Описание набора данных

**«База данных муниципальных образований»**

# История изменений

|  |  |  |  |
| --- | --- | --- | --- |
| **Дата** | **Версия** | **Автор** | **Описание изменений** |
| 12.11.2020 | 1.0 | Уксусова Арина Сергеевна | Документ создан |
| 13.11.2020 | 1.1 | Копыток Витовт Константинович | Расширено описание в Разделе 5 о проведенных проверках. |

Оглавление

[История изменений 1](#_Toc54800971)

[Раздел 1. Общее описание набора данных 3](#_Toc54800972)

[Раздел 2. Источники данных 5](#_Toc54800973)

[Раздел 3. Внесенные в набор данных изменения по сравнению с источниками 8](#_Toc54800974)

[Описание изменений, внесенных в набор, по сравнению с источниками 8](#_Toc54800975)

[Раздел 4. Структура набора данных и примеры обработки 9](#_Toc54800976)

[Codebook 9](#_Toc54800977)

[Параметры запросов API 10](#_Toc54800978)

[Стандартные процедуры обработки 10](#_Toc54800979)

[Раздел 5. Известные ограничения данных 11](#_Toc54800980)

[Описание проведенных проверок и их результатов 11](#_Toc54800981)

[Качество данных 11](#_Toc54800982)

[Полнота данных 11](#_Toc54800983)

[Раздел 6. Ответственный за набор данных 13](#_Toc54800984)

# Раздел 1. Общее описание набора данных

|  |  |
| --- | --- |
| **Атрибут** | **Значение** |
| ***Общие сведения*** | |
| Полное наименование набора данных | База данных муниципальных образований |
| Сокращенное наименование набора данных | БДМО |
| Краткое описание набора данных | База данных муниципальных образований включает в себя консолидированные и предварительно обработанные данные муниципальной статистики Росстата за 2006-2018 годы. Данные муниципальной статистики собраны из открытых источников: раздел «Базы данных» Федеральной службы государственной статистики РФ, база данных «Показатели муниципальной статистики». Сбор данных осуществлялся напрямую с сайта с помощью web scraping методов. |
| Доступные форматы | csv |
| Размер набора данных для скачивания | 1.6 ГБ |
| Словари и справочники | ОКТМО (Общероссийский классификатор территорий муниципальных образований) |
| Ключевые слова (теги) | муниципальные образования, муниципальная статистика, экономика, социальная сфера |
| Тематика набора данных | Статистика |
| Стандарт публикации | <http://opendata.gosmonitor.ru/standard/3.0> |
| Краткое описание источников наборов данных | Данные муниципальной статистики, характеризующие состояние экономики и социальной сферы муниципального образования, представленные на официальном сайте Федеральной службы государственной статистики РФ (Росстат). |
| ***Периодичность публикации и обновления*** | |
| Дата размещения | 28.09.2020 |
| Покрываемый временной период | 2006-2018 |
| Периодичность обновления | не обновляется |
| Дата последнего обновления набора данных | 01.01.2020 |
| Характер обновления | не обновляется |
| Дата следующего обновления набора данных | не обновляется |
| ***Дополнительные сведения*** | |
| Язык набора | русский |
| Лицензия, под которой публикуется набор данных | CC BY-NC-SA (Creative Commons)  https://creativecommons.org/licenses/by-nc-sa/4.0/deed.ru  Допускается использование данных только в научно-исследовательских целях. Не допускается копирование и распространение данных. |
| Цитирование набора данных | "База данных муниципальных образований". **Источник: Росстат; обработка:** Инфраструктура научно-исследовательских данных, АНО «ЦПУР», 2020. (Ссылка на набор данных: <https://data-in.ru/data-catalog/datasets/115/>) |
| Режим доступа | Открытый (у Вас есть возможность работать с данными на своем личном ПК, в том числе и скачивать данные) |
| Геоданные (пространственная привязка набора данных) | да |

# Раздел 2. Источники данных

**«База данных муниципальных образований»**

|  |  |
| --- | --- |
| **Атрибут** | **Значение** |
| Полное наименование источника данных | База данных показателей муниципальных образований |
| Сокращенное наименование источника данных | БД ПМО |
| Владелец источника данных | Федеральная служба государственной статистики РФ (Росстат) |
| Краткое описание источника данных | В соответствии с разделом 1.33 «Муниципальная статистика» Федерального плана статистических работ (утв. распоряжением Правительства Российской Федерации от 6 мая 2008 г. № 671-р с последующими изменениями) разрабатывается официальная статистическая информация, находящаяся в компетенции Росстата, в разрезе муниципальных образований. База данных статистических показателей, характеризующих состояние экономики и социальной сферы муниципального образования (БД ПМО), включающая информацию по каждому из 21,9 тысячи муниципальных образований (по состоянию на 1 января  2018 г.), сформирована на основании статистической информации, разрабатываемой по муниципальным образованиям.  При размещении официальной статистической информации в БД ПМО обеспечивается исполнение требований Федерального закона Российской Федерации от 29 ноября 2007 г. № 282-ФЗ «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» (с изменениями в ред. Федерального закона Российской Федерации от  02 июля 2013 г. № 171-ФЗ) в части гарантии конфиденциальности первичных данных респондентов.  Данные муниципальной статистики размещаются на официальном сайте Росстата в формате html. Данные были собраны с сайта Федеральной службы государственной статистики с помощью web-scraping методов и обработаны для преобразования в машиночитаемый формат. |
| Ссылка на источник данных | Раздел «Базы данных» на официальном сайте Федеральной службы государственной статистики РФ (Росстат): <https://rosstat.gov.ru/storage/mediabank/munst.htm> |
| Перечень используемых классификаторов | ОКТМО |
| Понятия, используемые в источнике | Муниципальное образование – это населенная территория, на которой осуществляется местное самоуправления в Российской Федерации, т.е. населением непосредственно и (или) через органы местного самоуправления решаются преимущественно вопросы местного значения. В соответствии с федеральным законом от 06.10.2003 N 131-ФЗ (ред. от 20.07.2020) «Об общих принципах организации местного самоуправления в Российской Федерации» (с изм. и доп., вступ. в силу с 28.08.2020) такими населенными территориями являются: городское или сельское поселение, муниципальный район, муниципальный округ, городской округ, городской округ с внутригородским делением, внутригородской район либо внутригородская территория города федерального значения  ОКТМО (Общероссийский классификатор территорий муниципальных образований) входит в состав национальной системы стандартизации Российской Федерации. ОКТМО предназначен для обеспечения систематизации и однозначной идентификации на всей территории Российской Федерации муниципальных образований и населенных пунктов, входящих в их состав, с отражением структуры и уровней территориальной организации местного самоуправления для решения задач сбора, автоматизированной обработки, представления и анализа информации в разрезе муниципальных образований в различных областях экономики, включая статистику, прогнозирование, налогообложение. (в соответствии с «ОК 033-2013. Общероссийский классификатор территорий муниципальных образований" (Том 1. Центральный федеральный округ) (утв. Приказом Росстандарта от 14.06.2013 N 159-ст) (с учетом Изменений 1/2013 - 418/2020)) |
| НПА, регулирующие сбор данных в источнике данных | * Федеральный закон от 29.11.2007 N 282-ФЗ (ред. от 18.04.2018) «Об официальном статистическом учете и системе государственной статистики в Российской Федерации»; * Распоряжение Правительства РФ от 06.05.2008 N 671-р (ред. от 05.11.2020) «Об утверждении Федерального плана статистических работ» (вместе с "Федеральным планом статистических работ"); * Федеральный закон от 02.07.2013 N 171-ФЗ «О внесении изменений в Федеральный закон «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» и отдельные законодательные акты Российской Федерации». |
| Ссылки на НПА | <http://www.consultant.ru/document/cons_doc_LAW_72844/>  <http://www.consultant.ru/document/cons_doc_LAW_111344/e3f5282084b6a32b834009ecedd4d19dacfdf386/>  <http://www.consultant.ru/document/cons_doc_LAW_148482/3d0cac60971a511280cbba229d9b6329c07731f7/> |
| Методология и изменения методологии для источника данных | Подробная информация о методологии формирования показателей муниципальной статистики представлена в Кратком методологическом комментарии к Базе данных показателей, характеризующих состояние экономики и социальной сферы муниципального образования (БД ПМО) |
| Ссылки на методологию | <http://www.gks.ru/free_doc/new_site/bd_munst/bdmo_met2019.doc> |

# Раздел 3. Внесенные в набор данных изменения по сравнению с источниками

|  |  |
| --- | --- |
| **Атрибут** | **Значение** |
| Факт внесения изменений по сравнению с источниками | Есть |

## Описание изменений, внесенных в набор, по сравнению с источниками

* Данные представлены в формате единой базы. Панельная структура оформлена в виде «длинного» формата, где единицей наблюдения (строка в базе) является «муниципалитет-год»;
* Изменены названия некоторых муниципальных образований для приведения к единой системе наименования и построения полных и непрерывных временных рядов;
* Проведена предварительная обработка данных по очистке несоответствующих значений. Такие значения были заменены на пропуски в данных.

# Раздел 4. Структура набора данных и примеры обработки

|  |  |
| --- | --- |
| **Атрибут** | **Значение** |
| Структура набора данных | 3 плоских таблицы:  таблица 1: 300784 наблюдений, 338 атрибутов  таблица 2: 286149 наблюдений, 2468 атрибутов  таблица 3: 124219 наблюдений, 4324 атрибутов |
| Ссылка на открытый репозиторий | нет |

## Codebook

**Таблица «База данных муниципальных образований»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Атрибут** | **Описание** | **Единица измерения** | **ОКЕИ** | **Формат** |
| mun\_type | Тип муниципального образования |  |  | string |
| municipality | Название муниципалитета |  |  | string |
| oktmo | Код муниципалитета в соответствии с ОКТМО |  |  | numeric |
| oktmo\_munr | Код муниципального района в соответствии с ОКТМО |  |  | numeric |
| rayon | Название муниципального района |  |  | string |
| region | Название субъекта РФ |  |  | string |
| year | Год |  |  | numeric |
| атрибут типа t8001004[[1]](#footnote-2) | Показатель муниципальной статистики нулевого уровня вложенности |  |  | string |
| атрибут типа t8001001\_11000[[2]](#footnote-3) | Показатель муниципальной статистики первого уровня вложенности |  |  | string |
| атрибут типа t8007018\_kategor16\_kultur3000009[[3]](#footnote-4) | Показатель муниципальной статистики второго уровня вложенности |  |  | string |

Все показатели, представленные в наборе данных, разбиты на три типа:

* показатели нулевого уровня вложенности;
* показатели первого уровня вложенности;
* показатели второго уровня вложенности.

Например, для группы показателей, характеризующих численность населения, показателем нулевого уровня вложенности является показатель «Численность населения, всего», первого уровня вложенности — «Численность мужского населения, всего», второго уровня вложенности — «Численность мужского населения в возрасте 25 лет, всего». В названиях показателей уровни вложенности также отражаются. Например, для показателей первого уровня вложенности их название состоит из двух частей, разделенных нижним подчеркиванием, первая часть названия совпадает с названием показателя соответствующего нулевого уровня вложенности. В названиях показателей второго уровня вложенности последовательно отражены названия показателей нулевого и первого уровня вложенностей, разделенные нижним подчеркиванием.

## Параметры запросов API

Интеграция через API не предусмотрена.

# Раздел 5. Известные ограничения данных

## Описание проведенных проверок и их результатов

Переменные, фиксирующие название населенных пунктов, были проверены на корректность написания и соответствие ОКТМО: найденные неточности были исправлены путем изменения названий некоторых муниципальных образований с целью приведения к единой системе наименований.

Данные были проверены на наличие несоответствующих смыслу показателей значений по тематическим группам (например, местный бюджет, финансовая деятельность, сельское хозяйство, население, организация отдыха, развлечений и культуры и др.), в которые входят отдельные показатели муниципальной статистики. Для каждого конкретного показателя, входящего в определенную тематическую группу, проверялся ряд условий, которым значение показателя должно соответствовать, исходя из его смысла. Среди них:

* + Значения должны быть целочисленными;
  + Значения должны быть не отрицательными;
  + Значения не должны превышать 100;
  + Значения должны быть положительными;
  + Значения должны быть бинарными.

При этом не обязательно, чтобы значения конкретного показателя удовлетворяли каждому из перечисленных выше условий. Информация о том, на соответствие каким условиям проверялись значения каждого конкретного показателя, содержится в колонках ***«nonnegative»*** (условие неотрицательности значений), ***«greater0»*** (условие положительности значений), ***«integer»*** (условие целочисленности значений), ***«leq100»*** (условие того, что значение не превышает 100), ***«binary»*** (условие бинарности значений) файла ***cleaning\_info.csv.*** Значение 1 в одной из этих колонок в конкретной строке означает, что по смыслу показатель, который отражен в соответствующей строке, должен удовлетворять этому условию.

Кроме того, значения некоторых переменных были проверены на наличие факта изменения в шкале измерения (единице измерения), когда такое изменение происходило (отметка о том, что такое изменение должно наблюдаться для конкретного показателя содержится в колонке ***«change»***).

Информация о результатах проведенных проверок представлена в файле ***cleaning\_info.csv***, который доступен для скачивания вместе с набором данных. Каждая строка этого файла соответствует конкретному показателю, для которого проводилась проверка

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Атрибут** | **Описание** | **Единица измерения** | **ОКЕИ** | **Формат** |
| indicator\_all | Наименование показателя. |  |  | string |
| indicator\_id | Идентификатор показателя. |  |  | string |
| indicator\_id\_combined | Идентификатор показателя (дополнительный). Является технической колонкой, использовалась для контроля совпадения идентификаторов одних и тех же показателей в разных субъектах РФ. |  |  | string |
| indicator\_id\_changed\_flag | Отметка о том, что для некоторых субъектов идентификатор показателя отличается от того, который используется для в остальных субъектах. |  |  | boolean |
| region | Наименование субъекта РФ, для которого встречаются рассматриваемые проблемы для приведенного в строке показателя. Наименование субъекта приведено в качестве примера, такие же проблемы могут встречаться в других субъектах. |  |  | string |
| group | Группа показателей |  |  | string |
| indicator | В случае небольшой разницы в названиях показателей в разных субъектах проводилась проверка на соотнесение этих показателей по смыслу. |  |  | string |
| indicator\_status | Отметка о том, что показатели из ***indicator\_all*** и ***indicator*** совпадают по смыслу. |  |  | string |
| unit | Единица измерения показателя. |  |  |  |
| anom | Текстовый комментарий о проведенных проверках. |  |  | string |
| nonnegative | Отметка о том, что по смыслу значения показателя должны быть неотрицательными. |  |  | boolean |
| greater0 | Отметка о том, что по смыслу значения показателя должны быть положительными. |  |  | boolean |
| integer | Отметка о том, что по смыслу значения показателя должны быть целочисленными. |  |  | boolean |
| leq100 | Отметка о том, что по смыслу значения показателя не должны превышать 100. |  |  | boolean |
| change | Отметка о том, что в определенный момент должно произойти изменение единицы измерения. |  |  | boolean |
| binary | Отметка о том, что по смыслу значения показателя должны быть бинарными. |  |  | boolean |
| nonnegative\_exclusion | Количество наблюдений, удаленных из выборки из-за несоблюдения условия на неотрицательность значения. Учитывались также удаленные вложения переменной. |  |  | numeric |
| leq100\_exclusion | Количество наблюдений, удаленных из выборки из-за несоблюдения условия на значения, не превышающие 100. Учитывались также удаленные вложения переменной. |  |  | numeric |
| integer\_exclusion | Количество наблюдений, удаленных из выборки из-за несоблюдения условия на целочисленность значений. Учитывались также удаленные вложения переменной. |  |  | numeric |
| num\_of\_vars | Количество вложений в индикатор (значение 1 соответствует отсутствию вложений, 2 — одному уровню вложенности, 3 — двум уровням вложенности). |  |  | numeric |

## Качество данных

По результатам проверки и дополнительной корректировки данных, не все найденные и возможные проблемы в данных были устранены.

В частности, при работе с набором данных следует учитывать, что, предобработка (очистка) данных была произведена только по трем условиям, указанным выше (за исключением атрибута «Доходы бюджета», для которого встречаются отрицательные значения и такие случаи не удалялись).

Не обработаны значения по атрибутам, для которых, исходя из смысла, значения могут быть только положительными. Возможность наличия нулевых значений целесообразно рассматривать для каждого атрибута отдельно. Также не произведены изменения внутри переменных, для которых в определенные моменты менялась шкала расчета/измерения Пример: до определенного момента переменная измерялась в метрах, затем стала измеряться в километрах: для более подробной информации см. атрибут ***«change»*** в файле ***cleaning\_info.csv***.

При анализе для каждой рассматриваемой переменной целесообразно строить график изменения среднего по времени, чтобы понять встречаются ли резкие пики/спады, которые могут свидетельствовать об изменении методологии.

Также данные не корректировались с точки зрения удаления экстремальных значений, если переменная не имеет ограничения сверху, поэтому при работе с данными следует проводить анализ на наличие выбросов.

## Полнота данных

В данных содержатся пропуски по некоторым из показателей муниципальной статистики. Это связано, во-первых, с качеством сбора данных в первоисточнике, во-вторых, с тем, что данные по ряду показателей могли собираться только за определенные промежутки времени (годы): например, данные по показателю «Количество населенных пунктов, не имеющих водопроводов (отдельных водопроводных сетей)» собирались в 2017-2018 гг., данные за период с 2006 по 2016 гг. не представлены в источнике.

Также на полноту данных влияет тот факт, что, набор данных в настоящий момент не предполагает обновления, поэтому последний год, за который представлены данные — 2018.

Следует учесть тот факт, что существует проблема смены названий муниципальных образований, упразднения или выделения новых муниципалитетов, изменение кодов ОКТМО, поэтому при попытке дополнить данные БДМО более свежими данными муниципальной статистики, следует учитывать то, что с момента сбора БДМО могли произойти изменения в кодах ОКТМО, в названиях и количестве муниципальных образований, но эти изменения не зафиксированы в БДМО.

# Раздел 6. Ответственный за набор данных

**Ответственный 1**

|  |  |
| --- | --- |
| **Атрибут** | **Значение** |
| ФИО ответственного лица | Уксусова Арина Сергеевна |
| Телефон ответственного лица | не указано |
| Электронная почта ответственного лица | [a.uksusova@cpur.ru](mailto:a.uksusova@cpur.ru) |

**Ответственный 2**

|  |  |
| --- | --- |
| **Атрибут** | **Значение** |
| ФИО ответственного лица | Казанцев Кирилл Ильич |
| Телефон ответственного лица | не указано |
| Электронная почта ответственного лица | k.kazantsev@cpur.ru |

**Ответственный 3**

|  |  |
| --- | --- |
| **Атрибут** | **Значение** |
| ФИО ответственного лица | Афанасенко Тарас |
| Телефон ответственного лица | не указано |
| Электронная почта ответственного лица | t.afanasenko@cpur.ru |

1. # Расшифровку атрибутов с различными уровнями вложенности можно найти в специальном файле BDMO\_variables.csv, который можно скачать с сайта «Инфраструктуры научно-исследовательских данных» ИНИД (<https://data-in.ru/>) вместе с набором данных «База данных муниципальных образований» по ссылке: <https://fs.data-in.ru/index.php/s/6ZrwEC5Ei48ascw/download>. Дополнительное описание данных и структуры файлов, представленных в каталоге ИНИД, можно найти в файле README.rtf, размещенном по адресу: <https://docs.data-in.ru/bdmo/csv/>. Скачать файл можно по ссылке: <https://fs.data-in.ru/index.php/s/7GGQRrZTxCfnPbq/download>.

   [↑](#footnote-ref-2)
2. См. сноску 1. [↑](#footnote-ref-3)
3. См. сноску 1. [↑](#footnote-ref-4)