживаемых форматах (CSV, JSON, XML). Это позволяет использовать данные в локальной среде для дальнейшего анализа и обработки.

Дополнительные функциональные возможности для поставщиков набора данных:

Поставщики набора данных имеют возможность не только просматривать и оставлять комментарии, но и отвечать на комментарии других пользователей под набором ОД.

4.2 Уровень подготовки пользователей

Пользователи, работающие с наборами ОД, должны обладать следующими навыками:

- Основы работы с электронными таблицами и текстовыми редакторами.
- Понимание основ структурированных данных и форматов данных.
- Базовые навыки работы с программным обеспечением для обработки данных.

4.3 Перечень документов к предварительному ознакомлению

Для успешной работы с наборами ОД и обеспечения их соответствия стандартам, рекомендуется ознакомиться с перечнем регламентирующих документов, охватывающих различные форматы данных. Ниже представлен список ключевых стандартов для каждого из форматов:

1. CSV (Comma-Separated Values)
 - RFC 4180: Common Format and MIME Type for Comma-Separated Values (CSV) Files.

2. XML (eXtensible Markup Language)
 - ГОСТ ISO/IEC 12785-2-2015
 - ГОСТ Р 70317-2022 Пространственные данные. Метаданные. Часть 3. Реализация

XML-схемы для основных понятий

- W3C Recommendation: XML Schema Part 2: Datatypes
3. JSON (JavaScript Object Notation)
 - ECMA-404 The JSON Data Interchange Standard.
 - RFC 7159: The JavaScript Object Notation (JSON) Data Interchange Format.
 - ISO/IEC 21778:2017: Information technology – The JSON data interchange syntax.
 - ГОСТ ИСО 8601 — 2001.
 4. HTML+RDFa (Resource Description Framework in attributes)
 - W3C Recommendation: RDFa 1.1 Primer – Third Edition.
 - W3C Recommendation: HTML+RDFa 1.1 – Support for RDFa in HTML4 and XHTML1.
 5. HTML+Microdata
 - W3C Recommendation: HTML Microdata – A mechanism for annotating HTML documents

with machine-readable data.

- W3C Recommendation: HTML5 – A vocabulary and associated APIs for HTML and XHTML.

4.4 Рабочее место и дистрибутивы

Для обеспечения эффективной работы с наборами ОД на Портале ОД, клиентское рабочее место должно соответствовать ряду технических требований. Эти требования охватывают как программное, так и аппаратное обеспечение.

Требования к программному обеспечению клиентского рабочего места:

1. Операционная система

Для корректной работы необходимо использовать одну из следующих операционных систем:

- Windows 10 и выше;
- Linux, включая операционные системы на базе ядра Linux (типа Astra, РЕД ОС, ALT),

включенные в реестр российского программного обеспечения.

2. Браузер

Для доступа к Порталу и его функциональным возможностям требуется один из следующих браузеров:

- Яндекс.Браузер, версии 21 и выше;
- Mozilla Firefox, версии 87 и выше;
- Google Chrome, версии 89 и выше.

3. *Офисные пакеты приложений*

Для обработки и анализа данных, а также для работы с документами, необходимо наличие одного из следующих офисных приложений:

- Open Office, включая отечественные пакеты типа «МойОфис», «Р7 – Офис»;
- Microsoft Office.

Требования к аппаратному обеспечению клиентского рабочего места:

4. *Мышь и клавиатура*

Для удобного взаимодействия с системой рекомендуется использовать стандартную мышь и клавиатуру. Это позволит эффективно выполнять операции по вводу и навигации.

5. *Интернет-соединение*

Для обеспечения стабильного доступа к Порталу и быстрой загрузки данных требуется надежное интернет-соединение.

5 Вспомогательные материалы

5.1 Эталонные примеры наборов ОД и структуры

Пример информации, необходимой для размещения в виде набора ОД (Таблица 10):

Таблица 10 – Эталонный пример набора

Эталонный пример структуры набора ОД в формате JSON:

Название	Округ	Район	Адрес	Телефон
ГБУК г. Москвы «Выставочный зал «Солянка ВПА»	Центральный административный округ	Басманный район	улица Солянка, дом 1/2, строение 2	(495) 621-55-72; (495) 621-59-61
ГБУК г. Москвы «Выставочный зал «Творчество»	Центральный административный округ	Таганский район	Таганская улица, дом 31/22	(495) 678-55-78
ГБУК г. Москвы «Московский выставочный зал «Галерея А3»	Центральный административный округ	район Арбат	Староконюшенный переулок, дом 39	(495) 697-14-56

Структура содержит следующие поля:

1. Основные параметры JSON схемы:
 - title: Название набора данных ("Showrooms").
 - description: Описание набора данных ("Showrooms").
 - type: Тип данных на верхнем уровне (в данном случае массив "array").
 - minItems: Минимальное количество элементов в массиве (здесь минимум 1 элемент).
 - Элементы массива:
 - items: Описание элементов массива.
 - description: Описание элемента массива ("Administrative district.").
 - type: Тип элемента массива (здесь объект "object").
 - Свойства объекта:
 - properties: Описание свойств объекта.
 - name: Название административного округа.
 - description: Описание свойства ("Administrative district name.").
 - type: Тип данных свойства (строка "string").
 - showroom: Список выставочных залов.
 - description: Описание свойства ("List of showrooms.").

- type: Тип данных свойства (массив "array").
- minItems: Минимальное количество элементов в массиве (здесь минимум 1 элемент).
- items: Описание элементов массива выставочных залов.
- description: Описание элемента массива ("Showroom.").
- type: Тип элемента массива (объект "object").
- Свойства объекта "showroom":
- properties: Описание свойств объекта "showroom".
- name: Название выставочных залов.
- description: Описание свойства ("Name of showroom.").
- type: Тип данных свойства (строка "string").
- area: Муниципальная территория выставочных залов.
- description: Описание свойства ("Municipal area of showroom.").
- type: Тип данных свойства (строка "string").
- address: Адрес выставочных залов (улица, здание, структура, если есть).
- description: Описание свойства ("Street, building, structure (if any), where showroom is located.").
- type: Тип данных свойства (строка "string").
- telephone: Контактные номера телефонов шоурума в формате (xxx) ууу-уу-уу, разделенные ";", где "xxx" — телефонный код города.
- description: Описание свойства ("Contact phone numbers of showroom in format (xxx) ууу-уу-уу, separated by ";", where "xxx" is phone area code.").
- type: Тип данных свойства (строка "string").
-

structure.json

```
{ "$schema": "http://json-schema.org/draft-04/schema#",
  "title": "Showrooms",
  "description": "Showrooms.",
  "type": "array",
  "minItems": 1,
  "items": {
    "description": "Administrative district.",
    "type": "object",
    "properties": {
```

```

"name":{
  "description":"Administrative district name.",
  "type":"string"},
"showroom":{
  "description":"List of showrooms.",
  "type":"array",
  "minItems":1,
  "items":{
    "description":"Showroom.",
    "type":"object",
    "properties":{
      "name":{
        "description":"Name of showroom.",
        "type":"string" },
      "area":{
        "description":"Municipal area of showroom.",
        "type":"string"},
      "address":{
        "description":"Street, building, structure (if any), where
showroom is located.",
        "type":"string"},
      "telephone":{ "description":"Contact phone numbers of showroom in format
(xxx) yyy-yy-yy, separated by \";\", where \"xxx\" is phone area code.",
        "type":"string" } } },
"required":[
  "name",
  "area",
  "address",
  "telephone" ] } },
"required":[
  "name",
  "showroom"] } }

```

5.2 Эталонный пример паспорта набора ОД

Эталонный пример паспорта набора приведён в Таблице 11:

Таблица 11 – Эталонный пример паспорта набора ОД

№	Название поля паспорта	Значение поля паспорта (пример)
1	Идентификационный номер	7712345678-showrooms
2	Наименование набора ОД	Список выставочных залов
3	Описание набора ОД	Список выставочных залов с указанием контактных данных по округам и районам.
4	Владелец набора ОД	Министерство культуры
5	Ответственное лицо	Иванов Иван Иванович
6	Телефон ответственного лица	74951234567
7	Адрес электронной почты ответственного лица	ivanov@example.com
8	Гиперссылка (URL) на открытые данные	data-2013-06-01T00-00-structure-2013-04-01T00-00.xml
9	Формат набора ОД	XML
10	Описание структуры набора ОД	structure-2013 -04-01T00-00.xsd
11	Дата первой публикации набора ОД	01.04.2013
12	Дата последнего внесения изменений	01.06.2013
13	Содержание последнего изменения	Изменение данных
14	Дата актуальности набора данных	01.07.2013
15	Ключевые слова, соответствующие содержанию набора данных	Выставка, зал, выставочный зал, культура
16	Гиперссылки (URL) на версии открытых данных	
17	Гиперссылки (URL) на версии структуры набора	
18	Версия методических рекомендаций	Методические рекомендации 4.0

6 Публикация наборов ОД на Портале ОД

6.1 Определение и границы роли «Поставщик»

Поставщик открытых данных (далее – поставщик) – государственный орган или орган местного самоуправления, осуществляющий публикацию (размещение) наборов ОД на Портале открытых данных Российской Федерации.

6.1.1 Регистрация/авторизация/верификация Поставщика

Регистрация/авторизация поставщиков на портале осуществляется при помощи ЕСИА. Для регистрации/авторизации на портале следует нажать на кнопку «Войти в личный кабинет» в шапке любой страницы Портала ОД.

После регистрации/авторизации, пользователь попадает в личный кабинет. Также в личный кабинет можно попасть с любой страницы Портала по нажатию на кнопку в правом верхнем углу страницы. Откроется страница вида, как изображено на рисунке ниже.

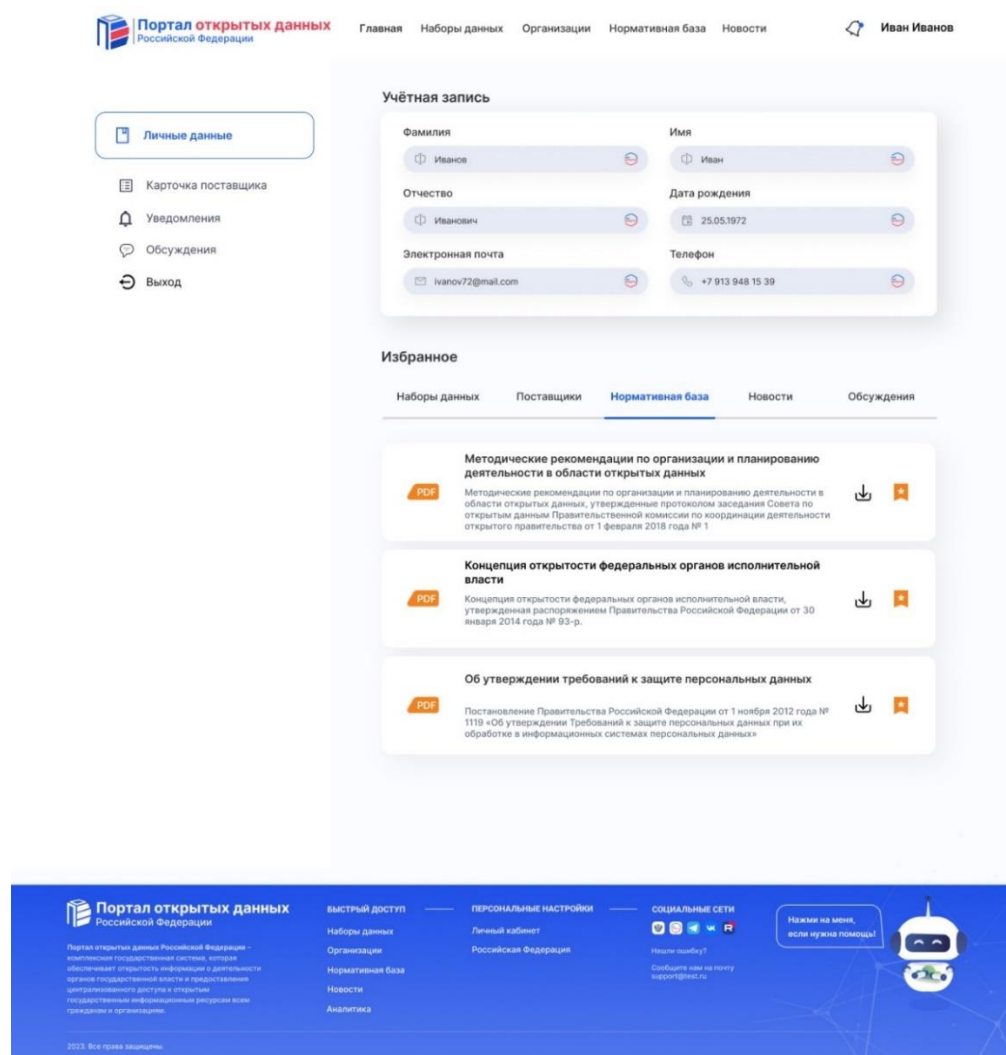


Рисунок 1 – Личный кабинет пользователя

Далее следует перейти на страницу «Карточка поставщика» путем выбора соответствующего раздела в меню, которое расположено слева.

Отобразится страница вида, как показано на рисунке ниже.

Портал открытых данных Российской Федерации

Главная Наборы данных Организации Нормативная база Новости Аналитика Иван Иванов

Личные данные

Карточка поставщика

Уведомления

Обсуждения

Выход

Данные организации

Наборы данных

Документы

Карточка поставщика

Организация

Общество с ограниченной ответственностью «Новые горизонты»

ИНН

7726344050

ОГРН

1157746571080

Территориальная принадлежность

Москва

КПП

1157746571080

Юридический адрес

г. Москва, шоссе Варшавское, д. 1А ком. 46

Контактная почта

ivanov72@mail.com

Контактный телефон

+7 913 948 15 39

Тип организации

☐ ФОИВ ☐ РОИВ ☐ ОМСУ ☐ Прочее

Дополнительно

Логотип

Загрузить иконку

Представитель

Иванов Илья Степанович

Описание деятельности

Сохранить изменения

Рисунок 2 – Карточка поставщика. Данные организации.

Поля данной формы используются в процессе верификации поставщика.

Часть данных о поставщике получается из ЕПГУ в автоматическом режиме в процессе регистрации/авторизации и не подлежит редактированию. В их числе:

- Наименование
- ИНН
- ОГРН
- КПП
- Юридический адрес
- Территориальная принадлежность

Кроме сведений, которые получены из ЕПГУ, пользователь должен указать в карточке поставщика следующие сведения:

- Контактная почта
- Контактный телефон
- Тип организации

- ФИО представителя
- Описание деятельности
- Логотип (загрузка графического изображения)

Все вышеуказанные атрибуты являются обязательными для заполнения.

Процесс верификации поставщика запускается после заполнения раздела «Данные организации» карточки поставщика и нажатия на кнопку «сохранить изменения».

Верификация поставщика осуществляется администратором системы на основании полей формы, которые были перечислены выше.

Публикация наборов ОД возможна только верифицированным поставщиком.

До принятия положительного решения о верификации поставщика, функционал вкладки «Наборы данных» недоступен.

В случае отрицательного решения по верификации, на данной странице будет отображено сообщение с причиной отказа, как показано на примере ниже.

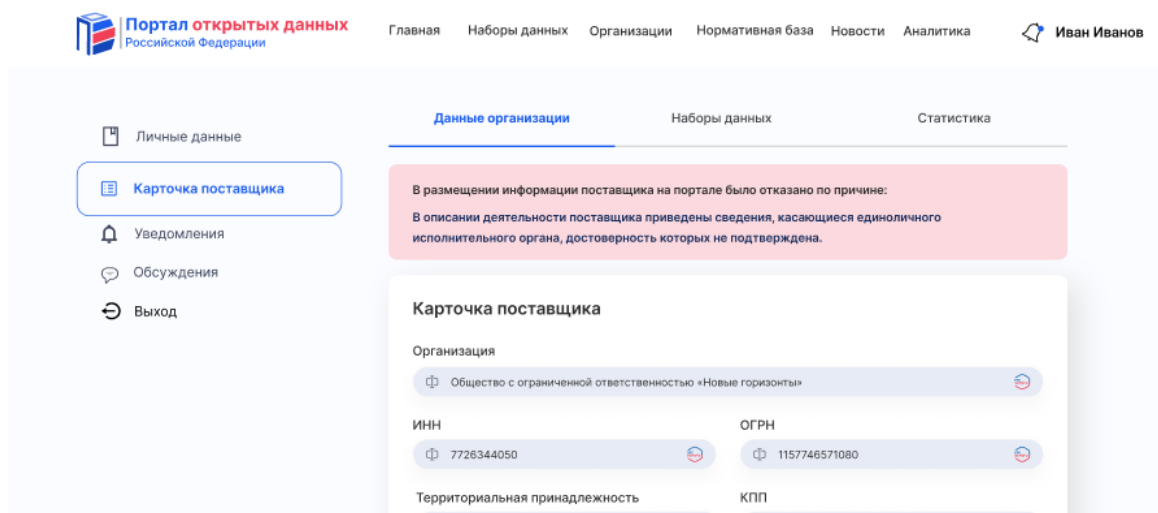


Рисунок 3 – Карточка поставщика. Данные организации.

В этом случае, пользователь имеет возможность исправить данные и повторно направить сведения на верификацию.

6.1.2 Карточка Поставщика открытых данных

Для просмотра наборов ОД, опубликованных поставщиком, необходимо перейти на вкладку «Наборы данных» карточки поставщика. Откроется страница вида, как показано на рисунке ниже.

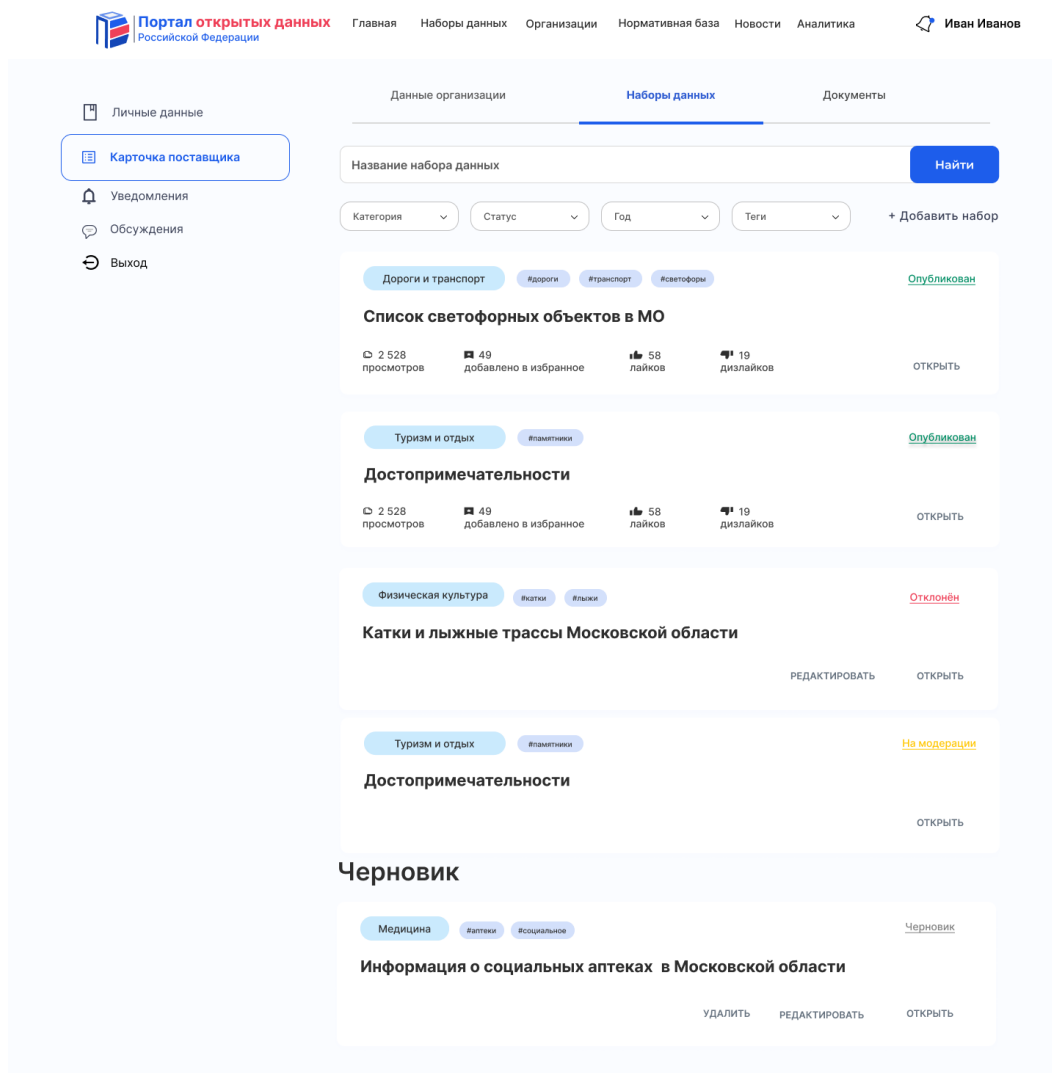


Рисунок 4 – Карточка поставщика. Наборы данных.

В данном разделе доступен перечень опубликованных поставщиком наборов ОД (помечены зеленым индикатором), а также наборов ОД, которые находятся на модерации (помечены желтым индикатором), в черновиках (помечены серым индикатором) или отклонены (помечены красным индикатором).

Карточка каждого опубликованного набора ОД имеет следующие атрибуты:

- Название;
- Категория;
- Теги;
- Количество просмотров;
- Количество добавлений в избранное;
- Количество лайков и дизлайков.

Наборы, которые не были приняты к публикации, доступны для редактирования. Для этого необходимо нажать на ссылку «Редактировать».

В случае, если была начата, но не была закончена процедура добавления нового набора ОД, он попадает в раздел черновики. Наборы ОД в черновиках доступны для удаления, редактирования и отправки на модерацию.

Содержимое каждого из наборов ОД можно просмотреть, для этого необходимо нажать на соответствующую ссылку.

Для поиска нужного набора стоит воспользоваться строкой поиска, а также фильтрами по категории, статусу, году или тегу. После ввода интересующего запроса следует нажать кнопку «Найти», после чего отобразится перечень всех наборов ОД поставщика, удовлетворяющих поисковому запросу.

Подробное описание процесса публикации наборов ОД описано в разделе 6.2 «Этапы публикации» Методических рекомендаций.

Для просмотра показателей деятельности и статистики поставщика на Портале ОД, необходимо перейти на вкладку «Статистика». Отобразится страница, как показано на рисунке ниже.

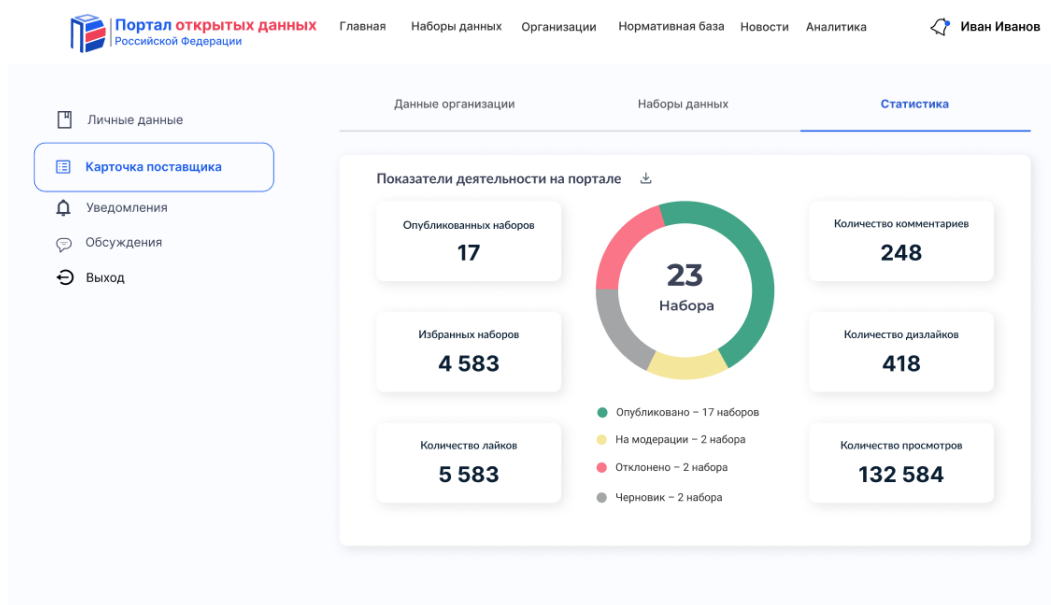


Рисунок 5 – Карточка поставщика. Статистика.

Для просмотра доступны следующие показатели:

- Общее количество наборов ОД
- Опубликованных наборов ОД
- Наборов ОД на модерации
- Отклонённых наборов ОД
- Черновики
- Количество добавлений наборов ОД в избранное
- Количество лайков наборов ОД

- Количество комментариев наборов ОД
- Количество дизлайков наборов ОД
- Количество просмотров наборов ОД

Показатели деятельности также доступны для загрузки. Для этого необходимо нажать на соответствующую иконку.

Выгрузка статистики деятельности поставщика доступна в формате PDF. Также функциональные возможности браузера позволяют отправить файл непосредственно на печать. (При условии наличия подключённых принтеров).

Сведения из карточки верифицированного поставщика являются общедоступными.

Они отображаются на странице поставщика, которая выглядит так, как показано на примере ниже.

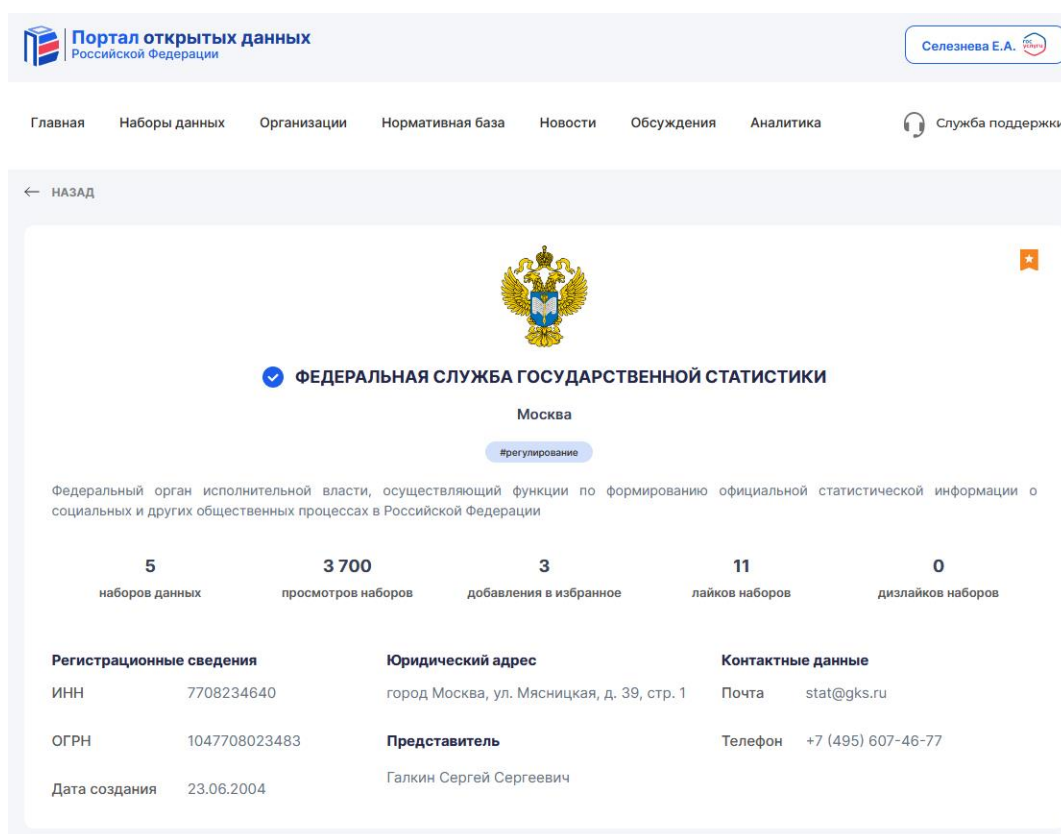


Рисунок 6 – Страница поставщика

На данной странице также отображается перечень всех опубликованных наборов ОД поставщика, как это показано на рисунке ниже.

Наборы поставщика		ПРОСМОТРЫ	В ИЗБРАННОМ	ОЦЕНКА	ОБНОВЛЕНИЕ	ФАЙЛ
ИНФОРМАЦИЯ О НАБОРЕ						
Здоровье	#здоровье					
Оценка состояния здоровья в настоящее время		361	1	2 3	21.12.2023	↓
Экономика	#численность занятых в россии					
Среднегодовая численность занятых в России (по укрупненным...		1 238	0	0 0	21.12.2023	↓
Экономика	#статистика сельского хозяйства					
ПОГОЛОВЬЕ КУР В ХОЗЯЙСТВАХ ГРАЖДАН В ГОРОДСКИХ...		607	1	0 0	21.12.2023	↓
Экономика	#тип населения					
Население, прошедшее диспансеризацию в последние два года		1 094	1	9 9	20.12.2023	↓
Экономика	#сельхоз					
ПОГОЛОВЬЕ СЕВЕРНЫХ ОЛЕНЕЙ В ХОЗЯЙСТВАХ НАСЕЛЕНИЯ		400	0	0 0	20.12.2023	↓

Рисунок 7 – Опубликованные наборы поставщика

По каждому набору ОД отображается:

- Категория
- Тэги
- Название
- Количество просмотров
- Количество добавлений в избранное
- Количество лайков
- Количество дизлайков
- Дата последнего обновления
- Ссылка для скачивания

Для выгрузки паспорта или набора ОД необходимо нажать на соответствующую иконку и выбрать один из форматов файла:

- CSV
- XML
- JSON

Выгрузка доступна в любом из форматов вне зависимости от того, в каком формате набор был загружен изначально.

6.2 Этапы публикации

Для добавления нового набора ОД нужно нажать кнопку «Добавить набор», после чего откроется форма выбора категории будущего набора ОД, как показано на рисунке ниже.

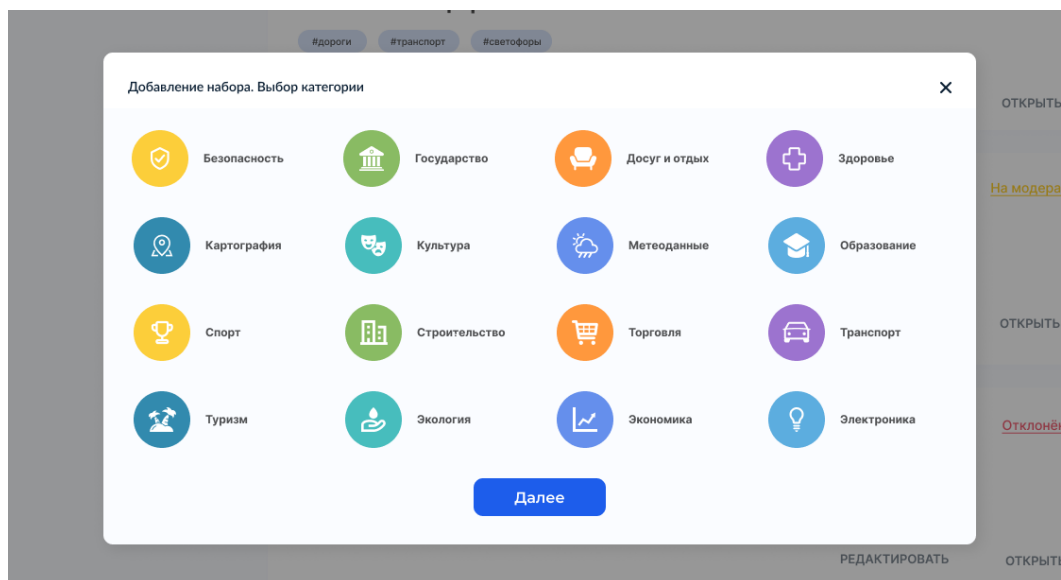


Рисунок 8 – Форма выбора категории нового набора.

Для выбора категории нужно нажать на ее иконку, после чего нажать кнопку «Далее», после чего откроется страница с формой заполнения данных, которая описана в разделе 6.2.1 Методических рекомендаций.

6.2.1 Паспорт набора ОД

На первом шаге заполняются поля паспорта набора ОД. Часть полей паспорта является предварительно заполненными и недоступными для редактирования:

- Владелец набора ОД;
- Ответственное лицо;
- Телефон;
- Электронный адрес;
- Дата первой публикации;
- Дата последнего внесения
- Версия методических рекомендаций

Портал открытых данных Российской Федерации

Главная Наборы данных Организации Нормативная база Новости Аналитика Иван Иванов

Добавление набора ОД

1 Паспорт * 2 Набор данных 3 Предпросмотр

Личные данные

Карточка поставщика

Уведомления

Обсуждения

Поддержка

Выход

Категория: Экономика

Идентификатор: Peoples

Наименование набора данных: Население, прошедшее диспансеризацию в последние два года

Описание набора данных: Информация о доли населения РФ, которые прошли диспансеризацию в последние два года по регионам

Владелец набора данных: Федеральная служба государственной

Ответственное лицо: Бутырев Владимир Викторович

Телефон: +74956074141

Электронный адрес: butyrevvv@gks.ru

Дата первой публикации: 20.12.2023

Дата последних изменений: 20.12.2023

Содержание последних изменений: Релиз

Ключевые слова (Теги): Тип населения, Пол, Регион

Дата актуальности: 20.12.2023

Версия методических рекомендаций: <https://data.gov.ru/metodicheskie-rekom>

Назад Далее

Рисунок 9 – Форма заполнения паспорта набора

Остальные поля формы необходимо заполнить вручную. Все поля являются обязательными для заполнения. При этом, необходимо соблюдать следующие условия заполнения:

- Идентификатор набора ОД должен представлять сокращенное англоязычное название набора ОД, указывается в одно слово (уникальное в пределах организации).
- Значение поля «Описание набора данных» должно составлять не менее 20 символов.
- Значение поля «Дата актуальности» должно содержать дату позднее, чем указана в поле «Дата последних изменений».

После заполнения всех полей, следует нажать на кнопку «Далее», после чего будет осуществлён переход на второй шаг, который описан в следующем разделе данного документа.

6.2.2 Набор ОД и его структура

После успешного прохождения первого шага, будет отображена страница, как показано на рисунке ниже.

Рисунок 10 – Второй шаг публикации данных

Далее необходимо загрузить файл набора ОД и файл структуры набора ОД в соответствующие поля формы. Загрузка может быть произведена двумя способами:

- Перемещением необходимых файлов в соответствующие области формы
- Нажатием на соответствующую область и выбор необходимого файла через проводник операционной системы.

После загрузки файлов любым из способов, указанных выше, будет произведен автоматический форматно-логический контроль. Подробное описание критериев форматно-логического контроля приведено в разделе 6.2.3 Методических рекомендаций.

В случае успешного прохождения форматно-логического контроля поле «Формат данных» будет заполнен автоматически. Также будут автоматически сформированы гиперссылки на набор ОД и структуру и заполнены соответствующие поля, как показано на рисунке ниже

Портал открытых данных
Российской Федерации

Главная Наборы данных Организации Нормативная база Новости Аналитика Иван Иванов

Личные данные

Карточка поставщика

Уведомления

Обсуждения

Выход

Добавление набора ОД

1 Паспорт * 2 Набор данных 3 Предпросмотр

Прикрепить набор и структуру данных

Набор открытых данных

Перетащите файл или выберите на компьютере

Peoples.csv

Структура набора данных

Перетащите файл или выберите на компьютере

Peoples_structure.csv

Гиперссылка на набор:
<https://portal-opendata.isands.ru/92ec620c>

Гиперссылка на структуру:
<https://portal-opendata.isands.ru/92ec620c>

Формат данных
CSV

Группа риска

☒ Отсутствуют персональные данные

☒ Отсутствуют конфиденциальные данные

☒ Отсутствуют геоданные

☒ Отсутствует информация ограниченного доступа

Связать набор

Адреса больниц новосибирской области, Адреса больниц и поликлиник Алтайского кра... ▾

Назад Далее

Рисунок 11 – Группы риска и связанные наборы

Также на данном шаге пользователю необходимо указать наличие или отсутствие групп риска публикации данного набора ОД. Для этого предусмотрены «чек-боксы» по каждой из групп риска публикации. Для подтверждения отсутствия определенной группы риска необходимо активировать соответствующий «чек-бокс».

Также, при необходимости пользователь имеет возможность связать размещаемый набор ОД с одним или несколькими ранее опубликованными наборами ОД данного поставщика. Данное поле не является обязательным к заполнению. В случае отсутствия ранее опубликованных данным поставщиком наборов ОД, выпадающий список будет пуст.

После заполнения полей данного шага, следует нажать на кнопку «Далее». Будет отображен предпросмотр паспорта набора ОД, как показано на примере ниже.

Предпросмотр

Структура 20.07.2021 / 31.12.2022 / 15.08.2023

Опубликовать

6.2.3 Форматно-логический контроль

После загрузки файлов набора ОД и структуры набора ОД автоматически запускается их проверка на форматно-логический контроль. Она выполняется по критериям, которые перечислены ниже.

- Один из следующих форматов файлов:
 - CSV
 - XML
 - JSON
- Форматы файлов набора ОД и структуры набора ОД должны совпадать
- Количество полей файла набора ОД и их описание в структуре набора ОД должны совпадать.
- Количество записей в файле набора ОД должно составлять не менее 2 строк и 2 столбцов;
- Файлы должны находиться в кодировке UTF-8

В случае несоответствия критериям форматно-логического контроля об этом будет выведено соответствующее информационное сообщение, пример которого представлен на рисунке ниже.

Портал открытых данных
Российской Федерации

Главная Наборы данных Организации Нормативная база Новости Аналитика Иван Иванов

Личные данные
Карточка поставщика
Уведомления
Обсуждения
Выход

Добавление набора ОД

1 Паспорт * 2 Набор данных 3 Предпросмотр

Загруженный набор открытых данных не прошёл проверку. Обнаружены несоответствия. Подробнее можно узнать в [Отчёте об ошибке](#). После корректировки файлов, вы можете снова их загрузить и нажать кнопку "Проверить".
С правилами заполнения набора и структуры можно ознакомиться [здесь](#).

Прикрепить набор и структуру данных

Набор открытых данных

Перетащите файл или выберите на компьютере
[Выбрать файл](#)

Гиперссылка на набор:

Структура набора данных

Перетащите файл или выберите на компьютере
[Выбрать файл](#)

Гиперссылка на структуру:

Формат данных

Группа риска

☒ Отсутствуют персональные данные ☒ Отсутствуют конфиденциальные данные

☒ Отсутствуют геоданные ☒ Отсутствует информация ограниченного доступа

Связать набор

Адреса больниц новосибирской области, Адреса больниц и поликлиник Алтайского кра...

[Назад](#) [Далее](#)

Рисунок 13 – Обнаружение несоответствий

Перечень несоответствий указан в отчете об ошибке. Для его загрузки необходимо нажать на ссылку «Отчёт об ошибке».

Структура отчета и пример его содержания описаны в разделе 3.3.1.2 Отчёт о валидации набора ОД.

После исправления несоответствий необходимо загрузить обновленные файлы набора ОД и структуры набора ОД и запустить проверку вновь.

6.2.4 Статусы набора ОД

Наборы данных на портале могут иметь следующие статусы:

- Черновик
- На модерации
- Опубликован
- Отклонён

Статус «Черновик» присваивается автоматически при выборе категории набора ОД поставщиком. В дальнейшем пользователь может вернуться к заполнению остальных данных.

Статус «На модерации» присваивается набору при отправке поставщиком набора ОД на модерацию.

Статус «Опубликован» присваивается администратором после проверки набора ОД на корректность и допустимость публикации.

Статус «Отклонен» присваивается после проверки набора ОД на корректность и допустимость. В случае получения такого статуса, поставщик набора ОД имеет возможность устранить выявленные несоответствия и повторно направить набор ОД на модерацию.

6.2.5 Модерация набора ОД

Модерация набора ОД осуществляется администратором системы. В процессе модерации администратор проверяет корректность заполнения паспорта набора ОД, файлов набора ОД и структуры набора ОД, а также оценивает указанные группы риска публикации набора ОД.

В случае принятия положительного решения, набор ОД публикуется на Портале ОД и становится общедоступным всем пользователям для просмотра, скачивания и комментирования.

Чтобы открыть опубликованный набор ОД в общедоступном режиме просмотра следует перейти в раздел «Карточка поставщика» личного кабинета и выбрать вкладку «Наборы данных». Будет отображена страница, как показано на рисунке ниже.

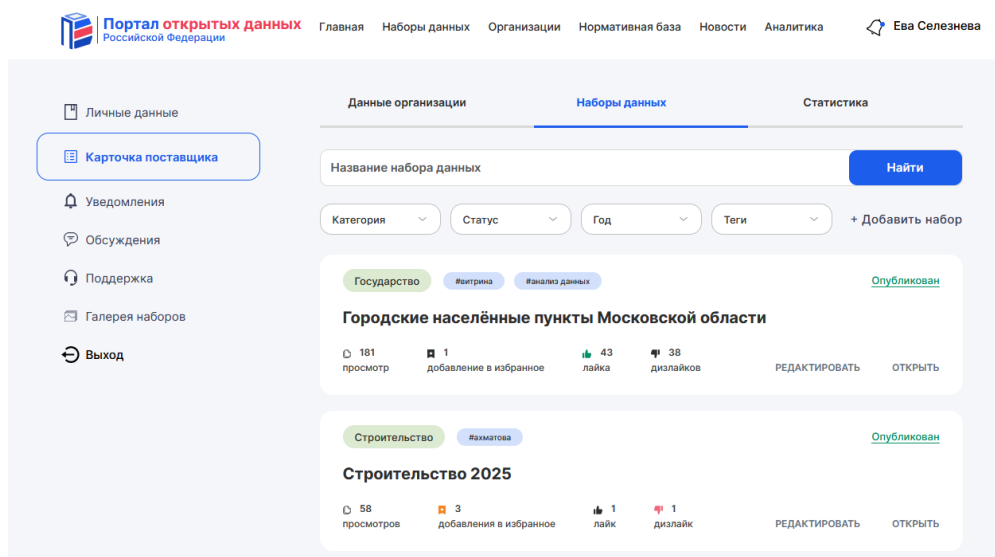


Рисунок 14 – Карточка поставщика. Наборы данных

Для просмотра опубликованного набора ОД, необходимо нажать на ссылку «Открыть» в соответствующем блоке.

Будет отображена страница следующего вида.

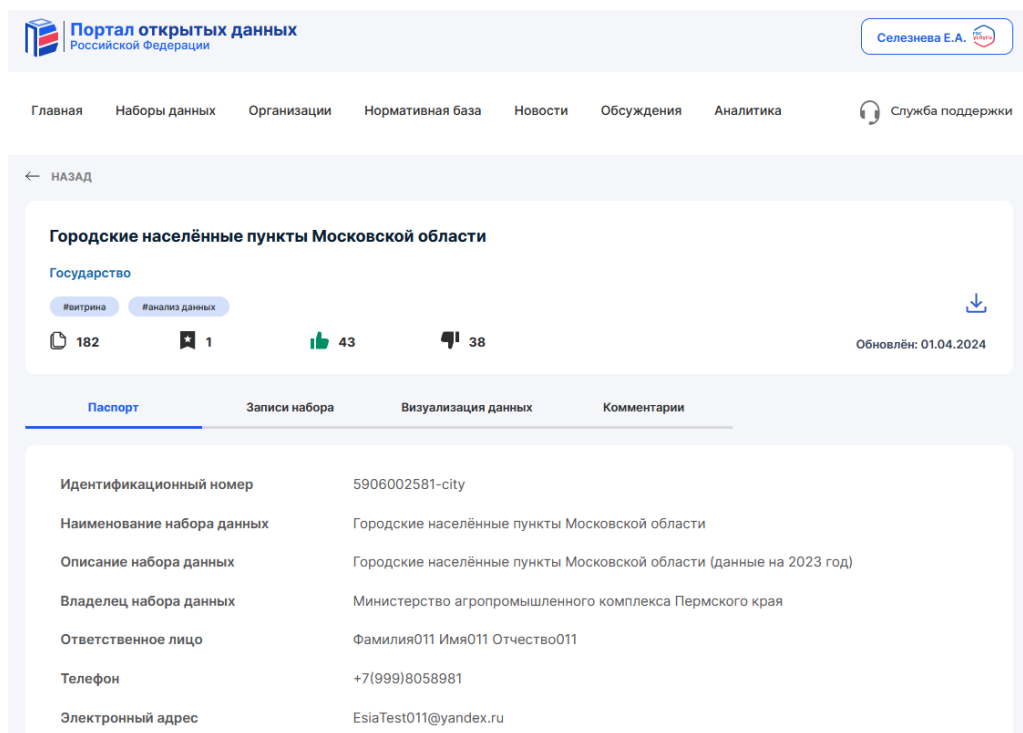


Рисунок 15 – Страница опубликованного набора данных

Набор ОД был опубликован на портале. Он доступен для просмотра и скачивания всем пользователям Портала ОД.

В случае если набор ОД не прошел модерацию, поставщик набора ОД имеет возможность устранить выявленные несоответствия и повторно направить набор ОД на модерацию. Причины, по которым набор ОД был отклонён, будут указаны в информационном блоке сверху набора. Для их просмотра необходимо открыть раздел «Карточка поставщика» личного кабинета, перейти на вкладку «Наборы данных», нажать «Открыть» в блоке соответствующего набора ОД.

При изменении статусов наборов ОД, пользователю портала в личный кабинет направляется соответствующее уведомление. Для просмотра уведомлений, необходимо в меню личного кабинета, выбрать раздел «Уведомления». Будет отображена страница, как показано на примере ниже.

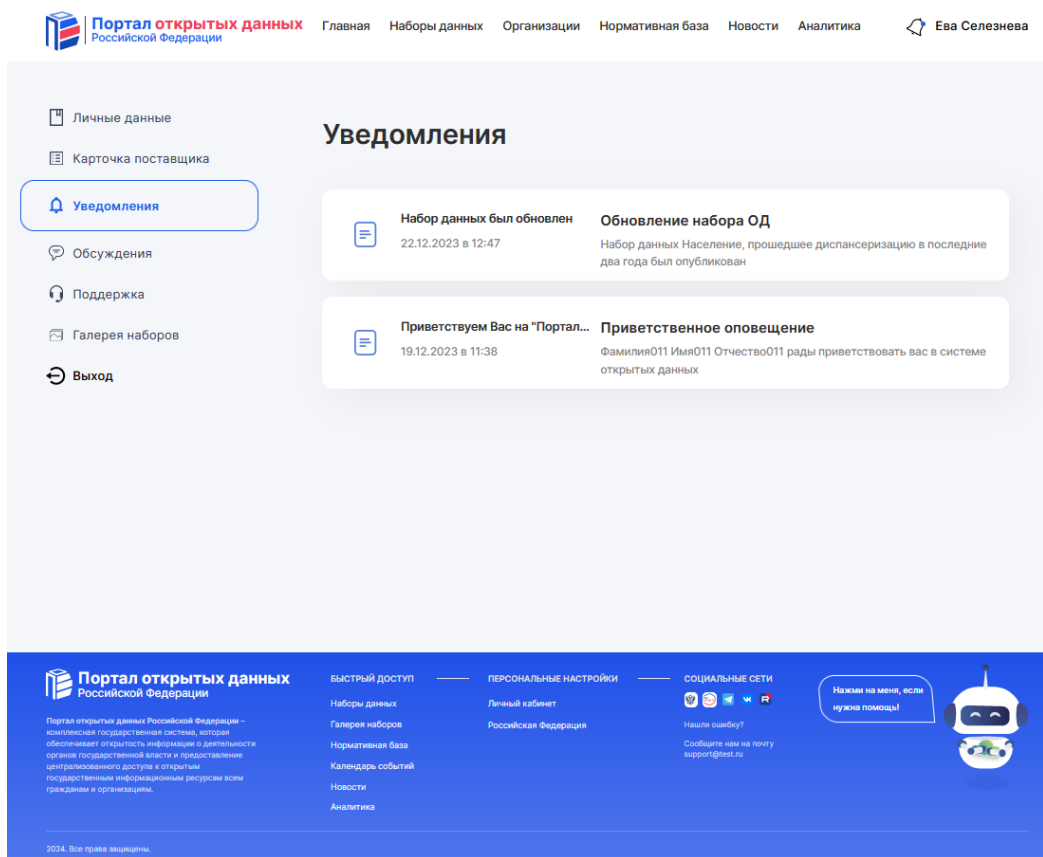


Рисунок 16 – Страница уведомлений

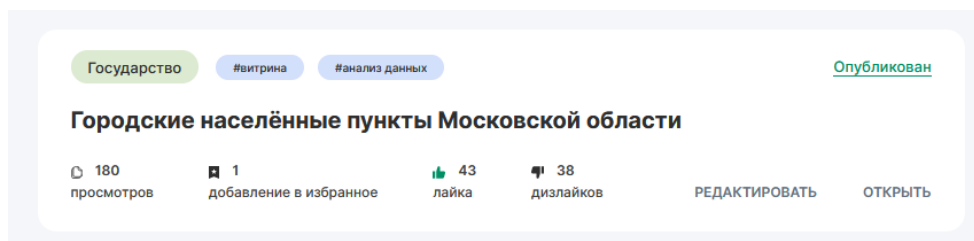
6.2.6 Статистика и показатели по набору ОД

Для каждого опубликованного на портале набора ОД собирается следующая статистики и показатели:

- Количество просмотров
- Количество добавлений в избранное
- Количество лайков
- Количество дизлайков

Добавление в избранное, постановку лайка или дизлайка набору ОД данных могут осуществлять только авторизованные пользователи Портала ОД. Просмотры осуществляются всеми пользователями Портала ОД.

Для просмотра статистики набора ОД, необходимо перейти в раздел «Карточка поставщика» и выбрать вкладку «Наборы данных». В блоке каждого опубликованного набора ОД приведены все собранные показатели на данный момент, как показано на рисунке ниже.



6.2.7 Изменение/обновление набора ОД

Для изменения/обновления набора ОД, необходимо перейти в раздел «Карточка поставщика» личного кабинета. Далее следует открыть вкладку «Наборы данных». Отобразится страница как показано на примере ниже.

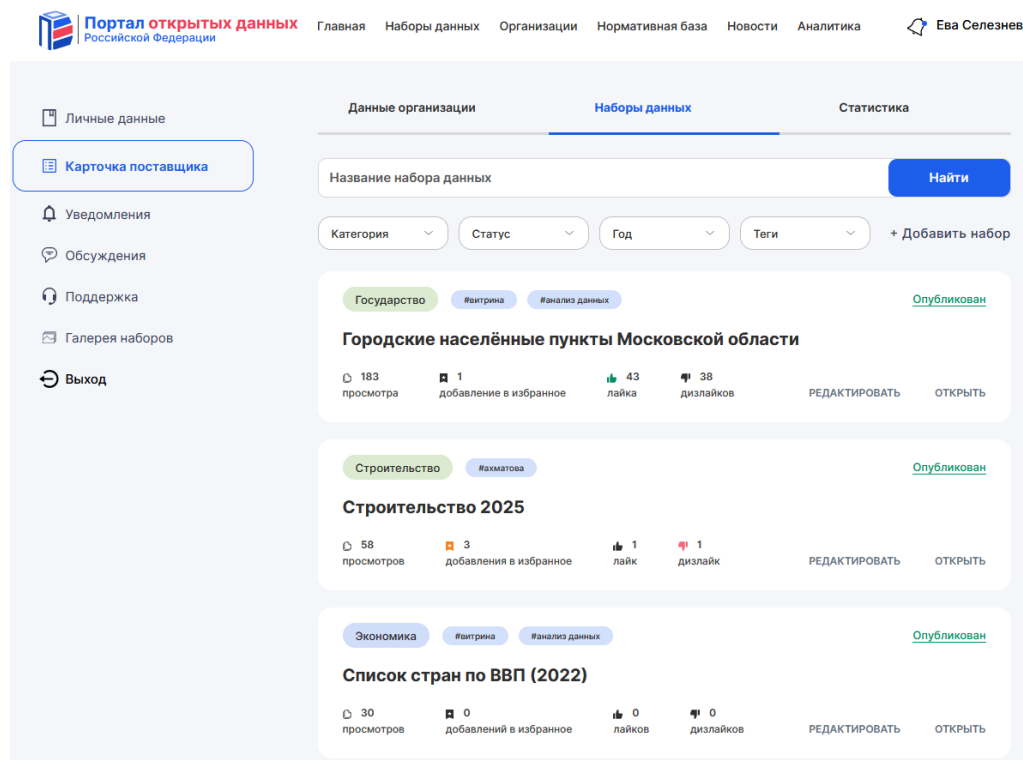


Рисунок 18 – Карточка поставщика. Наборы данных

Для изменения/обновления набора ОД необходимо нажать на ссылку «Редактировать» в блоке соответствующего набора ОД.

Процесс изменения/обновления набора ОД идентичен процессу публикации набора ОД и детально описан в пунктах 6.2.1-6.2.8 Методических рекомендаций.

При изменении каких-либо данных в наборе ОД, он повторно отправляется на модерацию к администратору.

6.2.8 Снятие с публикации

Снятие наборов ОД с публикации осуществляется администратором Портала ОД с учетом Приказа Минэкономразвития России № 247. В случае возникновения необходимости снять набор ОД с публикации, следует подать заявку в службу технической поддержки с указанием набора ОД, который необходимо снять с публикации, а также причины снятия. Для этого следует в меню личного кабинета выбрать раздел «Поддержка» и далее выбрать «Техническая поддержка». Отобразится форма, пример которой приведен на рисунке ниже.

Портал открытых данных
Российской Федерации

Селезнева Е.А.

Главная Наборы данных Организации Нормативная база Новости Обсуждения Аналитика Служба поддержки

Руководства и инструкции
Запрос на раскрытие данных
Отзывы и предложения
Стать поставщиком данных
Техническая поддержка

Обращение в службу технической поддержки

В целях оперативного рассмотрения вашего обращения просим максимально точно изложить суть вопроса и имеющиеся факты.

Фамилия
Селезнева

Имя
Ева

Отчество
Алексеевна

Электронная почта
EsiaTest011@yandex.ru

Телефон
+7(999)8058981

Опишите суть вашего обращения

Перетащите файл или выберите на компьютере
[Выбрать файл](#)

Отправить

Рисунок 19 – Обращение в службу технической поддержки

6.3 Варианты добавления наборов ОД

На Портале ОД предусмотрено 3 варианта добавления наборов:

- Ручной (Личный кабинет Поставщика)
- Автоматизированный (HTML-парсер)
- Автоматический (через REST Open API)

6.3.1 Ручной вариант добавления наборов ОД

6.3.1.1 Сценарий добавления набора ОД через Портал ОД

Добавление набора ОД поставщиком через портал состоит из следующих основных шагов:

1. Заполнение данных об организации и верификация (Если верификация еще не пройдена)
2. Выбор категории набора ОД
3. Заполнение данных паспорта набора ОД
4. Загрузка файлов набора ОД и структуры набора ОД
5. Указание отсутствия/наличия групп риска публикации набора ОД
6. Указание связанных наборов ОД (при необходимости)
7. Предпросмотр паспорта набора ОД предлагаемого к публикации набора ОД
8. Отправка набора ОД на модерацию.

Детальное описание процесса публикации набора ОД через Портал ОД описано в разделе 6.2 Методических рекомендаций.

6.3.1.2 Отчёт о валидации набора ОД

Отчет о валидации набора ОД включает в себя все несоответствия, которые были выявленных при форматно-логическом контроле файлов набора ОД и структуры набора ОД. Файл отчета может содержать следующие строки:

- Отсутствуют наборы ОД
- Отсутствуют файлы структуры набора ОД
- Проблема загрузки файла набора ОД по ссылке
- Отсутствуют файлы структуры набора ОД для актуального набора ОД
- Файлы набора ОД и структуры набора ОД имеют несоответствующие друг другу форматы,
- Количество полей, описанных в файле структуры набора ОД, отличается от количества полей в наборе ОД
- Некорректная структура XML или XSD. Line number: %s, Column number: %s. %s
- Еще дополнительно ошибок структуры XML или XSD в количестве: %
- Формат одного или нескольких файлов не поддерживается
- Данный файл не в кодировке UTF-8

6.3.2 Автоматизированный (HTML-парсер)

6.3.2.1 Правила ведения HTML страницы открытых данных в ГИС Поставщика

Требования к публикации паспорта набора ОД в машиночитаемом формате CSV:

1. Требования к расположению:
 - Машиночитаемый формат паспорта набора ОД представлен отдельным файлом в формате CSV.
 - Файл имеет название «meta.csv».
2. Требования к формату:
 - Соответствие файла стандарту RFC – «Common Format and MIME Type for Comma-Separated Values (CSV) Files» (<http://tools.ietf.org/html/rfc4180>).
3. Дополнительные требования к CSV:
 - Разделителем полей является знак «,» (запятая).
 - Ограничителем строк является знак «"» (универсальная двойная кавычка).
 - Разделителем целой и дробной части чисел является знак «.» (точка).
 - Кодировка файла – «UTF-8»

4. Требование к структуре данных:
 - По горизонтали перечислены атрибуты полей реестра наборов ОД, по вертикали перечислены поля реестра наборов ОД.
 - Наличие следующих обязательных атрибутов поля реестра наборов ОД:
5. Атрибут поля «Наименование поля реестра наборов ОД»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «property».
 - Значением атрибута является наименование поля реестра набора ОД.
6. Атрибут поля «Заголовок поля реестра наборов ОД»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «title».
 - Значением атрибута является заголовок поля реестра наборов ОД.
7. Атрибут поля «Значение поля реестра наборов ОД»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «value».
 - Значением атрибута является значение поля реестра наборов ОД.
8. Атрибут поля «Формат поля реестра наборов ОД»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «format».
 - Значением атрибута является формат значения поля реестра наборов ОД.
9. Наличие следующих обязательных полей реестра наборов ОД:
 - Должно быть установлено следующее обязательное наименование поля: «standardversion».
 - Должен быть установлен следующий обязательный заголовок поля: «Версия методических рекомендаций».
 - Значением поля является ссылка на версию методических рекомендаций, которой соответствует публикация этого реестра наборов ОД.
 - Атрибут «Формат поля реестра наборов ОД» неприменим для этого поля, поэтому значением формата поля является пустая строка.
10. Поле паспорта «Набор открытых данных» (данное поле повторяется столько раз, сколько опубликовано наборов ОД):
 - «<identifier>». Где <identifier> - это идентификационный номер паспорта набора ОД.
 - Должен быть установлен следующий обязательный заголовок поля: «<title>».Где <title> — это наименование набора ОД.
 - Значением поля является ссылка на машиночитаемый формат паспорта набора ОД.
 - Значением формата поля является машиночитаемый формат набора ОД.
11. Все наборы перечислены в одном файле, постраничное представление наборов ОД в нескольких CSV файлах не допускается. Пояснение: визуальное представление реестра

наборов ОД может быть разбито на несколько страниц и иметь сложную навигацию в случае большого количества наборов ОД, но при этом машинное представление должно быть представлено только одним файлом по указанному адресу и никак не зависит от страниц визуального представления.

Meta.csv (пример)

property,value
standardversion,<http://opendata.gosmonitor.ru/standard/3.0>
identifier,7712345678-showrooms
title,Список выставочных залов
description,Список выставочных залов с указанием контактных данных по округам и районам.

creator,Министерство культуры
created,20130401
modified,20130601
subject,"Выставка,зал,выставочный зал,культура"
format,xml
provenance,"Исправлены незначительные ошибки, добавлены новые выставочные залы"
valid,20130701
publishername,Иванов Иван Иванович
publisherphone,+74951234567
publishermailbox,ivanov@example.com
data-20130601T0000-structure-
20130401T0000,<http://example.com/opendata/7712345678-showrooms/data-20130601T0000-structure-20130401T0000.xml>
data-20130501T0000-structure-
20130401T0000,<http://example.com/opendata/7712345678-showrooms/data-20130501T0000-structure-20130401T0000.xml>
data-20130406T0000-structure-
20130401T0000,<http://example.com/opendata/7712345678-showrooms/data-20130406T0000-structure-20130401T0000.xml>
data-20130401T0000-structure-
20130401T0000,<http://example.com/opendata/7712345678-showrooms/data-20130401T0000-structure-20130401T0000.xml>
structure-20130401T0000,<http://example.com/opendata/7712345678->

Требования к публикации паспорта набора ОД в машиночитаемом формате XML:

1. Требования к расположению:
 - Машиночитаемый формат реестра наборов ОД представлен отдельным файлом в формате XML.
 - Файл имеет название «meta.xml».
2. Требования к формату:
 - Соответствие файла стандарту W3C – «Extensible Markup Language (XML) 1.1 (Second Edition)» (<http://www.w3.org/TR/xml11/>).
 - Кодировка файла указывается в объявлении XML если объявление XML используется, иначе используется кодировка «UTF-8».
3. Требование к структуре данных:
 - Соответствие файлу структуры паспорта набора ОД – «meta-schema.xsd» (Содержимое файла "meta-schema.xsd" приложено ниже).

Требования к публикации паспорта набора ОД в машиночитаемом формате JSON:

1. Требования к расположению:
 - Машиночитаемый формат паспорта набора ОД представлен отдельным файлом в формате JSON.
 - Файл имеет название «meta.json».
2. Требования к формату:
 - Соответствие файла стандарту JSON – (<http://json.org/json-ru.html>).
 - Кодировка файла – «UTF-8».
3. Требование к структуре данных:
 - Соответствие файлу структуры реестра наборов ОД – «meta-schema.json».

6.3.2.2 Сценарий добавления набора через HTML-парсер;

Парсер обеспечивает возможность сбора паспортов, самих наборов ОД с официальных сайтов, порталов, информационных систем государственной власти и местного самоуправления в целях помещения их в единое хранилище открытых данных. Настройка перечня источников осуществляется из закрытого контура Системы

Добавление набора ОД через html-парсер осуществляется администратором портала, который через специализированную форму Панели администратора запускает механизм получения данных. После этого, система выполняет переход на указанный URL адрес страницы.

При парсинге html-страницы производятся ряд проверок, которые затрагивают как паспорт набора ОД, так и файл набора ОД и структуры набора ОД.

В частности, паспорт набора ОД проверяется на точное соответствие критериям, которые указаны в пункте 6.2.1 Методических рекомендаций. Проверяются следующие поля паспорта набора ОД:

- Идентификационный номер
- Наименование набора ОД
- Описание набора ОД
- Владелец набора ОД
- Ответственное лицо
- Телефон ответственного лица
- Адрес электронной почты ответственного лица
- Гиперссылка (URL) на открытые данные
- Формат набора ОД
- Описание структуры набора ОД
- Дата первой публикации набора ОД
- Дата последнего внесения изменений
- Содержание последнего изменения
- Дата актуальности набора ОД
- Ключевые слова, соответствующие содержанию набора данных
- Гиперссылки (URL) на версии открытых данных
- Гиперссылки (URL) на версии структуры набора ОД
- Версия методических рекомендаций

Также выполняется проверка файла набора ОД и структуры набора ОД на соответствие следующим критериям:

- Один из следующих форматов файлов:
- CSV
- XML
- JSON
- Форматы файлов набора ОД и структуры набора ОД должны совпадать
- Количество полей файла набора ОД и их описание в структуре набора ОД должны совпадать.
- Количество записей в файле набора ОД должно составлять не менее 10
- Файлы должны находиться в кодировке UTF-8

6.3.2.3 Отчёт об итогах парсинга

По итогам парсинга набора с внешнего ресурса формируется отчет. В случае, если при парсинге набора ОД не было выявлено нарушений, паспорт, а также файлы набора ОД и структуры набора ОД импортируются на Портал ОД. Набору ОД присваивается статус «На модерации». После чего, администратор Портала ОД имеет возможность провести его проверку и принять решение о публикации. В случае, если при парсинге набора ОД были выявлены несоответствия, информация об этом заносится в отчёт. В отчете содержатся перечень всех выявленных несоответствий критериям паспорта набора ОД, а также требованиям, предъявляемым к набору ОД и его структуре.

6.3.3 Автоматизированный

6.3.3.1 Структура запроса данных в формате json;

Структура запроса данных в формате json с требуемыми типами данных по каждому из параметров предоставлена ниже.

```
{
  "datasetId": "string",
  "name": "string",
  "category": "3fa85f64-5717-4562-b3fc-2c963f66afa6",
  "description": "string",
  "organization": "3fa85f64-5717-4562-b3fc-2c963f66afa6",
  "contactName": "string",
  "contactEmail": "string",
  "contactPhone": "string",
  "format": "3fa85f64-5717-4562-b3fc-2c963f66afa6",
  "creationDate": "2024-07-12",
  "lastChangeDate": "2024-07-12",
  "actualDate": "2024-07-12",
  "lastChangeDescription": "string",
  "data": [
    {
      "fileId": "3fa85f64-5717-4562-b3fc-2c963f66afa6",
      "source": "string",
      "changeContent": "string",
      "createdDate": "2024-07-12",
      "structureLink": "string",
      "structureId": "3fa85f64-5717-4562-b3fc-2c963f66afa6"
```

```

    }
  ],
  "structure": [
    {
      "fileId": "3fa85f64-5717-4562-b3fc-2c963f66afa6",
      "source": "string",
      "createdDate": "2024-07-12"
    }
  ],
  "dataSource": "string",
  "risks": [
    "3fa85f64-5717-4562-b3fc-2c963f66afa6"
  ],
  "tags": [
    "3fa85f64-5717-4562-b3fc-2c963f66afa6"
  ],
  "joinedDataset": [
    "3fa85f64-5717-4562-b3fc-2c963f66afa6"
  ]
}

```

Расшифровка:

Параметр «datasetId» должен содержать в себе идентификатор набора ОД.

Параметр «name» должен содержать наименование набора ОД

Параметр «category» должен содержать идентификатор категории набора ОД

Параметр «description» должен содержать описание набора ОД

Параметр «organization» должен содержать идентификатор организации публикатора набора ОД

Параметр «contactName» должен содержать ответственное лицо

Параметр «contactEmail» должен содержать адрес электронной почты ответственного лица

Параметр «contactPhone» должен содержать телефон ответственного лица

Параметр «format» должен содержать идентификатор формата файлов набора ОД и структуры набора ОД

Параметр «creationDate» должен содержать дату первой публикации набора ОД

Параметр «lastChangeDate» должен содержать дату последнего внесения изменений

Параметр «actualDate» должен содержать дату актуальности набора ОД

Параметр «lastChangeDescription» должен содержать последнее изменение

Массив «data» должен содержать в себе все версии файла набора ОД и включать параметры по каждому из них:

- Параметр «fileId» должен содержать идентификатор файла набора ОД
- Параметр «source» должен содержать ссылку на источник файла набора ОД
- Параметр «changeContent» должен содержать изменение набора ОД
- Параметр «createdDate» должен содержать дату публикации
- Параметр «structureLink» должен содержать ссылку на структуру
- Параметр «structureId» должен содержать идентификатор структуры

Массив «structure» должен содержать в себе все версии файла набора ОД данных и включать параметры по каждому из них:

- Параметр «fileId» должен содержать идентификатор файла структуры набора ОД
- Параметр «source» должен содержать ссылку на источник файла структуры набора ОД
- Параметр «createdDate» должен содержать дату публикации

Параметр «dataSource» должен содержать ссылку на ресурс-источник

Массив «risks» должен содержать в себе перечень идентификаторов рисков публикации набора ОД

Массив «tags» должен содержать в себе перечень идентификаторов ключевых слов набора ОД

Массив «joinedDataset» должен содержать в себе перечень идентификаторов ранее опубликованных наборов ОД.

6.3.3.1 Процедура получения доступа к интеграции

Для получения доступа к интеграции следует направить запрос по любому официальному каналу связи в Министерство экономического развития Российской Федерации.

В запросе следует указать реквизиты и контактные данные государственного органа или органа местного самоуправления, от лица которого подаётся запрос. Также в запросе должны быть указаны параметры (Наименование, url-адрес) информационной системы, интеграция с которой запрашивается в запросе.

Памятка: Портал открытых данных Российской Федерации развернут и функционирует на основании компонентов единой цифровой платформы Российской Федерации «ГосТех».

Доступ по сети внешних ресурсов до информационных систем, размещенных в контуре ЕЦП «ГосТех» регулируется нормативными документами, а также внутренними регламентами оператора ЕЦП «ГосТех» и не входит в зону ответственности Министерства экономического развития Российской Федерации.

В этой связи, для получения доступа к интеграции может потребоваться ряд дополнительных технических и организационных действий от оператора внешней информационной системы, совокупность и порядок которых устанавливается оператором ЕЦП «ГосТех» и которые будут действовать на момент запроса интеграции.