

НАЦИОНАЛЬНАЯ СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ
ЕДИНЫЕ ТРЕБОВАНИЯ ПО УПРАВЛЕНИЮ ГОСУДАРСТВЕННЫМИ
ДАННЫМИ

Часть 4. Требования по информационному взаимодействию

Книга 4. Порядок публикации государственных данных

Версия 2.1

СОДЕРЖАНИЕ

1.	ОБЩИЕ СВЕДЕНИЯ.....	5
1.1.	Нормативная база.....	5
1.2.	Назначение документа.....	5
1.3.	Общие положения.....	5
2.	ПУБЛИКАЦИЯ ГОСУДАРСТВЕННЫХ ДАННЫХ.....	6
2.1.	Извлечение данных.....	6
2.2.	Преобразование данных	7
2.3.	Загрузка данных	8
2.4.	Базовые методы обновления данных в рамках регламентной загрузки данных в витрины данных.....	9

ИЗМЕНЕНИЯ

Версия	Дата	Автор	Изменения
2.1	05.02.2020	Воронов Ю.А.	Исходный документ

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

Сокращение	Описание
ГИР	Государственный информационный ресурс
ЕИП НСУД	Федеральная государственная информационная система «Единая информационная платформа национальной системы управления данными»

1. ОБЩИЕ СВЕДЕНИЯ

1.1. Нормативная база

Порядок публикации государственных данных разработан с учетом положений Концепции создания и функционирования национальной системы управления данными, утвержденной распоряжением Правительства Российской Федерации от 3 июня 2019 г. № 1189-р.

1.2. Назначение документа

Документ содержит комплекс требований к операторам государственных информационных ресурсов (ГИР) – поставщикам государственных данных по загрузке данных, относящихся к сформированному виду государственных данных, из ГИР в витрину данных с использованием средств ЕИП НСУД.

1.3. Общие положения

В настоящем документе приведено общее описание порядка публикации (первоначальной и регламентной загрузки) государственных данных в витрины данных, сформированные в соответствии с положениями Части 4, Книга 3 «Порядок формирования витрины данных» Единых требований по управлению государственными данными.

Алгоритмы публикации данных делятся на два класса:

1) Алгоритмы первоначальной загрузки данных в витрины данных, предназначенные для первичной загрузки данных в пустые (ранее сформированные) витрины данных.

2) Алгоритмы регламентной загрузки данных в витрины данных, предназначенные для обновления уже существующих (ранее загруженных) данных в витринах данных.

2. ПУБЛИКАЦИЯ ГОСУДАРСТВЕННЫХ ДАННЫХ

Публикация государственных данных в ранее сформированной витрине данных состоит из следующих стадий:

- 1) Извлечение данных – осуществляется извлечение данных из ГИР – источника данных;
- 2) Преобразование данных – осуществляется преобразование данных в соответствии с целевой структурой данных витрины данных;
- 3) Загрузка данных – осуществляется загрузка данных в витрину данных.

Каждая из перечисленных стадий состоит из ряда этапов.

При этом отдельные этапы могут быть полностью идентичными для обоих классов алгоритмов загрузки (например, единые правила преобразования данных и для первоначальной, и для регламентной загрузки данных). Приведенные процессы инициирует Оператор ГИР.

2.1. Извлечение данных

Извлечение данных из ГИР должно осуществляться за несколько последовательных этапов. Типовые этапы стадии извлечения данных перечислены в таблице 1.

Для извлечения данных с учетом особенностей ГИР состав типовых этапов может быть сокращен или расширен оператором ГИР.

Таблица 1 – Типовые этапы извлечения данных

№	Условное наименование этапа	Назначение этапа
1	Table input	Извлечение данных из базы данных с применением SQL-запросов
2	CSV file input	Извлечение данных из файла формата CSV
3	Fixed file input	Извлечение данных из текстового файла с фиксированной шириной столбца
4	Text file input	Извлечение данных из набора файлов формата CSV либо набора файлов с фиксированной шириной столбца
5	Get data from XML	Извлечение данных из файла формата XML
6	JSON input	Извлечение данных из файла формата JSON
7	Microsoft Excel input	Извлечение данных из книг формата Microsoft Excel

2.2. Преобразование данных

Преобразование данных, извлеченных из ГИР, должно осуществляться за несколько последовательных этапов. Типовые этапы стадии преобразования данных перечислены в таблице 2.

Для преобразования данных с учетом особенностей ГИР состав типовых этапов может быть сокращен или расширен оператором ГИР.

Таблица 2 – Типовые этапы преобразования данных

№	Условное наименование этапа	Назначение этапа
1	Add constant	Добавление в набор данных нового атрибута с постоянным значением
2	Calculator	Добавление в набор данных нового атрибута со значением, определяемым на основании иных атрибутов или условий исполнения этапа обработки
3	Concat fields	Добавление в набор данных нового атрибута, значение которого формируется путем объединения значений иных атрибутов и разделителя
4	Replace in string	Замена значений в текстовых строках
5	Select fields	Мэппинг атрибутов и конвертация их типа
6	Set field value	Замена значения атрибута значением из иного атрибута
7	Set field constant	Замена значения атрибута предопределенным значением
8	Split field to row	Разделение атрибута, содержащего множественные значения с разделителями, на несколько строк, содержащих каждое из значений этого атрибута
9	Split fields	Разделение атрибута, содержащего множественные значения с разделителями, на несколько новых атрибутов
10	String operations	Выполнение стандартных операций со строковыми значениями (приведение к одному регистру, удаление специальных символов, удаление пустых символов в начале и в конце строки и т.д.)
11	Strings cur	Получение части строки
12	Sort rows	Сортировка записей
13	Unique rows	Отбор только уникальных строк
14	Value mapper	Автозамена для текстовых значений
15	Clone row	Клонирование строки
16	If field value is null	Автозамена пустых значений на предопределенные значения
17	Null if	Автозамена определенных значений на пустые значения

№	Условное наименование этапа	Назначение этапа
18	Dummy (do nothing)	Пустой этап обработки, может использоваться для целей организации ветвлений, а также для отладки алгоритмов загрузки с целью промежуточного объединения результатов параллельной работы нескольких предыдущих этапов обработки
19	Filter rows	Фильтрация строк
20	Execute SQL script	Выполнение SQL скрипта
21	Formula	Добавление в набор данных нового атрибута со значением, определяемым с использованием встроенного языка формул
22	Modified JavaScript value	Добавление в набор данных нового атрибута со значением, определяемым с использованием скрипта JavaScript
23	Regex evaluation	Проверка значения с использованием регулярных выражений
24	Database Lookup	Поиск в базе данных строки по определенному ключевому атрибуту

2.3. Загрузка данных

Загрузка извлеченных из ГИР и преобразованных данных, должна осуществляться за несколько последовательных этапов. Типовые этапы стадии загрузки данных перечислены в таблице 3.

Для загрузки данных с учетом особенностей ГИР состав типовых этапов может быть сокращен или расширен оператором ГИР.

Таблица 3 – Типовые этапы загрузки данных

№	Условное наименование этапа	Назначение этапа
1	Delete	Удаление строк в базе данных по значениям определенных столбцов, используемых в качестве ключевых
2	Insert / Update	Обновление строк в базе данных по значениям определенных столбцов, используемых в качестве ключевых, либо добавление новых строк в случае отсутствия в базе данных таких ключевых значений
3	Table output	Добавление в базу данных новых записей (с использованием INSERT)
4	Update	Обновление данных в базе данных (с использованием UPDATE)

2.4. Базовые методы обновления данных в рамках регламентной загрузки данных в витрины данных

В результате регламентной загрузки должно быть осуществлено обновление данных в существующих витринах данных. В связи с наличием ранее загруженных данных применяются следующие базовые методы обновления данных в витринах данных в зависимости от специфики ГИР:

- 1) добавление только новых данных (без возможности обновления и удаления данных) – должно осуществляться с использованием этапа загрузки «Table output»;
- 2) обновление существующих данных и добавление новых данных (без возможности удаления данных) – должно осуществляться с использованием этапа загрузки «Insert / Update»;
- 3) обновление и удаление существующих данных, добавление новых данных должно осуществляться с использованием сочетания этапов загрузки «Insert / Update» и «Delete»;
- 4) удаление всех ранее существующих данных и создание новых записей – должно осуществляться с использованием этапа загрузки «Table output» с параметром «Truncate table».

После публикации государственных данных, относящихся к виду государственных данных, в витрине данных, автоматически средствами ЕИП НСУД заполняются служебные атрибуты описания вида государственных данных:

- признак публикации данных вида государственных данных оператором ГИР в витрине данных принимает значение «Да»;
- проставляется фактическая дата публикации государственных данных.