

Правила работы с системой, данными и др. вопросы

Подготовка исходных файлов

За основу берем <http://dataprotocols.org/tabular-data-package/>

1. Имя файла должно отображать значение датасета и записаны английским языком или латиницей (английский язык предпочтительней).
2. Основной формат данных - csv (в excel формате): разделитель - запятая, поля со спец символами оборачиваются в кавычки, кавычки экранируются кавычками, кодировка - UTF-8.
3. Во входном csv-файле в полях значений должны быть или цифры или ничего. Символы вроде "-" или пробелы недопустимы. Десятичные знаки отделяются точкой.
4. Заголовки колонок должны быть записаны английским языком или латиницей (английский язык предпочтительней) в стиле myField (в нижнем регистре, слова разделяются подчеркиванием). В описании ресурса необходимо приложить легенду, расшифровывающую названия полей.
5. Пример [неправильного и правильного оформления](#) набора данных для загрузки на портал.
6. Заголовок поля с датой должен иметь название "date", в не зависимости от того, содержаться ли в нем только года, или полные даты.
7. Заголовок поля с датой и временем должен иметь название "datetime", в не зависимости от точности времени.
8. **Предложение на обсуждение:** Широта и долгота обозначается как latitude, longitude ИЛИ массивом ИЛИ geojson.

Удаление ресурсов, групп и датасетов

Для правильного удаления элементов, их необходимо "отвязать" друг от друга перед удалением.

Удаление ресурсов из датасетов

Зайти на страницу редактирования датасета, убрать связь с ресурсом, сохранить. Не наоборот!

Зайти на страницу еще раз. Удалить ресурс.

Удаление датасетов

Отвязать все ресурсы от датасета

Отвязать датасет от групп

Удалить датасет

Загрузка данных и удаление набора данных

При загрузке данных необходимо точно указать разделитель данных. Это касается как загрузки ресурса на сайт, так и загрузки ресурса в Хранилище данных.

Для правильного удаления набора данных необходимо предварительно удалить все отношения между объектами. Например, перед удалением набора данных нужно отвязать его от всех связанных с ним ресурсов.

Только датасеты привязываются к группам, ресурсы не привязываются (хотя такое поле может присутствовать) - они наследуют привязку родителя.

3. Предложение обсудить представление гео данных, а также затронуть адреса:

1. geojson

1. загружаем все данные в табличном виде (кроме pdf и тд)
2. гео объект описывается с помощью сжатого (без пробельных символов) [geojson](#) в соответствующем поле, для точки допустимо как пара полей lat/lon
3. гео объектом может являться любой geojson объект за исключением Feature, FeatureGroup и CRS
4. координаты должны представляться в проекции spherical merkator (или web merkator, данная проекция используется по умолчанию для leaflet или google maps)
5. дополнительно можно указываться степень упрощения геометрии tolerance при содержании одних и тех же объектов с разным упрощением, если геометрия без упрощения, то это поле должно быть пустым

2. address

1. при указании адреса должен указываться полный адрес
страна, область, район, город, улица (ОСМ), дом, квартира (квартира может быть комнатой, кабинетом и тд) в отдельном поле
2. при масштабе дома квартира не указывается, при масштабе улицы дом и тд не указывается, при масштабе города улица и тд не указывается, при масштабе района город и тд не указывается, при масштабе области район и тд не указывается, при масштабе страны область и тд не указывается

3. есть неоднозначность с одинаковыми названиями улиц в одном городе, так что возможно в этих случаях потребуется дополнительное поле, можно решать по мере проявления
4. по умолчанию должны соблюдаться официальное административное деление, иначе данная информация должна указываться для ресурса/датасета
 1. минск не входит в минскую область, те область, район, город для основного датасета будут одинаковыми
 2. минск, областные города и некоторые другие не входят в состав районов, те район, город для них будут одинаковыми
 3. для описанных выше примеров имеет смысл сделать отдельные датасеты с включением городов в область и регионы, с соответствующей информацией
3. поля и атрибуты:
 1. все загружаемые данные должны задаваться в utf-8 кодировке
 2. атрибуты описываются в стиле camelCase на латинице (желательно на английской)
 3. для идентификаторов OSM **osmid**: node - **n12345**, way - **w12345**, relation - **r12345**
 4. для страны - **country**
 5. для области - **region**
 6. для района - **subregion**
 7. для города - **city**
 8. для улицы - **street**
 9. для дома - **house**
 10. для квартиры - **apartment**

Документация по API портала

Документация находится здесь:

Иное

Словарь терминов

Набор данных (dataset)

Ресурс

Файл

API

метаданные (metadata)

Данные (data)