

Публикация открытых данных
органами государственной власти РФ:
детали процесса публикации наборов
открытых данных



Москва, 2013

Содержание семинара

- Введение
 - ✓ Основные понятия
 - ✓ Нормативно-правовое обеспечение
- Способы публикации наборов открытых данных
- Процесс публикации наборов открытых данных
- Возможные типы исходных данных
- Структура и строение набора открытых данных
- Зависимость формата структуры набора открытых данных от формата открытых данных
- Процесс формирования набора открытых данных
- Способы преобразования информации в открытые данные
- Процесс преобразования в открытые данные
- Мастер-класс

ОСНОВНЫЕ ПОНЯТИЯ

Открытые данные – информация, размещаемая ее обладателями в сети "Интернет" в формате, допускающем автоматизированную обработку без предварительных изменений человеком в целях повторного ее использования (ст.7 п.4 Федерального закона от 27.07.2006 №149-ФЗ).

Набор открытых данных (набор данных) – систематизированная совокупность открытых данных и метаданных о них.

Машиночитаемые данные – данные, представленные в форматах, пригодных для автоматического или автоматизированного использования.

Метаинформация (метаданные) – структурированные данные, представляющие собой характеристики описываемых сущностей для целей их идентификации, поиска, оценки, управления этими сущностями.

Открытая лицензия - документ, представляет собой свод прав и ограничений рассматриваемого объекта.

Нормативно-правовое обеспечение

Федеральные законы

1. Федеральный закон от 27.07.2006 N 149-ФЗ (ред. от 02.07.2013) "Об информации, информационных технологиях и о защите информации "
2. Федеральный закон от 09.02.2009 N 8-ФЗ (ред. от 07.06.2013) "Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления»

НПА Правительства

1. Постановление Правительства РФ от 24.11.2009 N 953 (ред. от 10.07.2013) "Об обеспечении доступа к информации о деятельности Правительства Российской Федерации и федеральных органов исполнительной власти"
2. Постановление 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»
3. Распоряжение Правительства Российской Федерации от 10 июля 2013 г. № 1187-р «О перечнях общедоступной информации, размещаемой в сети «Интернет» в форме открытых данных»

Ведомственные НПА

1. Приказ Министерства связи и массовых коммуникаций Российской Федерации от 27 июня 2013 г. № 149 «Об утверждении требований к техническим, программным и лингвистическим средствам, необходимым для размещения информации в сети «Интернет» в форме открытых данных, а также для обеспечения ее использования»

Методическое обеспечение

- Национальный план действий Российской Федерации в области открытых данных
- Концепция открытых данных
- План мероприятий («дорожная карта») «Открытые данные Российской Федерации»
- План действий по реализации принципов Хартии по открытым данным
- Методические рекомендации по публикации открытых данных государственными органами и органами местного самоуправления и технические требования к публикации открытых данных (Версия 2.3)
- Методические рекомендации по публикации открытых данных государственными органами и органами местного самоуправления и технические требования к публикации открытых данных (Версия 3.0)

Способы публикации

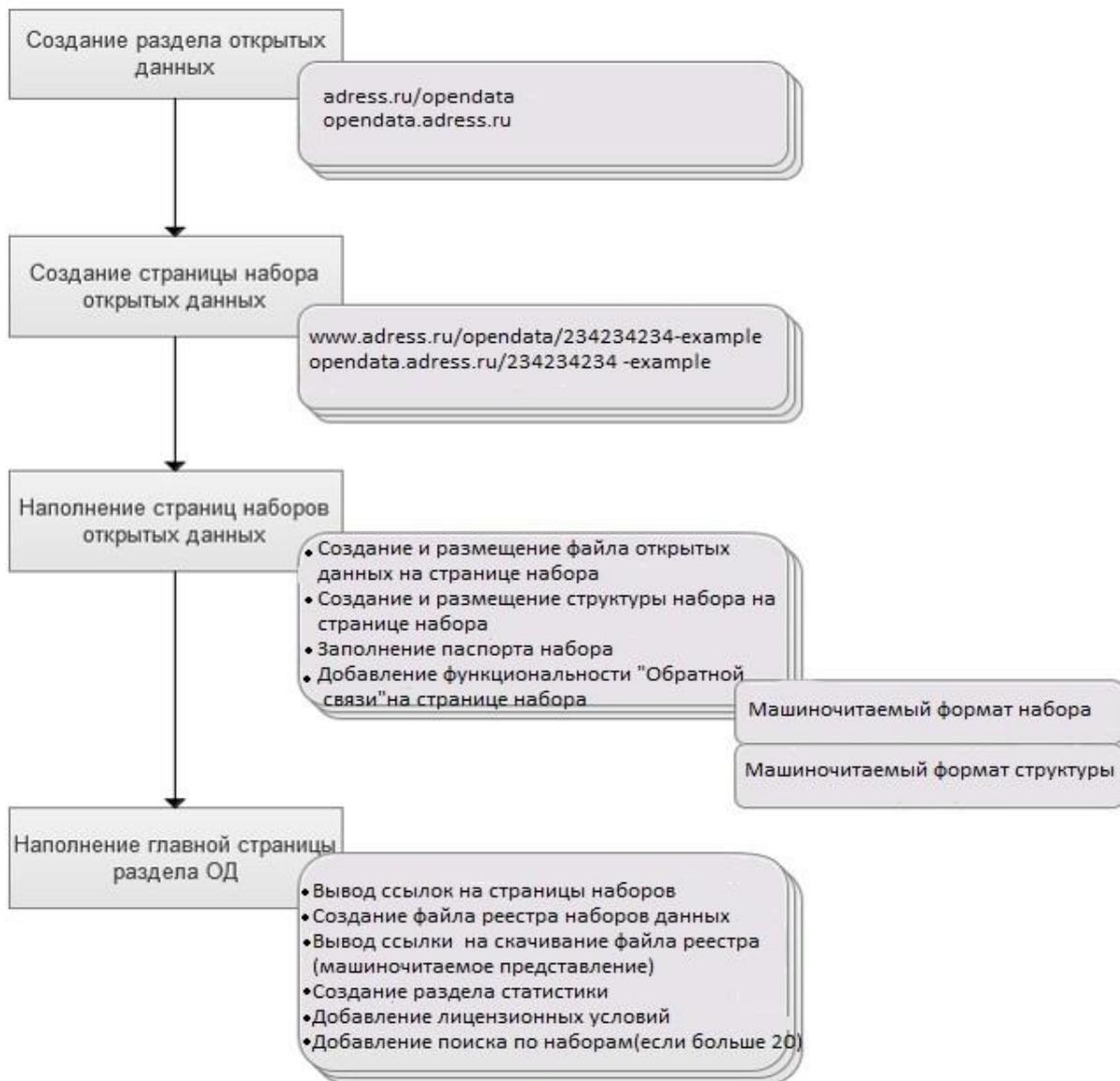
Обязательные способы публикации:

- ✓ через раздел открытых данных на официальном сайте государственного органа
- ✓ через портал открытых данных государственного органа
- ✓ через портал открытых данных РФ

Альтернативные способы публикации:

- ✓ через сайт третьей стороны
- ✓ через ftp-сервер
- ✓ через файлообменный протокол BitTorrent

Процесс публикации набора открытых данных



Возможные типы исходных данных

Данные

Линейные

Список:

Аистов Александр Алексеевич
Бобров Борис Борисович
Воробьева Валентина Владиславовна

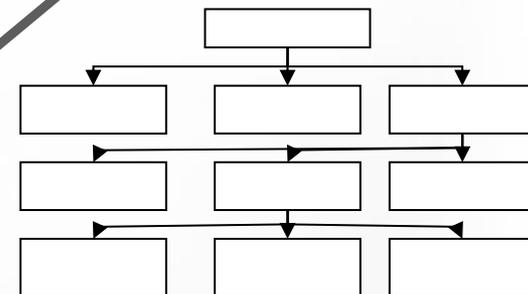
Таблица:

Тематические

(например, пространственные данные)

Объект	Долгота	Широта

Иерархические

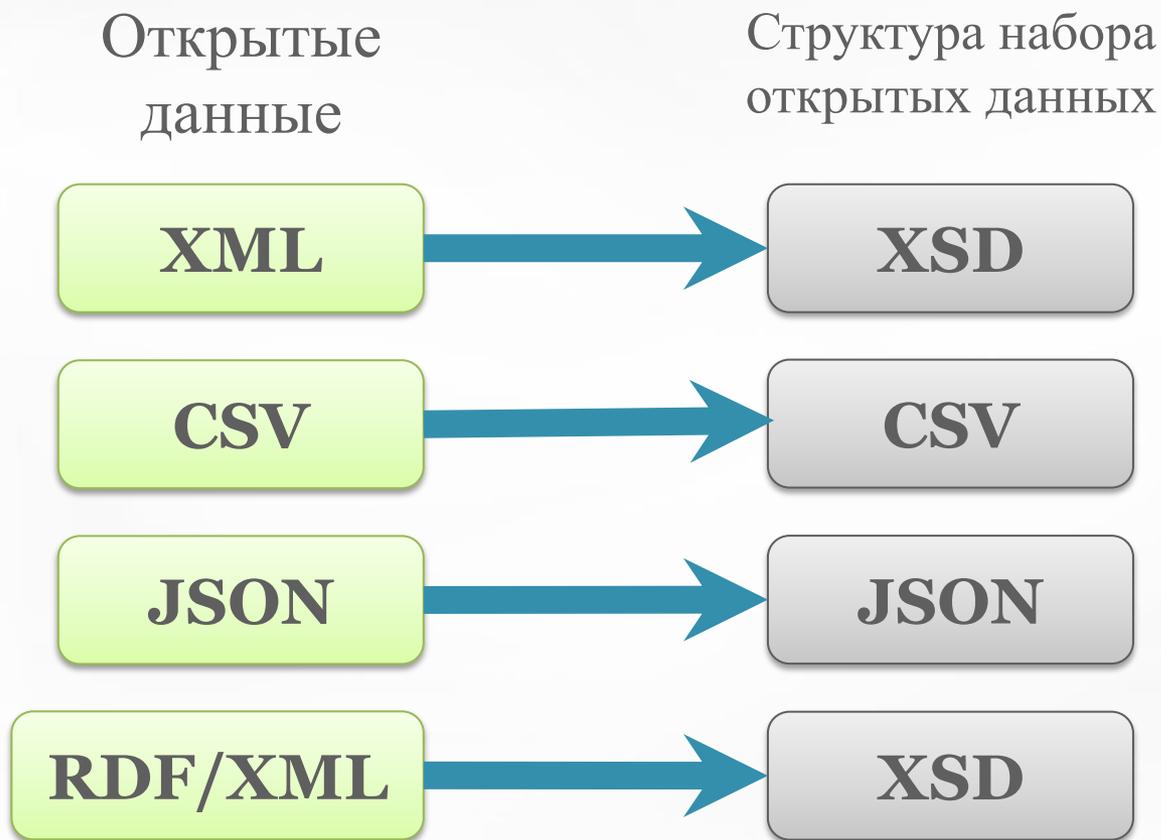


- Организационная структура
- Файловая система
- Группа задач в проекте;
- Классификация языковых терминов

Структура и строение набора открытых данных



Зависимость формата структуры набора открытых данных от формата открытых данных



Процесс формирования набора открытых данных



Способы преобразования исходных данных

Способы

```
graph TD; A[Способы] --> B[Ручной]; A --> C[Автоматизированный режим]; B --> D[Компетенции:]; C --> E[Инструменты:]; D --> D1[Навыки работы с XML, JSON, API, RDF]; D --> D2[Базовые навыки программирования]; E --> E1[Excel (табличные данные)]; E --> E2[Распознаватели (например: Fine Reader)]; E --> E3[Конвертеры];
```

Ручной

Компетенции:

- Навыки работы с XML, JSON, API, RDF
- Базовые навыки программирования

Автоматизированный режим

Инструменты:

- Excel (табличные данные)
- Распознаватели (например: Fine Reader)
- Конвертеры

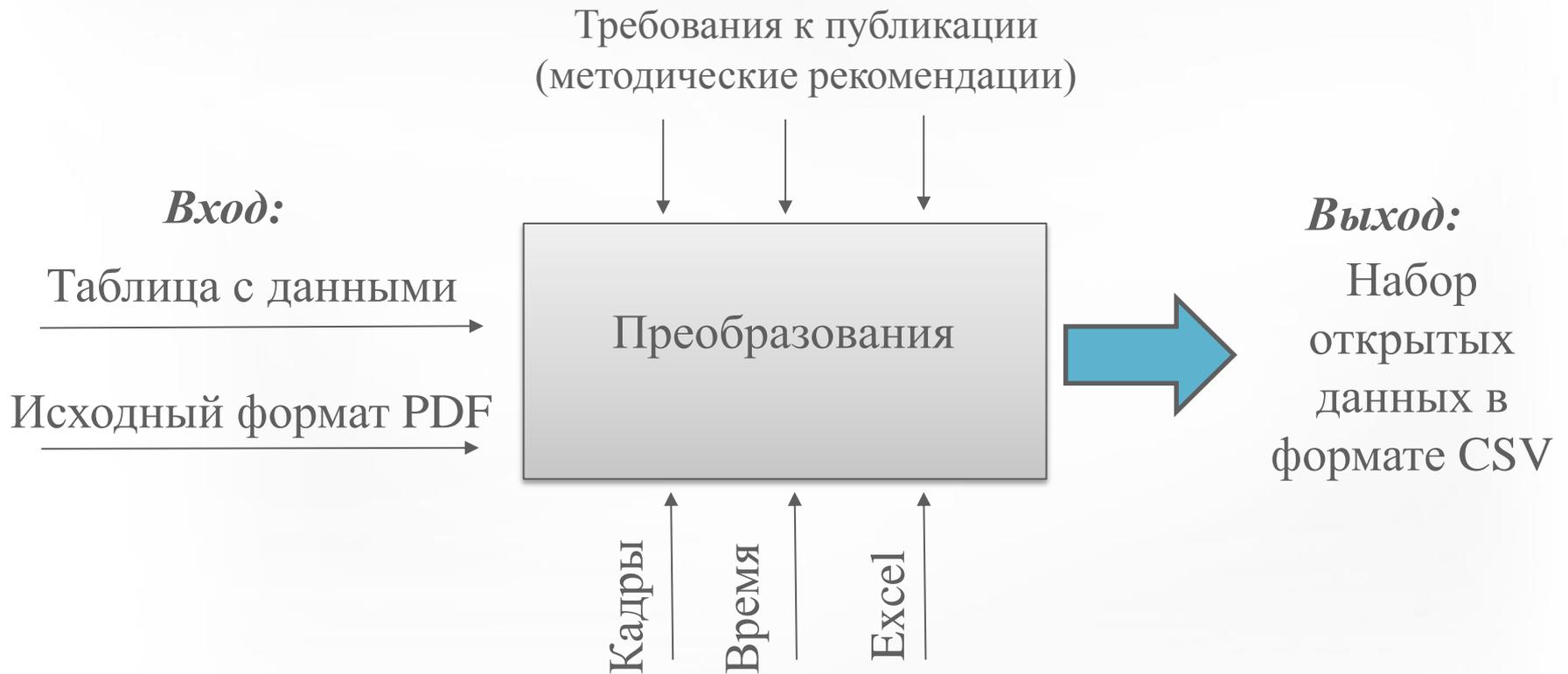
Процесс преобразования в открытые данные



Мастер-класс

Кейсы можно скачать по следующей
ссылке : <http://opendata.gosmonitor.ru/>

Кейс №1



Требования к публикации набора открытых данных в машиночитаемом формате CSV

1. Требования к форматам файла набора и файла структуры набора открытых данных:

- 1) Соответствие файла стандарту RFC – «Common Format and MIME Type for Comma-Separated Values (CSV) Files» (<http://tools.ietf.org/html/rfc4180>).
- 2) Кодировка файла – «UTF-8».

2. Требование к структуре файла набора открытых данных:

- 1) Файл набора открытых данных соответствует файлу структуры набора открытых данных.

Требования к публикации набора открытых данных в машиночитаемом формате CSV

Требование к структуре файла структуры набора открытых данных:

- По горизонтали перечислены атрибуты полей набора открытых данных, по вертикали перечислены все поля набора открытых данных.
- Наличие следующих обязательных атрибутов поля набора открытых данных:
 - ✓ Атрибут поля «Наименование поля набора открытых данных»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «field name»
 - Значением атрибута является наименование поля набора открытых данных
 - ✓ Атрибут поля «Английское описание поля набора открытых данных»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «english description»
 - Значением атрибута является подробное описание поля набора открытых данных на английском языке
 - ✓ Атрибут поля «Русское описание поля набора открытых данных»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «russian description»
 - Значением атрибута является подробное описание поля набора открытых данных на русском языке
 - ✓ Атрибут поля «Формат поля набора открытых данных»:
 - Должен быть установлен следующий обязательный заголовок атрибута: «format»
 - Значением атрибута является формат поля набора открытых данных на английском языке

Решение кейса

Действия:

1. Выбрать исходные данные для публикации
2. Перенести данные в инструмент обработки (например: Excel)
3. Сохранить в формате CSV

Таблица №1 Пример материала для публикации в виде открытых данных

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
ГБУК г. Москвы «ГВЗ «Ковчег»	Северный административный округ	Тимирязевский район	улица Немчинова, дом 12	(499) 977-00-44; (499) 977-00-88

Таблица в формате xls

Microsoft Excel Worksheet - Excel

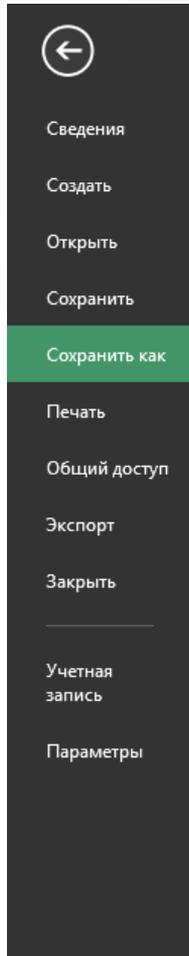
ФАЙЛ ГЛАВНАЯ ВСТАВКА РАЗМЕТКА СТРАНИЦЫ ФОРМУЛЫ ДАННЫЕ РЕЦЕНЗИРОВАНИЕ ВИД Foxit Reader PDF Foxit PDF

Вставить Шрифт Выравнивание Число Условное форматирование Форматировать как таблицу Стили Ячейки Вставить Удалить Формат Сортировка Найти и фильтр выделите Редактирование

A1 : X ✓ f Название

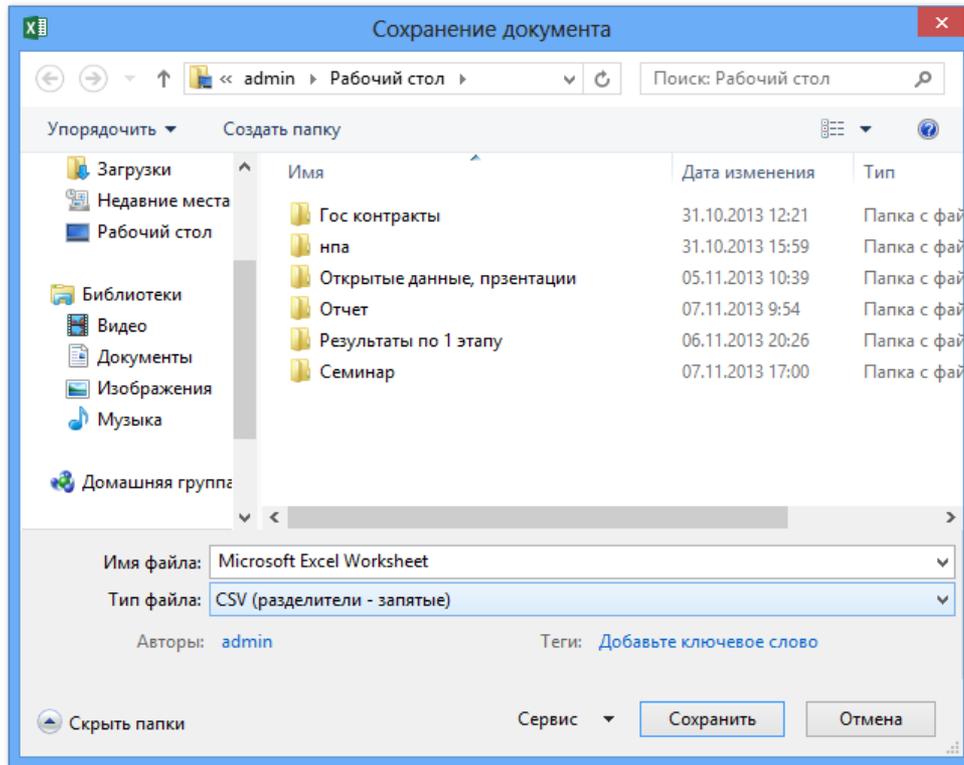
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Название	Округ	Район	Адрес	Телефон												
2	ГБУК г. Москвы «Выставочный зал «Солянка ВПА»	Центральный административный округ	Басманный район	улица Солянка, дом 1/2, строение 2	(495) 621-55-72; (495) 621-59-61												
3	ГБУК г. Москвы «Выставочный зал «Творчество»	Центральный административный округ	Таганский район	Таганская улица, дом 31/22	(495) 678-55-78												
	ГБУК г. Москвы «Московский	Центральный		Старокожуховский	(495)												

Сохранение данных в формате CSV



Microsoft Excel Worksheet - Excel

Сохранить как



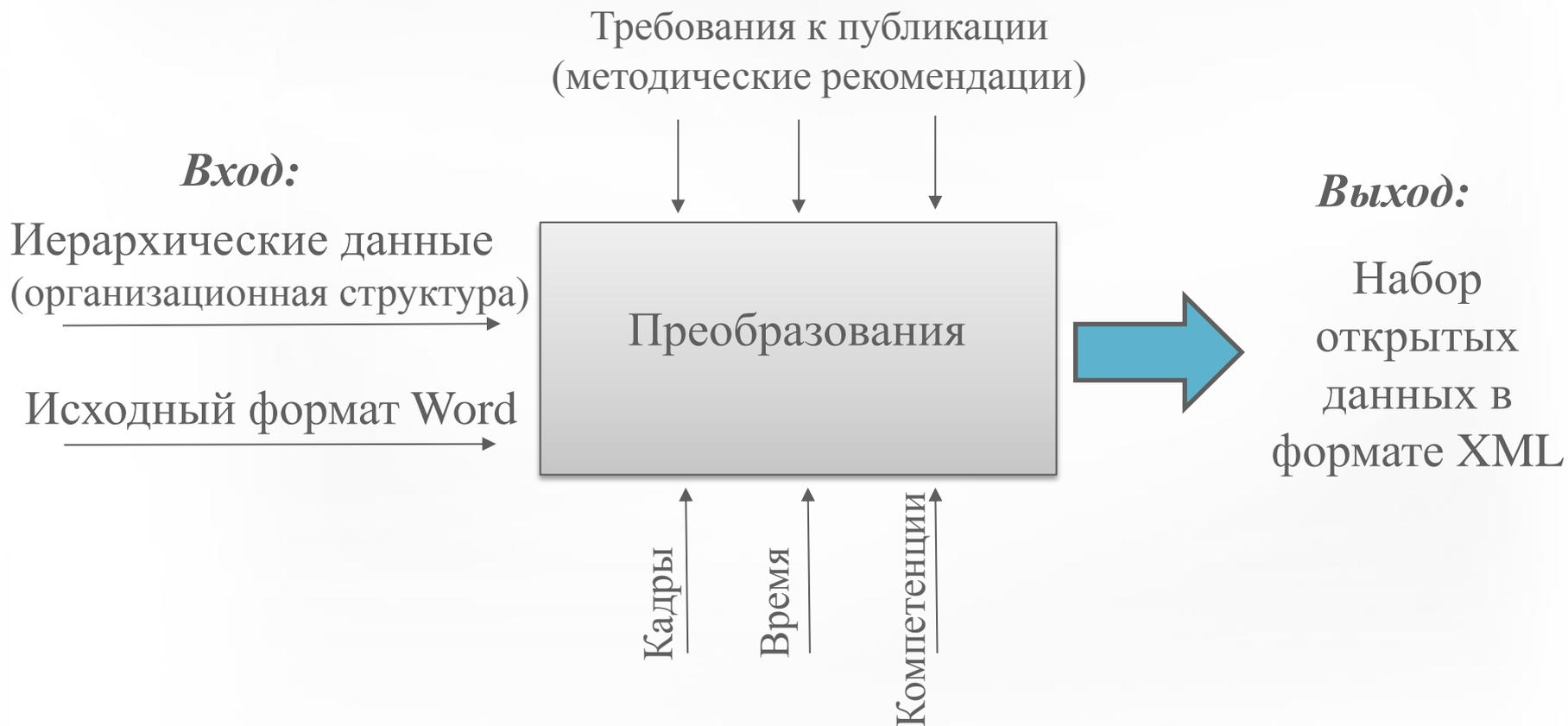
Результат: данные в машиночитаемом формате CSV

```
Набор данных в машиночитаемом формате CSV — Блокнот
Файл Правка Формат Вид Справка
name; district; area; address; telephone
ГБУК г, Москвы «Выставочный зал «Солянка ВПА»;Центральный административный округ;Басманный район;улица Солянка, дом 1/2, строение 2;"(495) 621-55-55"
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 АЗ»;Центральный административный округ;район Арбат;Старокожухинский переулок, дом 39;"(495) 697-55-55"
ГБУК г, Москвы «ГВЗ «Ковчег»;Северный административный округ;Тимирязевский район;улица Немчинова, дом 12;"(499) 977-00-44; (499) 977-00-88"
```

Описание структуры набора открытых данных

- field name, english description,russian description, format name,Name of showroom.,Название выставочного зала.,string
- district, Administrative district of showroom.,Административный округ в котором расположен выставочный зал.,string
- area, Municipal area of showroom.,Муниципальный район в котором расположен выставочный зал.,string
- address, "Street, building, structure (if any), where showroom is located.", "Улица, дом, строение (если существует), где расположен выставочный зал.",string
- telephone, "Contact phone numbers of showroom in format (xxx) ууу-уу-уу, separated by "; ", where ""xxx"" is phone area code.", "Контактные телефоны выставочного зала в формате (xxx) ууу-уу-уу, разделенные между собой знаком "; ", где ""xxx"" - телефонный код города.",string

Кейс №2



Требования к публикации набора открытых данных в машиночитаемом формате XML

1. Требования к форматам файла набора и структуры открытых данных:

1) Соответствие файла стандарту W3C – «Extensible Markup Language (XML) 1.1 (Second Edition)» (<http://www.w3.org/TR/xml11/>).

2) Кодировка файла указывается в объявлении XML если объявление XML используется, иначе используется кодировка «UTF-8».

2. Требование к структуре файла набора открытых данных:

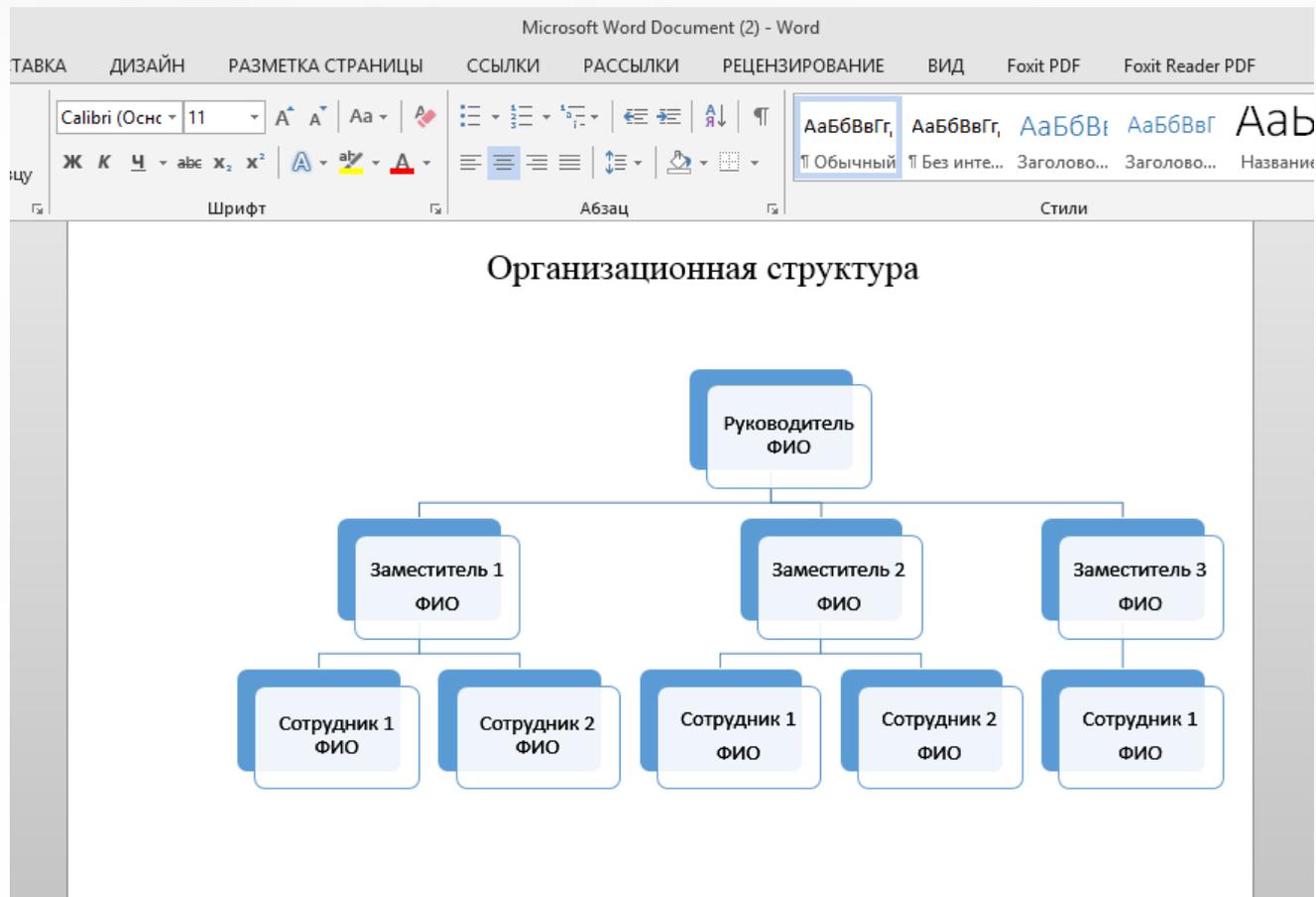
1. Файл набора открытых данных соответствует файлу структуры набора открытых данных.

Решение кейса

Действия:

1. Выбрать исходные данные для публикации
2. Перевести данные в XML ручным способом;
3. Составить структуру полученных данных в XML;
4. Проверить на соответствие полученных данных составленной структуре.

Пример исходных иерархических данных для публикации (в формате Word)



Описание данных в XML ручным способом

```
<?xml version="1.0" encoding="utf-8"?>
<root>
  <employee>
    <name>ФИО</name>
    <job>Руководитель</job>
  <emps>
    <employee>
      <name>ФИО</name>
      <job>Заместитель 1</job>
    <emps>
      <employee>
        <name>ФИО</name>
        <job>Работник 1</job>
      </employee>
      <employee>
        <name>ФИО</name>
        <job>Работник 2</job>
      </employee>
    </emps>
  </employee>
  <employee>
    <name>ФИО</name>
    <job>Заместитель 3</job>
  <emps>
    <employee>
      <name>ФИО</name>
      <job>Работник 1</job>
    </employee>
  </emps>
</root>
```

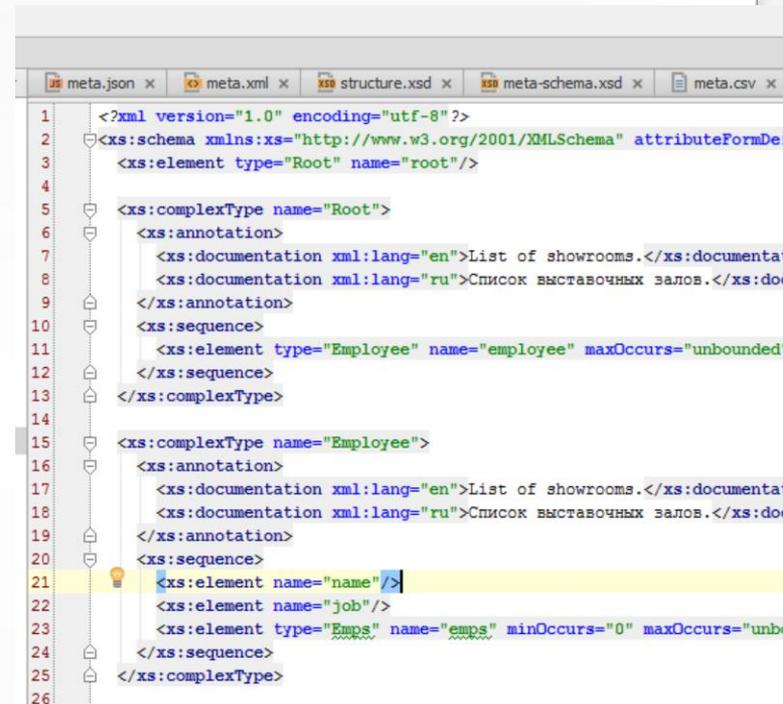
```
1 <?xml version="1.0" encoding="utf-8"?>
2 <root>
3   <employee>
4     <name>Петя</name>
5     <job>Руководитель</job>
6   <emps>
7     <employee>
8       <name>Вася</name>
9       <job>Заместитель 1</job>
10    <emps>
11      <employee>
12        <name>Коля</name>
13        <job>Работник 1</job>
14      </employee>
15      <employee>
16        <name>Коля</name>
17        <job>Работник 2</job>
18      </employee>
19    </emps>
20  </employee>
```

Описание структуры данных в XML ручным способом

```
<xs:complexType name="Root">
  <xs:annotation>
    <xs:documentation xml:lang="en">List of showrooms.</xs:documentation>
    <xs:documentation xml:lang="ru">Список выставочных залов.</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element type="Employee" name="employee" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>
```

```
<xs:complexType name="Employee">
  <xs:annotation>
    <xs:documentation xml:lang="en">List of showrooms.</xs:documentation>
    <xs:documentation xml:lang="ru">Список выставочных залов.</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="name"/>
    <xs:element name="job"/>
    <xs:element type="Emps" name="emps" minOccurs="0" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>
```

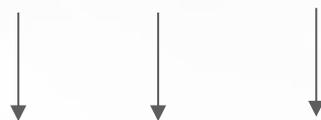
```
<xs:complexType name="Emps">
  <xs:annotation>
    <xs:documentation xml:lang="en">List of showrooms.</xs:documentation>
    <xs:documentation xml:lang="ru">Список выставочных залов.</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element type="Employee" name="employee" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>
</xs:schema>
```



```
1 <?xml version="1.0" encoding="utf-8" ?>
2 <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" attributeFormDe
3   <xs:element type="Root" name="root"/>
4
5   <xs:complexType name="Root">
6     <xs:annotation>
7       <xs:documentation xml:lang="en">List of showrooms.</xs:documenta
8       <xs:documentation xml:lang="ru">Список выставочных залов.</xs:do
9     </xs:annotation>
10    <xs:sequence>
11      <xs:element type="Employee" name="employee" maxOccurs="unbounded
12    </xs:sequence>
13  </xs:complexType>
14
15  <xs:complexType name="Employee">
16    <xs:annotation>
17      <xs:documentation xml:lang="en">List of showrooms.</xs:documenta
18      <xs:documentation xml:lang="ru">Список выставочных залов.</xs:do
19    </xs:annotation>
20    <xs:sequence>
21      <xs:element name="name"/>
22      <xs:element name="job"/>
23      <xs:element type="Emps" name="emps" minOccurs="0" maxOccurs="unb
24    </xs:sequence>
25  </xs:complexType>
26
```

Кейс №3

Требования к публикации
(методические рекомендации)

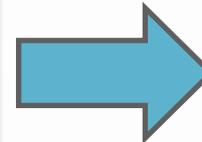
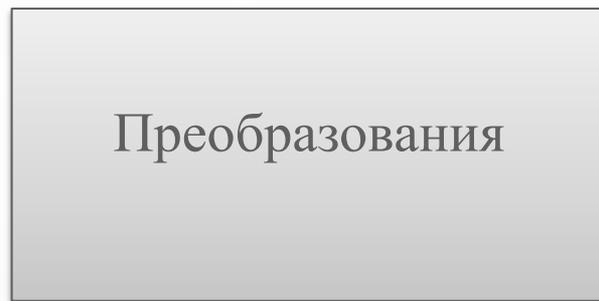


Вход:

Геоданные



Исходный формат PDF



Выход:

Набор
открытых
данных в
формате
JSON

Кадры



Время



Конвертер



Требования к публикации набора открытых данных в машиночитаемом формате JSON

- 1) Требования к форматам файла набора и файла структуры набора открытых данных:
 - ✓ Соответствие файла стандарту JSON – (<http://json.org/json-ru.html>).
 - ✓ Кодировка файла – «UTF-8».

- 2) Требование к структуре файла набора открытых данных:
 - Файл набора открытых данных соответствует файлу структуры набора открытых данных.

Решение кейса

Действия:

1. Выбрать исходные данные для публикации
2. Перенести данные в Excel
3. Описать поля: название, метро, адрес, широта, долгота
4. Определить долготу и широту
(<http://api.yandex.ru/maps/tools/getlonglat/>)
5. Данные из Excel скопировать в конвертер
(<http://shancarter.github.io/mr-data-converter/>)
6. Поставить Разделитель (Delimiter) – Tab; Десятичный разделитель (Decimal Sign) – Dot (точка).

Пример геоданных в формате PDF

Исходные данные:

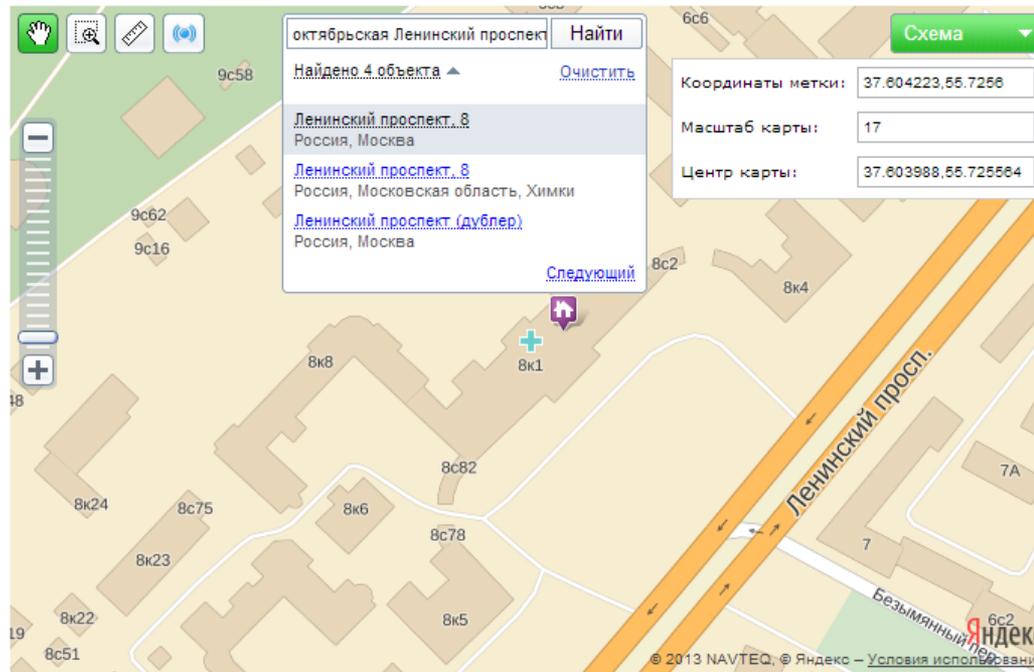
1. Городская клиническая больница № 1 им. Н.И. Пирогова Октябрьская г. Москва, Ленинский проспект, д. 8 236-80-66, 236-91-15 236-65-35
2. Городская клиническая больница № 2 им. Н.И. Пирогова Октябрьская г. Москва, Ленинский проспект, д. 8 236-85-78, 236-73-20 236-45-17

Определение координат места

Определение координат

Введите в строке поиска название географического объекта или перетащите метку в то место, координаты которого вы хотите узнать.

В правом верхнем углу карты будут показаны: координаты метки в формате "долгота, широта", текущий масштаб карты и координаты центра карты.



Перенос данных в Excel

The screenshot shows the Microsoft Excel interface with the following elements:

- Title Bar:** Microsoft Excel Worksheet - Excel
- File Menu:** ФАЙЛ
- Home Tab:** ГЛАВНАЯ
- Other Tabs:** ВСТАВКА, РАЗМЕТКА СТРАНИЦЫ, ФОРМУЛЫ, ДАННЫЕ, РЕЦЕНЗИРОВАНИЕ, ВИД, Foxit Reader PDF, Foxit PDF
- Font Group:** Calibri, 11, Bold (Ж), Italic (К), Underline (Ц), Color (A), Background Color (D)
- Paragraph Group:** Alignment (center, left, right, justified), Bullets, Numbering
- Number Group:** General (Общий), Percentage (%), Thousand Separator (000), Decimal Places (0,00)
- Styles Group:** Conditional Formatting (Условное форматирование), Format as Table (Форматировать как таблицу), Cell Styles (Стили ячеек)
- Clipboard Group:** Paste (Вставить), Delete (Удалить), Format (Формат)
- Editing Group:** Sum (Σ), Sort & Filter (Сортировка и фильтр), Find & Select (Найти и выделить)
- Formula Bar:** E10, X, Y, fx
- Worksheet Grid:** Columns A-I, Rows 1-8. Data is present in rows 1-3.

	A	B	C	D	E	F	G	H	I
1	Название	Метро	Адрес	Телефон	Долгота	Широта			
2	Городская клиническая больница № 1 им. Н.И. Пирогова	Октябрьская	Ленинский прос:	236-80-66	37.604223	55.7256			
3	Городская клиническая больница № 2 им. Н.И. Пирогова	Октябрьская	Ленинский прос:	236-80-66	37.604223	55.7256			
4									
5									
6									
7									
8									

Работа с конвертером

Mr. Data Converter

I will convert your Excel data into one of several web-friendly formats, including HTML, JSON and XML.

Fork me on [github](#).

SETTINGS

Delimiter: Auto Comma Tab

Decimal Sign: Dot Comma

First row is the header

Transform: downcase upcase none

Include white space in output

Indent with: tabs spaces

Input CSV or tab-delimited data. Using Excel? Simply copy and paste. No data on hand? [Use sample](#)

Название	Метро	Адрес	телефон	Долгота	Широта
Городская клиническая больница №1 им. Н.И. Пирогова	Октябрьская	Ленинский проспект, д.8	37.604223, 55.7266	236-80-66	236-91-15
Городская клиническая больница №1 им. Н.И. Пирогова	Октябрьская	Ленинский проспект, д.8	37.604223, 55.7266	236-80-66	236-91-15

Output as **JSON - Row Arrays**

```
[
  ["Городская клиническая больница №1 им. Н.И. Пирогова", "Октябрьская", "Ленинский проспект, д.8", "236-80-66", "236-91-15", 37.604223, 55.7266],
  ["Городская клиническая больница №1 им. Н.И. Пирогова", "Октябрьская", "Ленинский проспект, д.8", "236-80-66", "236-91-15", 37.604223, 55.7266]
]
```

Спасибо за внимание!