

# Упътване за публикуване на данни в [opendata.government.bg](https://opendata.government.bg)

*и технически насоки за отваряне на данни от държавната  
администрация*

## **СЪДЪРЖАНИЕ**

### **1. АКРОНИМИ**

### **2. ЕЛЕКТРОННИ ФОРМАТИ ЗА ДАННИ**

2.1. Сравнение на формати

2.2. Отворени и затворени формати

2.2.1. Отворени формати

2.2.2. Затворени формати

2.3. Машинно-четими формати

2.4. Метаданни

2.5. Структурирани, полуструктурирани и неструктурирани формати

### **3. РАЗШИРЕНИ ОПИСАНИЯ НА ПОПУЛЯРНИ ФОРМАТИ**

3.1. CSV (Comma Separated Values)

3.2. JSON

3.3. RDF

3.4. XML

3.5. WMS

### **4. КАЧВАНЕ НА ДАННИ В ПОРТАЛА**

## 1. АКРОНИМИ

Акроним	Описание
ФСД	Фиксирана схема на данните
ГСД	Гъвкава (нефиксирана) схема на данните
ПНФ	Първа нормална форма
СУБД	Система за управление на бази данни

## 2. ЕЛЕКТРОННИ ФОРМАТИ ЗА ДАННИ

### 2.1. Сравнение на формати

Формат	Отворен	Машинно-четим	Мета-данни	Структуриран	Подходящи приложения	Подходящ метод за достъп
CSV	Да	Да	Не	Да	Таблицы в ПНФ с ФСД	Файл
HTML	Да	Да	Да	Полу	Форматирани данни	Файл
JPEG	Да	Не	Не	Не	Растрни изображения	Файл
JSON	Да	Да	Да	Да	Сериализирани данни с ФСД	Уеб-услуга (API)
KML	Да	Да	Да	Да	Гео-информация	Файл
PDF	Да	Не	Не	Не	Документи (вкл. подписани)	Файл
RDF	Да	Да	Да	Да	Свързани данни	Файл
TIFF	Да	Не	Не	Не	Растрни изображения	Файл
TXT	Да	Да	Не	Не	Неструктуриран текст	Файл
WMS	Да	Да	Да	Да	Гео-информация	Уеб-услуга (API)
XLS	Не	Да	Не	Полу	Таблицы в ПНФ с ФСД	Файл
XLSX	Да	Да	Не	Полу	Таблицы в ПНФ с ФСД	Файл
XML	Да	Да	Да	Да	Данни с ГСД	Уеб-услуга (API), Файл

### ПРИЕМЛИВИ ФОРМАТИ ЗА ПУБЛИКУВАНЕ В ОТВОРЕН ФОРМАТ: CSV, JSON, KML, RDF, WMS, XML

### 2.2. Отворени и затворени формати

#### 2.2.1. Отворени формати

Дори ако информацията е предоставена в електронен, машинно-четим формат, може да съществуват технически и/или юридически пречки за повторното използване и обработка на данните.

За да бъде считан един формат за „отворен“, то той трябва да отговаря на някои от следните условия:

- Съществуват публично достъпни и безплатни софтуерни библиотеки с Отворен код, осигуряващи функционалност за интерпретирането на формата и създаването на

независими приложения, които да могат да разчитат дадения формат, без каквито и да било ограничения произтичащи от авторски права или свързани с права върху интелектуална собственост.

- Съществуват публично достъпни и безплатни технически спецификации, позволяващи интерпретирането на формата и създаването на независим софтуер, който да може да разчита дадения формат, без каквито и да било ограничения, произтичащи от авторски права или свързани с права върху интелектуална собственост.

Обикновено за отворените формати, които са стандартизирани от международни организации, има публикувани технически спецификации, документация и примерни софтуерни библиотеки, подходящи за вграждане в приложения, без ограничения върху употребата.

Публикуването на данни в отворени формати от страна на организациите от общественения сектор улеснява създаването на полезни софтуерни приложения, системи и услуги, които дават възможност за повторната употреба на данните и информацията, за цели, различни от първоначално предвидените.

### **2.2.2. Затворени формати**

Един формат се счита за „затворен“, ако:

- Не съществуват публично достъпни и безплатни технически спецификации и/или софтуерни библиотеки, позволяващи интерпретирането на формата и създаването на независим софтуер, който да може да разчита дадения формат, или такива съществуват, но са налице лицензионни ограничения или други ограничения произтичащи от авторски права или права върху интелектуална собственост.
- Форматът е патентован, и въпреки че техническите му спецификации са публично достъпни и безплатни, и че свободната му употреба е позволена, използването му може да породви бъдещи претенции или ограничителни зависимости от притежателя на патента.

Публикуването на данни в затворени формати от страна на организациите от общественения сектор създава затруднения и потенциални рискове за ползвателите, тъй като използването и обработката на данните предполага зависимост от конкретни производители на софтуер или притежатели на патенти, които могат да налагат заплащането на авторски и/или лицензионни възнаграждения и/или да променят правата за ползване на формата по собствено усмотрение.

От гледна точка на принципите за Отворено управление, публикуването на данни и информация за повторна употреба трябва да се извършва в машинно-четими отворени формати.

### **2.3. Машинно-четими формати**

Всички електронни формати са машинно-четими, но с оглед на целите и ползите от публикуване на данни за повторна употреба, дефиницията е разширена - **машинно-четим е електронен формат за данни, който е структуриран по начин, по който, без да се преобразува в друг формат, позволява софтуерни приложения да**

**идентифицират, разпознават и извличат специфични данни, включително отделни факти и тяхната вътрешна структура.**

Например масово използваният файлов формат **PDF (Portable Document Format)** осигурява множество предимства при разпространение на документи, които могат да бъдат отваряни и визуализирани на разнообразни устройства, работещи с различни операционни системи и приложен софтуер, но този формат **не е подходящ за обмен на структурирани данни**, тъй като те няма как да бъдат извлечени и интерпретирани еднозначно по машинен път.

Често пъти PDF форматът се използва за съхранение на сканирани текстови документи, които са представени като набор от растерни графични изображения, което прави невъзможно извличането дори на неструктурирания текст, който изобразяват, без да се приложи допълнителна обработка, като например машинно оптично разпознаване на знаци (OCR). Запазването на структурирани таблици с данни в PDF формат е друга честа практика, която също създава сериозни пречки за повторно използване на данните, тъй като структурата на таблиците се губи при трансформиране в PDF формат.

Други формати като **CSV** и **JSON** са много по-подходящи за публикуване на информация за повторна употреба, тъй като представят данните в структуриран вид и са лесни за интерпретиране и машинна обработка.

## **2.4. Метаданни**

Някои електронни формати могат да съдържат **референция (URL) към външна схема**, съдържаща метаданни, описващи структурата и предназначението на същинските данни. Интерпретирането и валидирането на данните може да се осъществява напълно автоматично само на база на правилата дефинирани в схемата, като по този начин се гарантира еднозначност при извличането на отделни данни, без да е задължително да има налична документация към набора от данни. Примери за такива формати са **XML, JSON, WMS**.

Други по-прости формати като **CSV** предоставят ограничени възможности за съхранение и обмен на метаданни, но за сметка на това са много по-компактни, лесни и бързи за употреба, без да изискват използването на специфични и сложни софтуерни инструменти. **Недостатъкът на форматите, които не могат да включват метаданни, може да бъде компенсиран посредством прилагане на техническа спецификация и/или допълнителен файл с метаданни, описващ структурата и предназначението на файла със същински данни, който да играе ролята на документация.**

## 2.5. Структурирани, полуструктурирани и неструктурирани формати

### 3. Разширени описания на популярни формати

#### 3.1. CSV (Comma Separated Values)

CSV е компактен файлов формат, подходящ за обмен на големи набори от данни с фиксирана структура, представени в първа нормална форма.

Форматът е подходящ за директно представяне на данни експортирани от таблици в СУБД. По аналогия с таблиците в СУБД, отделните записи са на отделни редове, с фиксирана еднаква структура, а полетата (колониите) във всеки запис са отделени с разделител (обикновено „запетая“).

CSV форматът е много прост и не може да съдържа метаданни, които да описват схемата (структурата) на всеки запис от данните. Стандартно, първият ред от CSV файла съдържа имената на полетата (колониите), но този ред не може да съдържа допълнителни метаданни за предназначението, типа и размерността на всяко поле (колона). **Това налага всеки набор от данни в CSV формат да бъде придружен с отделен набор от машинно-четими метаданни или описателна документация, която да бъде интерпретирана от разработчика на софтуер.**

```
1 Код,Дата,Име на фирма,Седалище,Адрес на управление
2 4093-НА,28.02.2001,"ЕТ ДЖАМБО-5-М","гр. Велики Преслав","ул. Александър Стамболийски 54"
3 3319-НА,04.07.2000,"ЕТ ВЕСЕЛА ДЕЯНОВА","гр. Кюстендил","ул. Княз Александър Батенберг 30, вх. А"
4 3627-НА,30.10.2000,"ЕТ СТОЯН КЕРКЕНЕЗОВ","гр. Пловдив","ул. Звезда 47, ет. 1, ап. 3"
5 3897-НА,15.01.2001,"ЕТ ТОДОР КИРОВ-72","гр. Пловдив","жк. Тракия, бл.81, вх. Б, ет. 4, ап. 16"
6 3706-НА,28.11.2000,"ЕТ ДАЯНА-ЯНКА ВАСИЛЕВА","гр. Сливен","кв. Даме Груев, бл. 15, вх. Б"
7 3906-НА,15.01.2001,"ЕТ ФЛОРА-ЦВЕТА ИЛИЕВА","гр. Сливен","жк. Дружба, бл.22, вх. В, ап. 10"
8 3150-НА,04.07.2000,"ЕТ ГЕОРГИ ГЕРОВ","гр. Шумен","пл. Александър Стамболийски 24"
```

**При интерпретирането на данни в CSV формат е изключително важно да има налична документация, която да дефинира еднозначно структурата на данните,** тъй като ако структурата на даден ред се различава (например поради изпуснатата стойност), прочитането на данните до края на файла е невъзможно, поради невъзможност за еднозначна интерпретация на последващите редове.

#### 3.2. JSON

JSON е прост формат, който е лесен за прочитане чрез стандартни методи налични в повечето съвременни езици за програмиране. Простотата на използване и по-малките изчислителни ресурси за интерпретиране на данните са основни предимства спрямо по-сложни формати като XML.

**Схемата (структурата) на данните е фиксирана, което позволява много бърза и икономична обработка,** и прави този формат подходящ за онлайн обмен данни между веб-услуги (web services) в реално време.

### 3.3. RDF

RDF е стандартен семантичен формат за обмен на данни, дефиниран в група препоръки на W3C. **Подходящ е за представяне на данни и насочени / наименувани връзки между тях. Този формат дава възможност и за сливане на свързани набори от данни, които са комбинирани от няколко източника с различни схеми (структури) на данните**, като позволява и еволюция на схемите на данните, без да се налага модификация на консуматорите на данни. Отделните данни могат да бъдат съхранени в XML, JSON или в друг сериализиран вид.

RDF улеснява използването на унифицирани ресурсни локатори и идентификатори (URL и URI), което позволява директното обвързване на съществуващи набори от данни, публикувани по различни инициативи за Отворени данни в Интернет.

RDF все още не е широко разпространен, но има възходяща тенденция за по-широкото му приложение, особено в инициативите за Отворено управление и Отворени данни.

### 3.4. XML

XML е широко използван формат за машинен обмен на данни и осигурява добри възможности за съхраняване на сложни структури от данни, като **позволява еднозначното им интерпретиране и валидиране на база схема, която съдържа метаданни описващи съответните правила**. XML е подходящ при представяне на сложни набори от данни - отделните записи с данни могат да бъдат с променлива (нефиксирана) структура, форматът може да съдържа допълнителни структури от данни като номенклатури и др.

Обработката на данни в XML формат, особено когато става дума за по-големи набори от данни, може да създаде редица затруднения, поради по-големите изисквания за изчислителни ресурси и памет при парсирането (parsing) на данните. Въпреки това XML форматът се използва в доста уеб услуги (API) с множество разширения базирани на XML протокола, които се използват за различни цели. Такива са RDF, WSDL, SOAP, OWL.

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <document>
3    <company id="4093-НА">
4      <date_established>28.02.2001</date_established>
5      <name>ЕТ ДЖАМЕО-5-М</name>
6      <address>
7        <location>гр. Велики Преслав</location>
8        <street>ул. Александър Стамболийски 54</street>
9      </address>
10   </company>
11   <company id="3319-НА">
12     <date_established>04.07.2000</date_established>
13     <name>ЕТ ВЕСЕЛА ДЕЯНОВА</name>
14     <address>
15       <location>гр. Кюстендил</location>
16       <street>ул. Княз Александър Батенберг 30, вх. А</street>
17     </address>
18   </company>

```

### 3.5. WMS

WMS (Web Map Service) е стандартен протокол и контейнерен формат за обмен на гео-реферирани картови изображения (изгледи), генерирани от картов сървър на базата на данни от ГИС.

## 4. КАЧВАНЕ НА ДАННИ В ПОРТАЛА

### Предназначение на упътването

Упътването описва какво е нужно, за да се качват отворени данни в портала [opendata.government.bg](http://opendata.government.bg) и как да се направи това.

Главната целева група за това упътване са различните държавни администрации.

Целта на качването на данните в портала е те да бъдат лесно достъпни, лесни за намиране и позволяващи автоматизирана употреба.

### Използвани понятия

В този документ:

- „структура на информация“ е синоним на „структура на файл“ и „структура на съдържание“. Използва се като определение на: начин, по който дадена информация е подредена във файл, тоест: в колко колони е файлът (ако е в



табличен вид) и в какъв ред са (първа колона - наименование, втора колона - сума, т.н.); ако има числа в колоните, в каква мерна единица са те и какво точно представляват стойностите изреждани в колоните. Ако файлът не е в табличен вид (каквито са **XML** и **JSON** форматите), тогава същото описание може да бъде направено, но вместо колони да се отнася до полетата във файла. Аналогично е за други формати.

- „набор от данни“ е синоним на „масиви от информация“, „масиви от данни“ и „набор“. Наборът от данни представлява колекция от еднороден тип информация - файлове в отворен формат с едно и също предназначение и структура. „Структура“ се използва според горното определение.
- „етикети“ е синоним на „ключови думи“ и „теми“. На портала [opendata.government.bg](http://opendata.government.bg) наборите от данни са ограничени до еднороден тип информация, но могат да бъдат асоциирани с различни теми. Темите, за които се отнасят данните в един набор, се задават като „етикети“.
- **API (АПИ)** е начин една информационна система да позволи на друга дадени действия със съхраняваната информация: да поиска информация, да добавя нова информация или да изтрива такава. Като всяко от тези действия може да бъде ограничено до дадени потребители.
- терминът „версия“ се отнася до наборите от данни в портала [opendata.government.bg](http://opendata.government.bg). Версията на един набор от данни е числов идентификатор (пример: версия 1.0, 2.0, 2.1). Може да съществуват два набора с еднакво име и еднакви „етикети“, но с разлика във версиите. И двата набора могат да съдържат еднаква по предназначение информация. Разликата във версиите индикира, че информацията в единия набор е с различна структура от тази в другия.
- „лиценз“ се използва като синоним за „стандартни условия за ползване“.

## Преди да качвате данни в [opendata.government.bg](http://opendata.government.bg)

е необходимо да имате:

1. Регистриран акаунт в сайта.
1. Да създадете своя организация чрез акаунта си.

### Как да получа акаунт в сайта?

Свържете се с администраторите на портала и поискайте да ви се издаде потребителско име и парола. Може да се свържете с администраторите на портала на:

- електронен пощенски адрес: [n.ivanova@government.bg](mailto:n.ivanova@government.bg)

- телефон: +359 2 940 2445, Нуша Иванова, дирекция „Модернизация на администрацията“, Министерски съвет“

След като имате акаунт влезте в портала, като въведете потребителското си име и парола на адрес:

<https://opendata.government.bg/user/login>

## Как да регистрирам организация?

Данните в портала са подредени в организации. Освен това, към качените данни се прилагат и [етикети](#) (виж използвани понятия). Тази подредба е важна, за да могат потребителите в портала да се ориентират и достигнат до данните бързо и лесно, така че да започнат да ги използват максимално бързо. Освен това, подредбата позволява различни администрации да работят в синхрон, без да пречат едни на други.

След като се сдобие с акаунт, отидете на [страница „Организации“](#) и натиснете червения бутон [„Добавяне на организация“](#). След това бивате заведени на страница за попълване на информация за организацията, където трябва да попълните следните 4 полета:

И / Създаване на организация

ляват

е се  
ане,  
уване на  
данни.  
гат да  
в гадена  
имост от  
здаване,  
икуване.

## Създаване на организация

Име:

\* URL адрес:

Описание:

You can use [Markdown formatting](#) here

Изображение:

\* Задължително поле

1) **Име на организацията.** [Задължително] За да бъде консистентен порталът, е препоръчително всички имена да са на кирилица;

2) **Интернет адрес на организацията в портала. [Желателно]** Това поле се попълва автоматично. Ако бихте искали по-добре разпознаваем адрес в портала от рода на „opendata.government.bg/organization/imeto-na-vashata-organizatsiya“, моля напишете го на латиница, с малки букви и тирета.

3) **Описание на организацията [Желателно]** Полето е желателно. Опитайте се да дадете релевантна за потребителите информация, например лице за контакт в случай че бъде намерена грешка в данните, лице за контакт в случай че бъдат поискани допълнително данни от същата администрация, или специфики в предоставяната информация, било то в подредбата/представянето или съдържанието;

4) **Изображение. [Желателно]** За да бъде консистентен порталът, е препоръчително организациите да изберат отличителни снимки, като входи, гербове, или фасада на някоя от сградите, в които се намират. Можете да качите снимка от вашия компютър или като въведете интернет адрес до снимката (URL).

След като натиснете бутона „Създаване на организация“, регистрирането на организация е завършено. Вие имате права да качвате данни в тази организация. Администраторите на платформата могат да асистират при нужда.

## Качване на данни

След като сте създали организация (или имате права да качвате данни в съществуваща организация) имате възможността да качвате масиви от данни. За целта влезте в страницата на вашата организация и натиснете бутона „Добавяне на набор от данни“. Ще бъдете насочени към следната страница:

1 данни / Нов набор от данни

Звляват ли?

данни

екция

(като

описание и

, достъпни

адрес.

ни са онова,

илте

търсят

1 Нов набор от данни

2 Добавяне на данни

**Заглавие:**

\* URL адрес: [opendata.government.bg/dataset/<dataset>](http://opendata.government.bg/dataset/<dataset>) Редакция

**Описание:**

You can use [Markdown formatting](#) here

**Етикети:**

**Лиценз:** Лицензът не е уточнен.

i Описание и допълнителна информация за лиценза можете да намерите на [opendefinition.org](http://opendefinition.org)

**Организация:** test

**Видимост:** Частен

**Източник:**

[Наборът от данни](#) (виж използвани понятия) представлява колекция от еднороден тип информация - файлове с едно и също предназначение и [структура](#) (виж използвани понятия).

Пример: „Резултати от европейски избори“ или „Списък с НЗОК аптеки“.

Грешка ще бъде, ако наборът от данни служи за тематично групиране на файлове с **различна [структура на съдържание](#)**, тоест с различни по брой или разместени колони (за пълна дефиниция на „структура“ виж секция [използвани понятия](#)). Тематичното свързване в [opendata.government.bg](http://opendata.government.bg) се осъществява чрез [етикети](#) (виж [използвания понятия](#))

В уеб формуляра на тази страница е нужно да попълните:

- **Заглавие** - Нека описанието на данните да е максимално подробно. Например, желателно е да посочите в заглавието и времевия интервал, в който качваните файлове в набора са актуални.
- **URL адрес** - Точно както при създаването на организациите
- **Описание** - Опитайте се да дадете релевантна за потребителите информация, за качваните данни

- **Етикети** - (виж използвани термини) Етикетите задават темите, за които се отнасят данните в набора. Изберете поне един етикет. След като започнете да пишете в полето ще Ви се покажат предложения за вече съществуващи етикети в портала. Ако има етикети тематично свързани с данните, които качвате, тогава изберете тях. Ако няма такива етикети, напишете ваш. Ако например данните са „Резултати от европейски избори“, удачни етикети ще бъдат „избори“, „европа“, „глас“, „вот“.
- **Лиценз** - (виж използвани термини) Това са **стандартните условия** за ползване. Изберете условия, които смятате за точни спрямо качваните данни. Препоръчва се „CreativeCommons CCZero“.
- **Организация** - Името на вашата организация или организацията, от чието име качвате данни
- **Видимост** - Винаги избирайте „публичен“, освен в редките случаи, когато искате само вие да можете да преглеждате качените данни.
- **Източник** - линк към данните на сайта на Вашата институция, в случай че данните вече са качени онлайн.
- **Версия** - виж секция използвани термини
- **Автор** - юридическо или физическо лице, което е създадо или събрало данните
- **Имейл на автора** - адрес на електронната поща на автора
- **Отговорник по поддръжка** - лицето, което отговаря за поддържането на данните в актуално състояние
- **Имейл на отговорник по поддръжка** - адрес на електронната поща на автора
- **Допълнително поле** - това са метаданните за качвания набор от данни. Може да бъде качен файл (текстови или друг), както и да бъде посочен линк към метаданни за файла с данните. Метаданните могат да съдържат информация за структурата и съдържанието на данните, мерни единици, методологични бележки и други.

В следващата стъпка от формуляра е нужно да качите първия файл във вашия набор от данни. След успешното създаване на набора от данни може да добавяте още файлове, както и да прилагате корекции върху качените. В съдържанието на вашите файлове (отделна колона или поле) трябва да посочите и конкретен **времеви интервал**, за който информацията е актуална - по-конкретен интервал от този на набора от данни. Препоръчва се времевият интервал да е посочен и в заглавието на файла, който се качва.

## Странични бележки

Вече описахте и качихте данни в платформата. От най-ценните качества на отворените данни е възможността за автоматичен достъп до тях от бизнесът и гражданите чрез **„АПИ“** (виж използвани термини).

Затова, след като качите данни, проверявайте дали данните ви са качени успешно. Успешно качените данни са тези, на чиято страница за преглед има зелен бутон с надпис „Data API“ изглеждащ по следния начин:



[9a5d-310491a2c457/down...](#)

Ако се случи така, че качите файл, и той не стане достъпен през API-то, тогава трябва да натиснете бутонът „DataStore“ и след това „Качване в DataStore хранилище“. Ако се покаже съобщение за грешка, най-вероятно файлът няма правилна структура на съдържание. Много често се случва да има проблем с имената на колоните.

Ще Ви се наложи първо да коригирате файла и след това да направите нов опит за качване в портала.