

Как и где искать открытые данные?

Ирина Радченко
кандидат технических наук, доцент
iradche@gmail.com
телеграм-канал @DataPlace

29 октября 2020

План

- Определения
- Порталы и агрегаторы данных
- Репозитории данных
- Платформы распространения данных
- Информационные ресурсы по поиску данных

Определения



Что такое Открытые данные?

Languages: de el en es fr he hr id is it ja ko lt lv my ne nl_BE pt_BR ro ru zh_CN
zh_TW

Это руководство посвящено **открытым** данным, но что же это в точности такое? В частности, что делает **открытые** данные **открытыми**, и о каких именно данных мы говорим.

Что значит открытость?

- **Доступность и читаемость:** данные должны быть доступны целиком не дороже разумной стоимости их воспроизведения; желательно через интернет. Формат данных должен быть удобным для чтения и изменения.
- **Повторное использование и распространение:** данные должны предоставляться на условиях, которые разрешают их повторное использование и распространение, в том числе – в комбинации с другими наборами данных.
- **Всеобщее участие:** каждый должен иметь возможность использовать и распространять данные. Не должно быть дискриминации областей применения, людей или групп. Например, ограничение «только для некоммерческого использования», которое запрещает «коммерческое» применение, или ограничение возможных областей применения (к примеру, только в образовании), недопустимы.

Если у вас возник вопрос, почему так важно чётко определить, что значит открытость и почему используется такое определение, есть простой ответ: **интероперабельность**.

Интероперабельность означает способность различных систем и организаций работать совместно (англ. inter-operate). В данном случае, речь идёт о способность к взаимодействию — или «перемешиванию» различных наборов данных.

Интероперабельность важна, поскольку позволяет различным компонентам работать вместе. Возможность выделения компонентов и их «сборки» в единое целое — ключевая для построения больших и сложных систем. Без интероперабельности это практически

— способность

Improve this page

Translate this page

Donate

Share

Improve this page

Translate this page

Donate

Share

Принципы FAIR

scientific data

View all Nature Research journals Search Q Login 

Explore our content ▾ Journal information ▾

nature > scientific data > comment > article

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons 

Scientific Data 3, Article number: 160018 (2016) | Cite this article

143k Accesses | 1752 Citations | 1592 Altmetric | Metrics

 An Addendum to this article was published on 19 March 2019

Download PDF 

Associated Content

Collection

Scientific data

Collection

Metadata quality

Sections References

Abstract

Comment

Additional Information

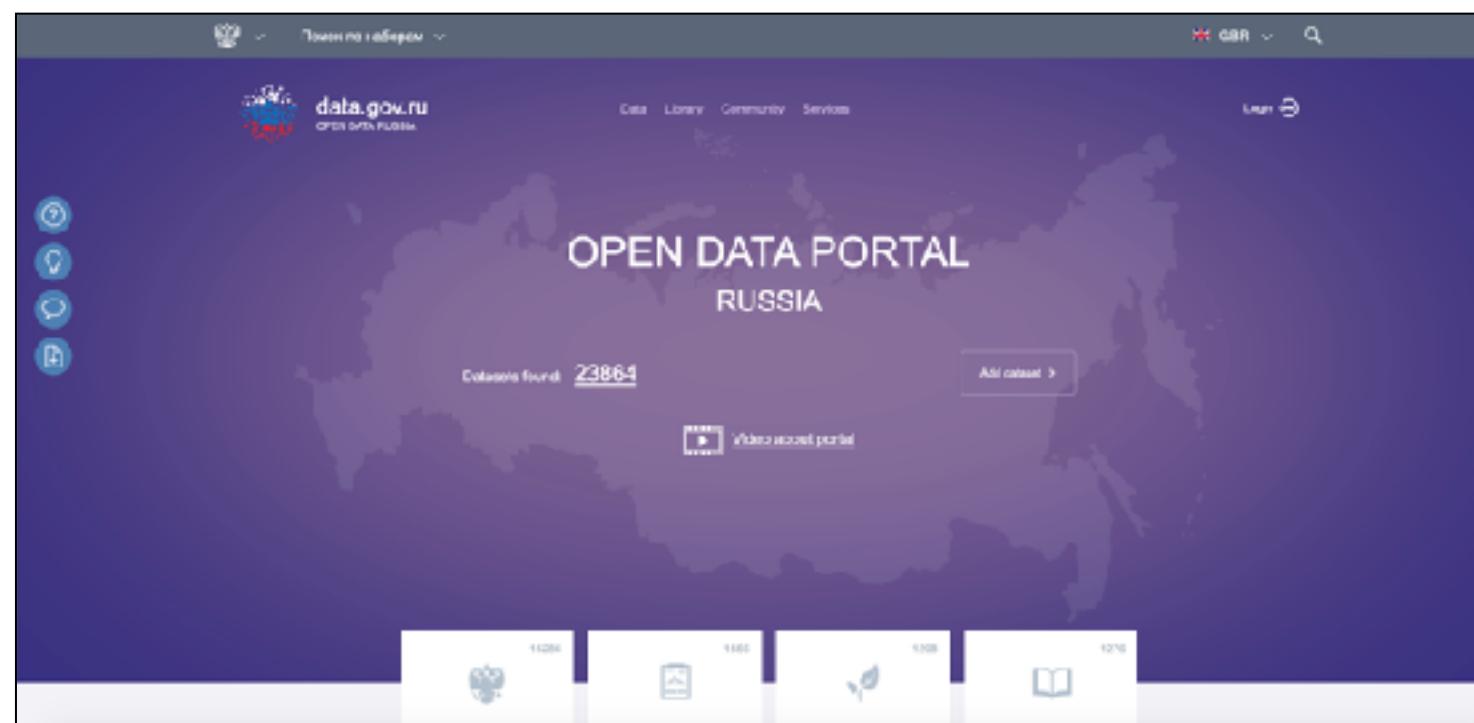
References

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and

<https://doi.org/10.1038/sdata.2016.18>

Порталы и агрегаторы данных

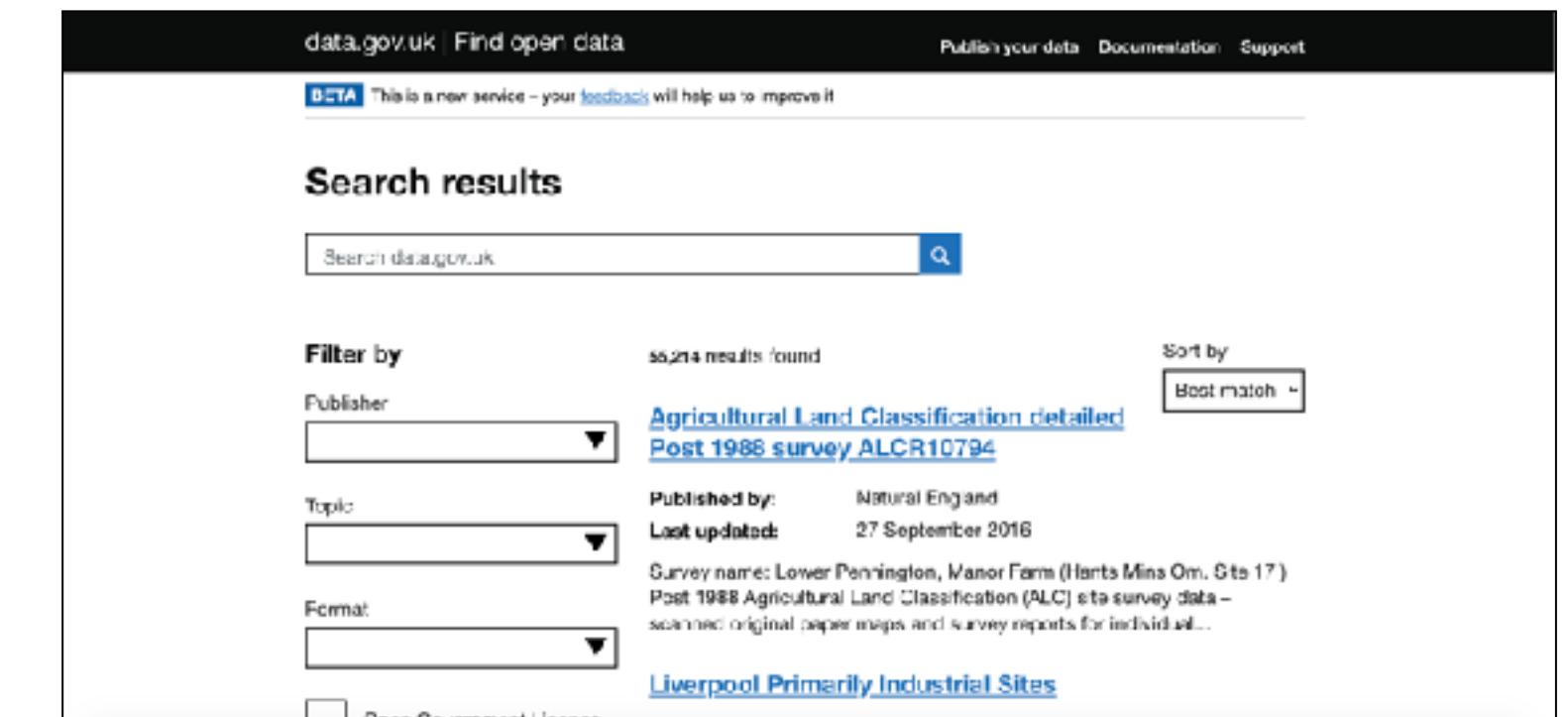
Порталы открытых данных



<https://data.gov.ru/>

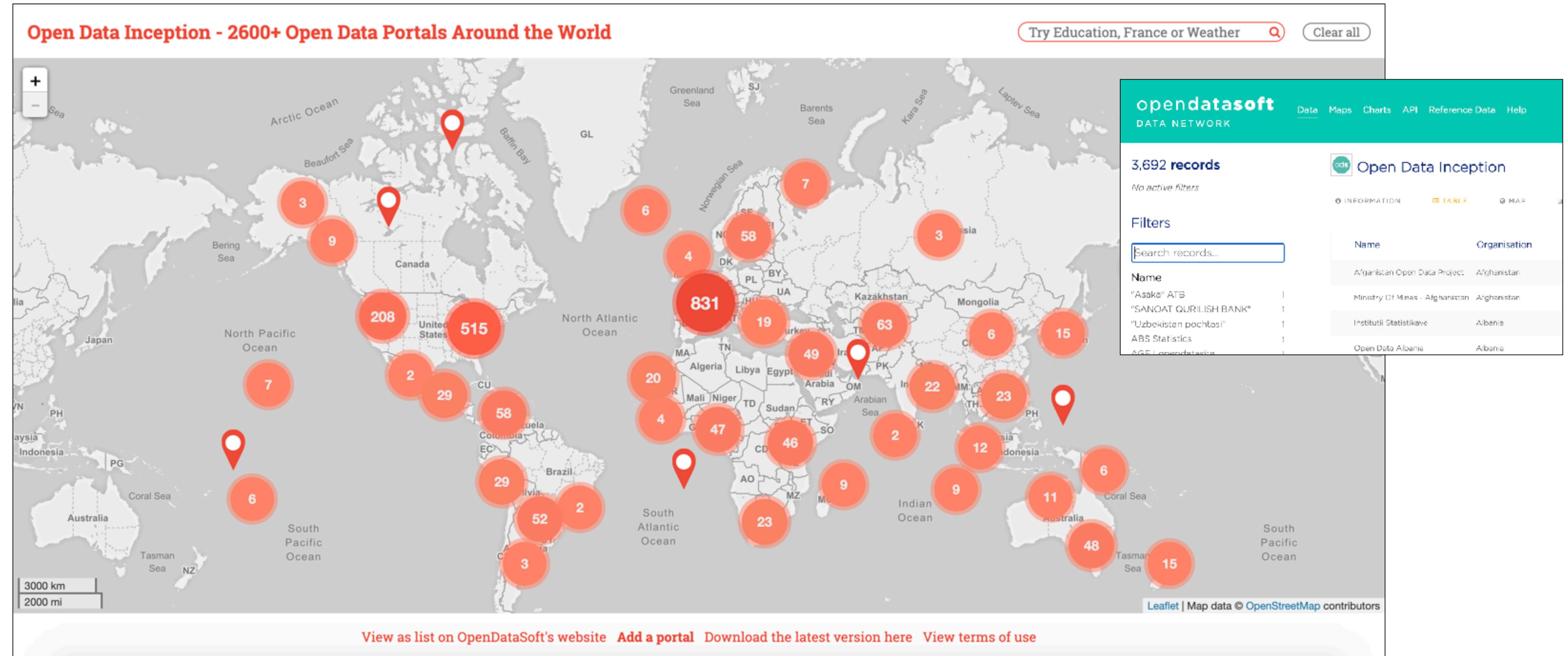


<https://www.data.gov/>



<https://data.gov.uk/>

Порталы открытых данных по всему миру



<https://opendatainception.io/>

Портал открытых данных World Bank

The screenshot shows the World Bank Open Data homepage. At the top left is the World Bank logo and a 'Data' link. At the top right are language links for English, Español, Français, العربية, and 中文. Below the header, there's a 'Start Here' link for new users and a navigation bar with icons for Home, DataBank, Microdata, Data Catalog, and a menu icon.

The main title 'World Bank Open Data' is centered above the subtitle 'Free and open access to global development data'. A search bar below the subtitle contains the placeholder 'Search data e.g. GDP, population, Indonesia'. To the right of the search bar is a 'Browse by Country or Indicator' link.

On the left, a 'MOST RECENT' section lists three articles:

- 'Online learning for Open Data in English and Spanish' by Tim Herzog and Kenneth Moreno, posted on Sep 16, 2020.
- 'Tracking the socioeconomic impacts of the pandemic in Nigeria: Results from the first three rounds of the Nigeria COVID-19 National Longitudinal Phone Survey' by Gbemisola Oseni, Amparo Palacios-Lopez, Kevin McGee, and Akuffo Amankwah, posted on Sep 16, 2020.
- 'Exploring links between democracy and...' (partially visible).

In the center, a 'WHAT YOU CAN LEARN WITH OPEN DATA' section features a large image of a scatter plot with a regression line, labeled 'Extreme Poverty' and 'The proportion of the world's population living in extreme poverty'.

On the right, a prominent graphic titled 'Understanding COVID-19 through data' displays a line graph showing the exponential growth of COVID-19 cases over time.

<https://data.worldbank.org/>

Портал открытых данных OECD

OECD.org Data Publications More sites ▾ News Job vacancies

OECD Data > Français

Find, compare and share the latest OECD data: charts, maps, tables and related publications ...

Search

OECD Data

Featured Charts

The latest OECD Interim Economic Outlook projects **global GDP to fall by 4.5%** this year, before picking up by 5% in 2021. See [more](#).

Quarterly GDP Total, Percentage change, previous period, Q2 2020 or latest available

Quarter	Percentage Change
Q1 2020	-25.2
Q2 2020	-20.4
Q3 2020	-17.1
Q4 2020	-16.4
Q1 2021	-13.8
Q2 2021	-12.8
Q3 2021	-11.5
Q4 2021	-11.4
Q1 2022	-11.0
Q2 2022	-10.6
Q3 2022	-9.7
Q4 2022	-9.7
Q1 2023	-9.1
Q2 2023	-7.9
Q3 2023	-7.0
Q4 2023	-6.9
Q1 2024	-6.9
Q2 2024	-3.2
Q3 2024	-3.2
Q4 2024	11.5

Browse by [topic](#) or [country](#)

[Search tips](#)
[Catalogue of OECD databases](#)

Latest news

Statistical news releases
See recent statistical news releases.

Compare your income
Try OECD's [Compare your income](#) to check your perception of income inequality against reality!

Statistical resources

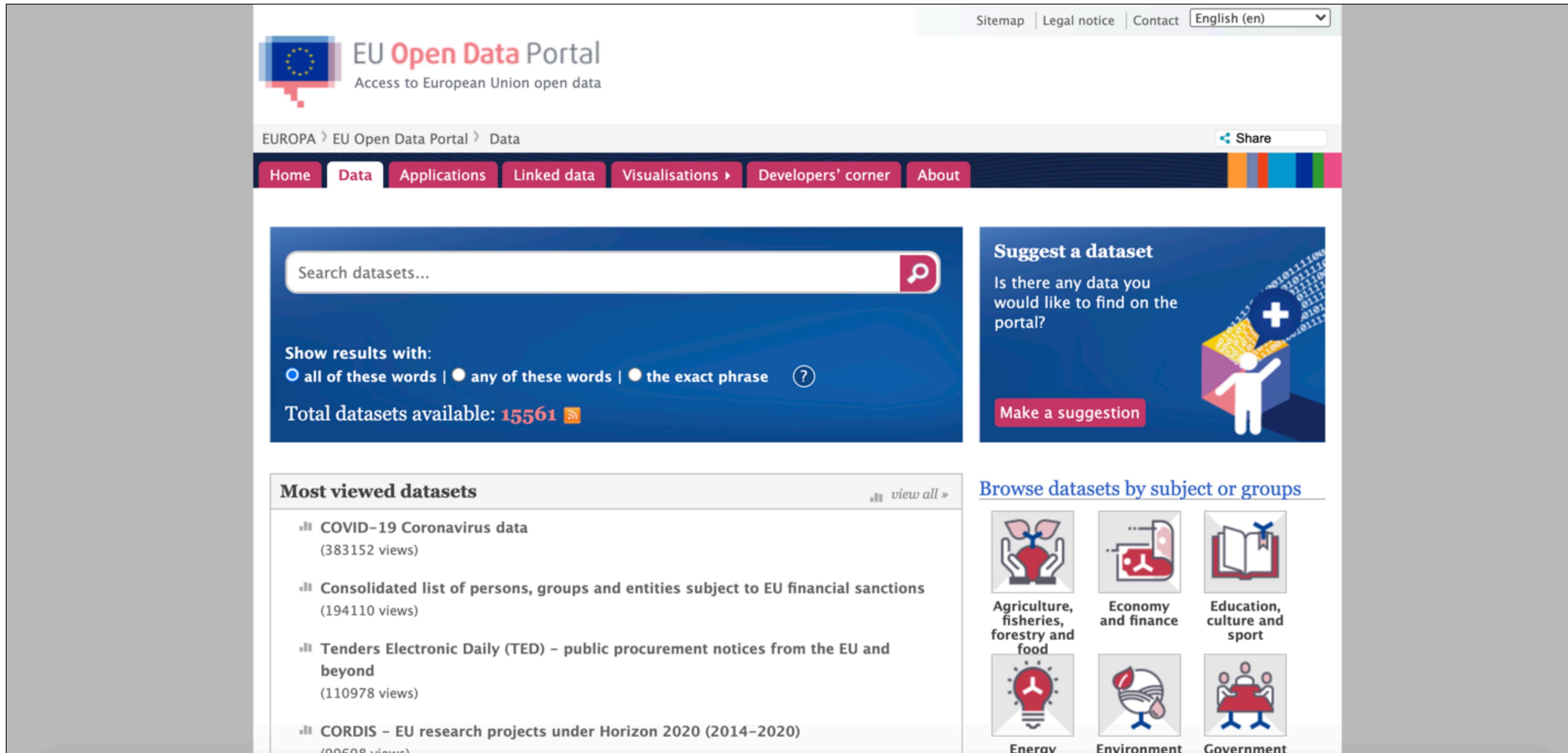
Database access

Households' financial and non-financial assets and liabilities - Annual and Quarterly

Country	United States
Assets	1
Liabilities	2

<https://data.oecd.org/>

EU open data portal



The screenshot shows the EU Open Data Portal homepage. At the top, there is a navigation bar with links to Sitemap, Legal notice, Contact, and a language dropdown set to English (en). Below the navigation bar is the EU Open Data Portal logo and the tagline "Access to European Union open data". The main search area features a search bar with placeholder text "Search datasets..." and a magnifying glass icon. Below the search bar are options for search operators: "all of these words" (radio button), "any of these words" (radio button), "the exact phrase" (radio button), and a help link (?). A message indicates "Total datasets available: 15561" with a feed icon. To the right, there is a "Suggest a dataset" section with a "Make a suggestion" button and an illustration of a person pointing at a stack of data cubes. Below the search area, there are two sections: "Most viewed datasets" and "Browse datasets by subject or groups". The "Most viewed datasets" section lists four datasets with their titles and view counts: COVID-19 Coronavirus data (383152 views), Consolidated list of persons, groups and entities subject to EU financial sanctions (194110 views), Tenders Electronic Daily (TED) – public procurement notices from the EU and beyond (110978 views), and CORDIS – EU research projects under Horizon 2020 (2014–2020) (99608 views). The "Browse datasets by subject or groups" section shows six categories with corresponding icons: Agriculture, fisheries, forestry and food (hands holding a plant), Economy and finance (a chart with a red arrow), Education, culture and sport (an open book), Energy (a lightbulb), Environment (a recycling symbol), and Government (two people at a desk).

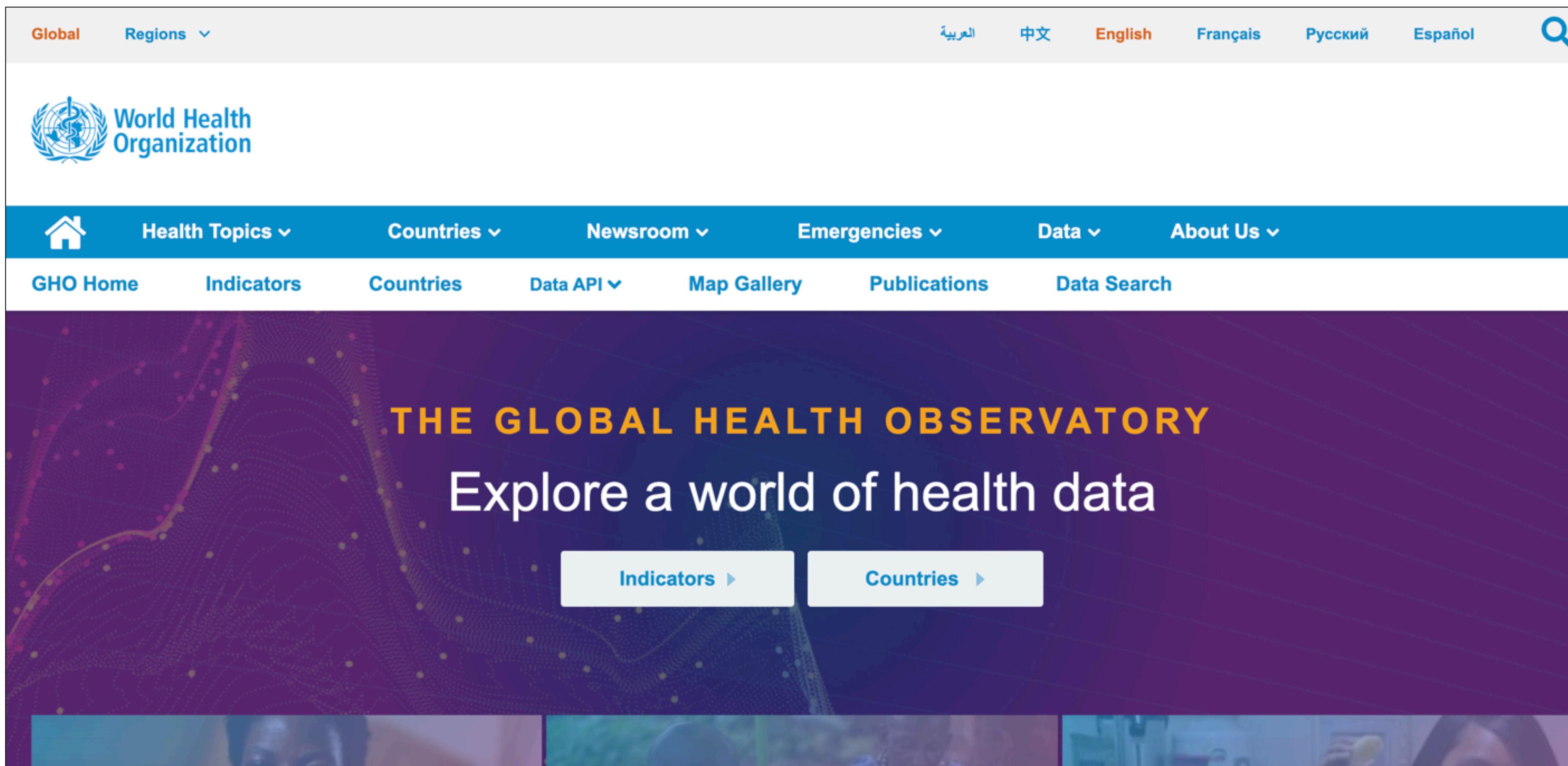
<https://data.europa.eu/euodp/en/data>

European Data portal: сборщик метаданных

The screenshot shows the European Data Portal's user interface. At the top, there is a navigation bar with links for Home, Data, Impact & Studies, Training, News & Events, and About. Below the navigation bar is a secondary menu with tabs for EU and international data, Country data (which is selected), SPARQL Search, Statistics, and Metadata Quality. On the left side, there is a map of Europe with a zoom control (+/-) and a search bar for filtering datasets by location. Next to the map is a search bar for ordering datasets by Last Modified. In the center, there is a search bar for "Search datasets..." and a message indicating "1115343 datasets found". To the right of the search bar, there are buttons for "Datasets Feed" and "Catalogues". Below this, a specific dataset is highlighted: "Journées Européennes du patrimoine 2020 (Liste)". The description for this dataset is: "Programme des Journées européennes du patrimoine du 18 au 20 septembre 2020 sur le territoire de Grand Chambéry. Les données sont extraites de la plateforme d'information touristique APIDAE.". The dataset has three download options: json, shp, and csv. It was created on 20.09.2020 at 06:00 and updated on the same day. At the bottom right, there is a logo for "Plateforme ouverte des données publiques françaises". On the far left, there is a sidebar with a "Settings" section containing "Operator" (set to AND), a toggle switch, and "OR".

<https://www.europeandataportal.eu/data/datasets?locale=en&minScoring=0>

Портал ВОЗ: Global Health Data



<https://www.who.int/data/gho>

Centers for Disease Control and Prevention

 Centers for Disease Control and Prevention
CDC 24/7: Saving Lives. Protecting People.TM Data.CDC.gov

Home Data Catalog Developers Video Guides [Sign In](#)

Search

Categories

- Administrative
- Biomonitoring
- Disability & Health
- Environmental Health & Toxicology
- Foodborne, Waterborne, and Related Diseases

Show All... View Types

- Calendars
- Charts

Featured Content

CDC COVID-19 Data Tracker

External Content



One-stop source for national and county-level cases, deaths, testing and other COVID-19 information.

919 Results

COVID-19 Case Surveillance Public Use Data

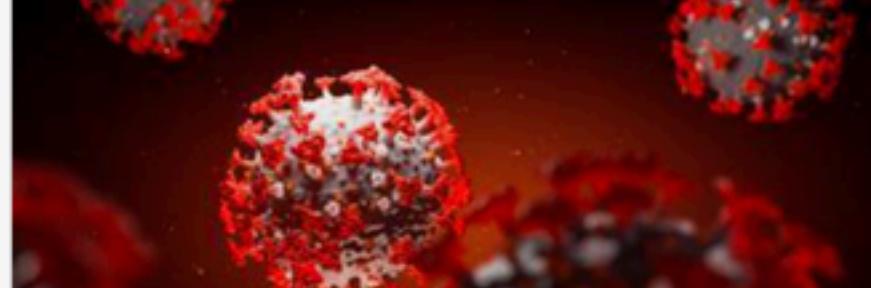
September 1, 2020 208K Views



The COVID-19 case surveillance system database includes patient-level data reported to U.S. states and autonomous reporting entities,...

COVIDView

External Content



A weekly surveillance summary of U.S. COVID-19 Activity.

Sort by Most Relevant

<https://data.cdc.gov/browse>

Федеральная служба государственной статистики

Федеральная служба
государственной статистики

f t v o Введите свой запрос Eng

О Росстате Статистика Публикации Респондентам Пресс-служба

Главная страница / Статистика В ИЗБРАННОЕ

Статистика

- Официальная статистика
- Переписи и обследования
- Методология и нормативно-справочная информация
- Интерактивные статистические сервисы

Анонсы

- Новости статистики
- Инфографика

Поделиться в соцсетях f t v o

<https://rosstat.gov.ru/statistic>

ЕМИСС государственная статистика

The screenshot displays the homepage of the EMISSTAT website. At the top, there is a dark header bar with the logo of the Federal Statistical Service of Russia (ФСТат) and the text "Е М И С С ГОСУДАРСТВЕННАЯ СТАТИСТИКА". Below the header are navigation links for "Показатели" (Indicators), "Ведомства" (Ministries), and a "Войти" (Log in) button. The main content area features a large title "Официальные статистические показатели" (Official statistical indicators) and a search bar with a placeholder "Поиск...". To the right, two large numerical boxes show "7166 показателей" (7166 indicators) and "65 ведомств" (65 ministries). Below this, on the left, is a box showing "Посетителей в день" (Visitors per day) with the number "1190" and a bar chart. On the right, under the heading "Популярные показатели" (Popular indicators), are four items: "Размер и состав денежных доходов и расходов населения" (87), "Экспорт отдельных товаров" (5), "Индексы потребительских цен на товары и услуги" (4), and "Средние потребительские цены (тарифы) на товары и услуги" (4). A dropdown menu "За неделю" is visible next to the popular indicators section.

Показатели

Ведомства

Войти

Официальные статистические показатели

Поиск...

Расширенный поиск

7166 показателей

65 ведомств

Посетителей в день

1190

Посетителей в неделю

15823

Популярные показатели

За неделю ▾

- 87 Размер и состав денежных доходов и расходов населения
- 5 Экспорт отдельных товаров
- 4 Индексы потребительских цен на товары и услуги
- 4 Средние потребительские цены (тарифы) на товары и услуги

<https://www.fedstat.ru/>

Проект ГосЗатрат

Онлайн справочник вопросов и ответов по коронавирусу COVID-19 [ХОЧУ НАЙТИ ОТВЕТЫ >](#)

[RSS](#) [f](#) [vk](#) [Twitter](#) [g+](#) [in](#) [Русский](#) [English](#)

ClearSpending

[Home](#) All news Субсидии Contracts Budgets Customers Suppliers Purchases Аномалии Выгрузка данных NEW Forum About the project

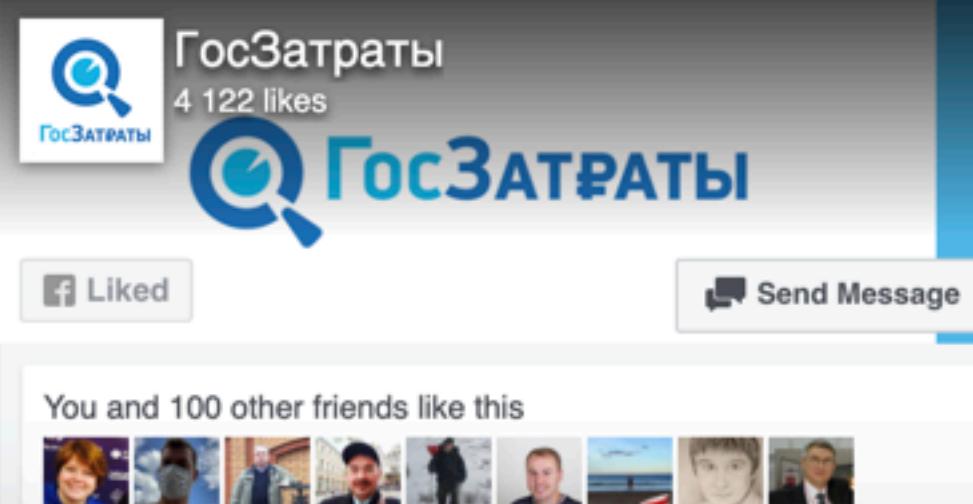
Contract search

Contract subject [Search](#)

Latest news



Мы в facebook



<https://clearspending.ru/>

Платформа DrivenData

The screenshot shows the homepage of the DrivenData platform. At the top, there is a navigation bar with links for COMPETITIONS, ABOUT, CAREERS, DRIVENDATA^{LABS}, BLOG, MY PROFILE, and LOG OUT. The logo 'DRIVENDATA' is on the left, featuring a colorful bar chart icon. Below the navigation, a large text area reads 'Data science competitions to build a better world'. Two buttons at the bottom left say 'I want to join a competition →' and 'I want to run a competition →'. To the right of the text is a large, colorful sunburst chart with many radial segments in various colors (green, yellow, orange, red, blue, purple) radiating from a central white circle.

<https://www.drivendata.org/>

Платформа Kaggle

The screenshot shows the Kaggle website interface. On the left is a sidebar with navigation links: Home, Compete, Data (which is selected), Notebooks, Discuss, Courses, Jobs, and More. The main content area has a search bar at the top. Below it, the title "Datasets" is displayed, followed by the sub-instruction "Find and use datasets or complete tasks. [Learn more.](#)". A prominent button labeled "+ New Dataset" is visible. In the center, there's a section titled "Create Public Datasets" with the sub-instruction "Open a dialogue, accept contributions, and get insights: improve your dataset by publishing it on Kaggle." Below this is a "Create Public Dataset" button. To the right of this central section is a cartoon illustration of two people working on a laptop. At the bottom of the main content area, there's a search bar for "Search 53,863 datasets", a "Feedback" link, and a "Filter" button. Below the search bar, there are tabs for "Public", "Your Datasets", and "Favorites", with "Public" being the active tab. The results list includes a dataset titled "Health Insurance Cross Sell Prediction" by Anmol Kumar, which was posted 9 days ago, is 6 MB in size, contains 10.0 files (CSV), and has 3 Files (CSV). There are 109 submissions for this dataset. To the right, there's a sidebar titled "Open Tasks" featuring a competition titled "Health Insurance Cross Sell Pre..." with 14 submissions. Below that is a "Competition" section with 17 submissions.

<https://www.kaggle.com/datasets>

Репозитории данных

Time Series Classification

Time Series Classification Home Datasets Algorithms Results Researchers Code Bibliography UEA Papers ▾ About Us

Train Size

[Less than 100 \(49\)](#)
[100 to 500 \(85\)](#)
[Greater than 500 \(44\)](#)

Test Size

[Less than 300 \(84\)](#)
[300 to 1000 \(49\)](#)
[Greater than 1000 \(45\)](#)

Length

[Less than 300 \(91\)](#)
[300 to 700 \(40\)](#)
[Greater than 700 \(47\)](#)

Classes

[Less than 10 \(134\)](#)
[10 to 30 \(34\)](#)
[Greater than 30 \(10\)](#)

Type

[Device \(10\)](#)
[ECG \(9\)](#)
[Image \(34\)](#)
[Motion \(25\)](#)

Dataset listing

The univariate and multivariate classification problems are available in three formats: Weka ARFF, simple text files and sktime ts format. Weka does not allow for unequal length series, so the unequal length problems are all padded with missing values. ts format does allow for this feature.

[Univariate Weka formatted ARFF files and .txt files \(about 500 MB\).](#)

[Univariate sktime formatted ts files \(about 300 MB\).](#)

[Multivariate Weka formatted ARFF files \(and .txt files\) \(about 2 GB\).](#)

[sktime formatted ts files \(about 1.5 GB\).](#)

Lists of the data, including which are unequal length, can be found [here](#). Details on loading sktime data with the Python package are [here](#).

To store multivariate series in ARFF we take advantage of relational attributes. These are fairly unintuitive, so we have provided an overview of this and other basic features of loading data and building classifiers [here](#).

These files provide a simple list of the data characteristics for [univariate](#) and [multivariate](#) problems.

To see how accurate different classifiers are on these data see the [results page](#).

More information on the datasets is given below. Problems with variable length series are listed as length 0.

Search

Dataset	Train Size	Test Size	Length	No. of Classes	Type
AbnormalHeartbeat	303	303	3053	5	AUDIO
ACSF1	100	100	1460	10	DEVICE
Adiac	390	391	176	37	IMAGE

<http://timeseriesclassification.com/dataset.php>

NLP Database

The Big Bad NLP Database

For database updates follow on [Twitter](#) or [Medium](#)

Want to add a dataset, edit? [Edit](#)

updated 09.19.20 584 datasets

search...

Dataset	Added	Lang	Description	Inst	Format	Task	Year	Creator	Source
Visual Genome	09.19.20	English	Dataset contains over 100K images where each image has an average of 21 objects, 18 attributes, and 18 pairwise relationships between objects.	100,000+	JSON	Visual Question Answering, Knowledge Base	2016	Krishna et al.	LINK PAPER
CraigsListBargain	09.19.20	English	Dataset contains 6,682 human-human dialogues where 2 agents negotiate the sale/purchase of an item.	6,682	JSON	Dialogue	2018	He et al.	LINK PAPER
A Multi-Turn, Multi-Domain Dialogue Dataset (KVRET)	09.19.20	English	Dataset contains 3,031 multi-turn dialogues in three distinct domains appropriate for an in-car assistant: calendar scheduling, weather information retrieval, and point-of-interest navigation.	3,301	JSON	Dialogue	2017	Eric et al.	LINK PAPER
CMU_ARCTIC	09.19.20	English	Dataset contains 1,150 utterances carefully selected from out-of-copyright texts from Project Gutenberg. The databases include US English male (bdl) and female (slt) speakers (both experienced voice talent) as well as other accented speakers.	1,150	WAV	Speech Recognition	2004	CMU	LINK PAPER
			Dataset includes recordings from twenty-four (24) non-native speakers of						

<https://datasets.quantumstat.com/>

Machine Learning Repository

UCI 

Machine Learning Repository
Center for Machine Learning and Intelligent Systems

About Citation Policy Donate a Data Set Contact
 Search
 Repository Web Google

[View ALL Data Set](#)

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 557 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. For a general overview of the Repository, please visit our [About page](#). For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to [contact the Repository librarians](#).

Supported By:  In Collaboration With: 

Latest News:	Newest Data Sets:	Most Popular Data Sets (hits since 2007):
<p>09-24-2018: Welcome to the new Repository admins Dheeru Dua and Efi Karra Taniskidou!</p> <p>04-04-2013: Welcome to the new Repository admins Kevin Bache and Moshe Lichman!</p> <p>03-01-2010: Note from donor regarding Netflix data</p> <p>10-16-2009: Two new data sets have been added.</p> <p>09-14-2009: Several data sets have been added.</p> <p>03-24-2008: New data sets have been added!</p> <p>06-25-2007: Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope</p>	<p>07-22-2020:  Facebook Large Page-Page Network</p> <p>07-17-2020:  Amphibians</p> <p>07-12-2020:  Early stage diabetes risk prediction dataset.</p> <p>06-28-2020:  Taiwanese Bankruptcy Prediction</p> <p>06-20-2020:  South German Credit (UPDATE)</p> <p>06-17-2020:  BitcoinHeistRansomwareAddressDataset</p>	<p>3533819:  Iris</p> <p>1923463:  Adult</p> <p>1484282:  Wine</p> <p>1326060:  Breast Cancer Wisconsin (Diagnostic)</p> <p>1309344:  Heart Disease</p> <p>1303973:  Wine Quality</p>

Latest News:
09-24-2018: Welcome to the new Repository admins Dheeru Dua and Efi Karra Taniskidou!
04-04-2013: Welcome to the new Repository admins Kevin Bache and Moshe Lichman!
03-01-2010: Note from donor regarding Netflix data
10-16-2009: Two new data sets have been added.
09-14-2009: Several data sets have been added.
03-24-2008: New data sets have been added!
06-25-2007: Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

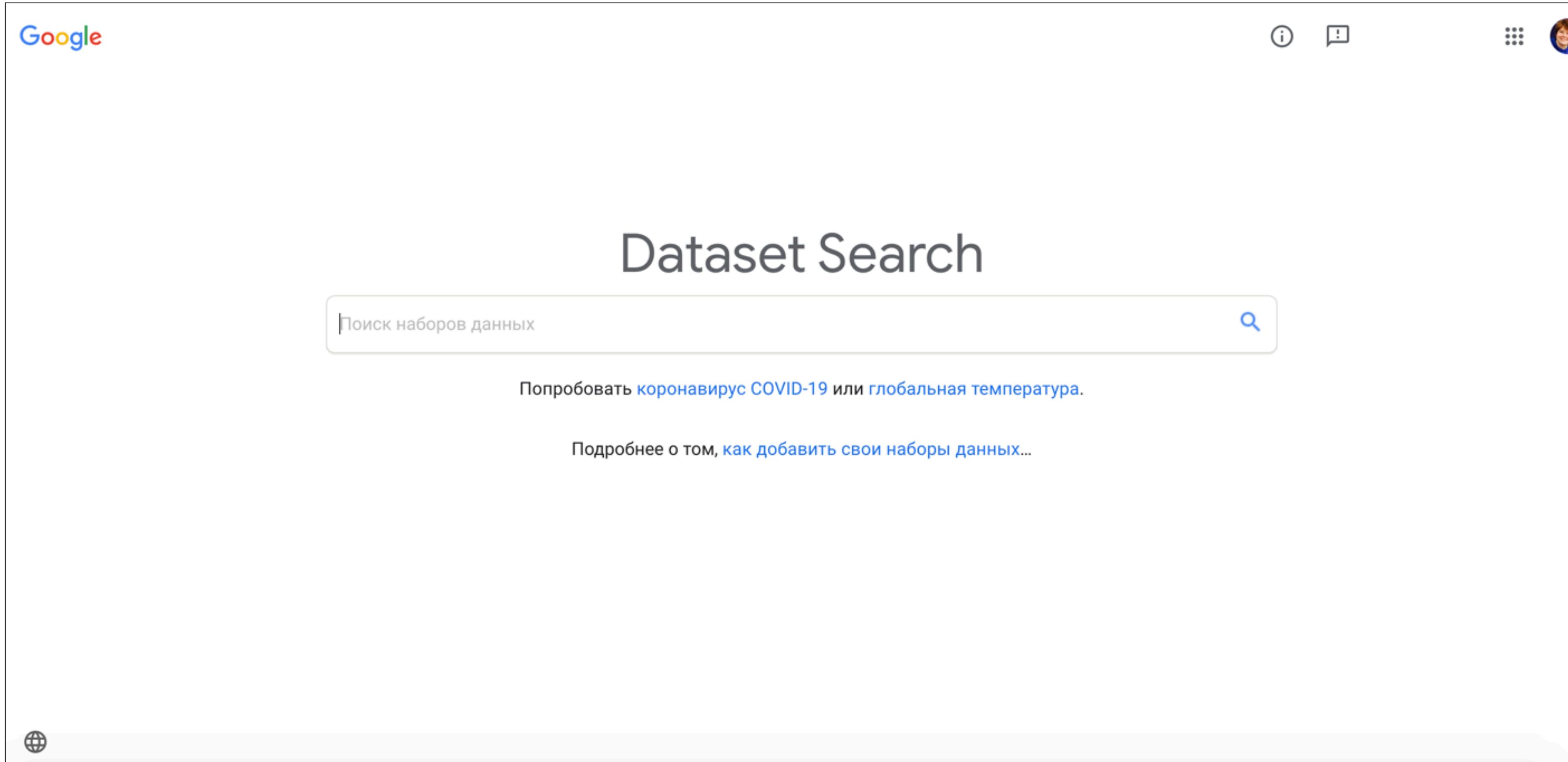
Featured Data Set: [ICU](#)
 Data Type: Multivariate, Time-Series

Newest Data Sets:
07-22-2020:  [Facebook Large Page-Page Network](#)
07-17-2020:  [Amphibians](#)
07-12-2020:  [Early stage diabetes risk prediction dataset.](#)
06-28-2020:  [Taiwanese Bankruptcy Prediction](#)
06-20-2020:  [South German Credit \(UPDATE\)](#)
06-17-2020:  [BitcoinHeistRansomwareAddressDataset](#)

Most Popular Data Sets (hits since 2007):
3533819:  [Iris](#)
1923463:  [Adult](#)
1484282:  [Wine](#)
1326060:  [Breast Cancer Wisconsin \(Diagnostic\)](#)
1309344:  [Heart Disease](#)
1303973:  [Wine Quality](#)

<https://archive.ics.uci.edu/ml/index.php>

Google: Проект Dataset Search



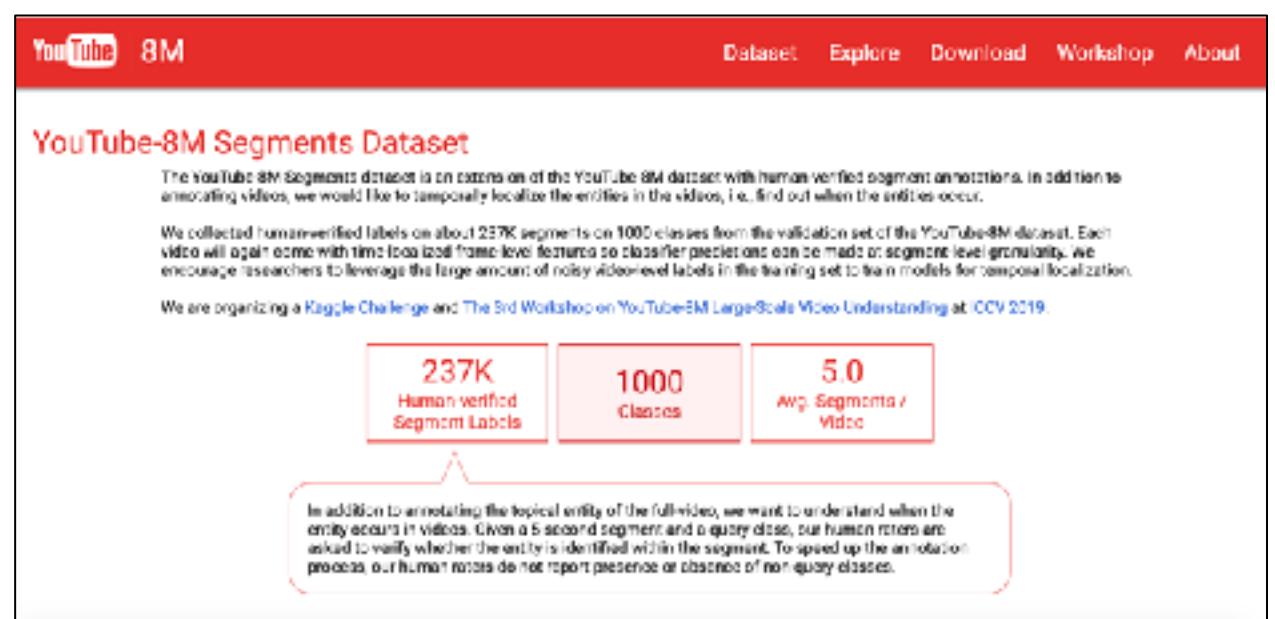
<https://datasetsearch.research.google.com/>

Google: BigQuery public datasets

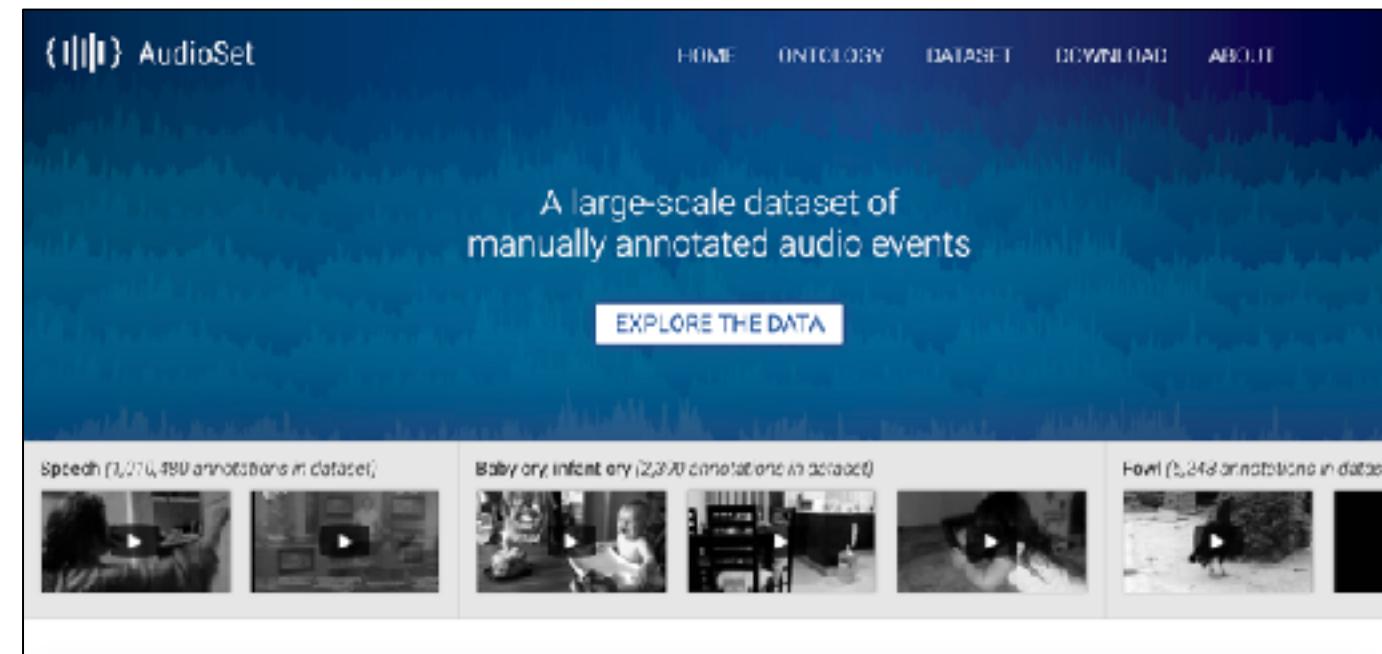
The screenshot shows the Google Cloud BigQuery public datasets documentation page. The page has a navigation bar at the top with links for Google Cloud, Why Google, Solutions, Products, Pricing, Getting Started, a search bar, Docs, Support, Language (set to English), Console, and a user profile icon. Below the navigation bar, there's a secondary navigation menu with links for Data analytics products, Guides, Reference, Support, Resources (which is underlined), Contact Sales, and Get started for free. On the left, there's a sidebar with a tree view of BigQuery resources: All resources, Pricing, Quotas and limits, Release notes, Public datasets (which is selected and highlighted in blue), Solution providers, and Service level agreement. The main content area has a breadcrumb trail: Data analytics products > BigQuery > Documentation > Resources. It features a title 'BigQuery public datasets' with a star rating of 5 stars and a 'Send feedback' button. A paragraph explains that the Cloud Public Datasets Program catalog is in Google Cloud Marketplace. It includes a 'Go to Datasets in the Cloud Marketplace' button. Another paragraph defines a public dataset as stored in BigQuery and made available through the Google Cloud Public Dataset Program, mentioning storage costs and query pricing details. The right sidebar contains a 'Contents' section with links to Before you begin, Public dataset locations, Accessing public datasets in the BigQuery web UI, Other public datasets, Sharing a dataset with the public, Sample tables, Contact us, and What's next.

<https://cloud.google.com/bigquery/public-data/>

Google: YouTube-8M Segments Dataset

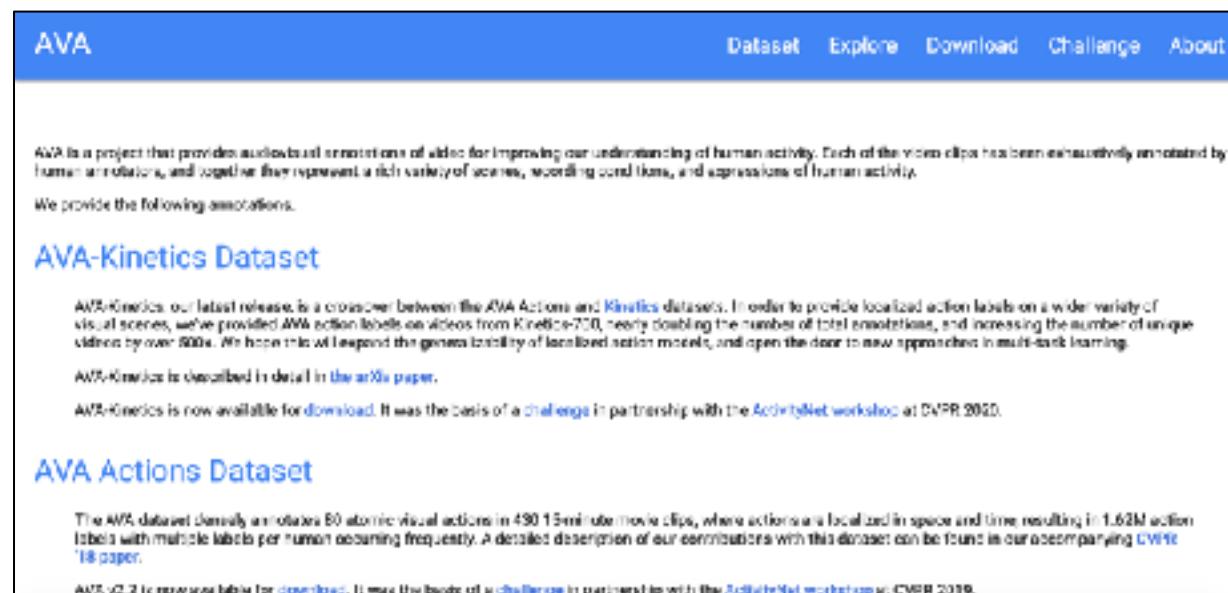


Google: AudioSet



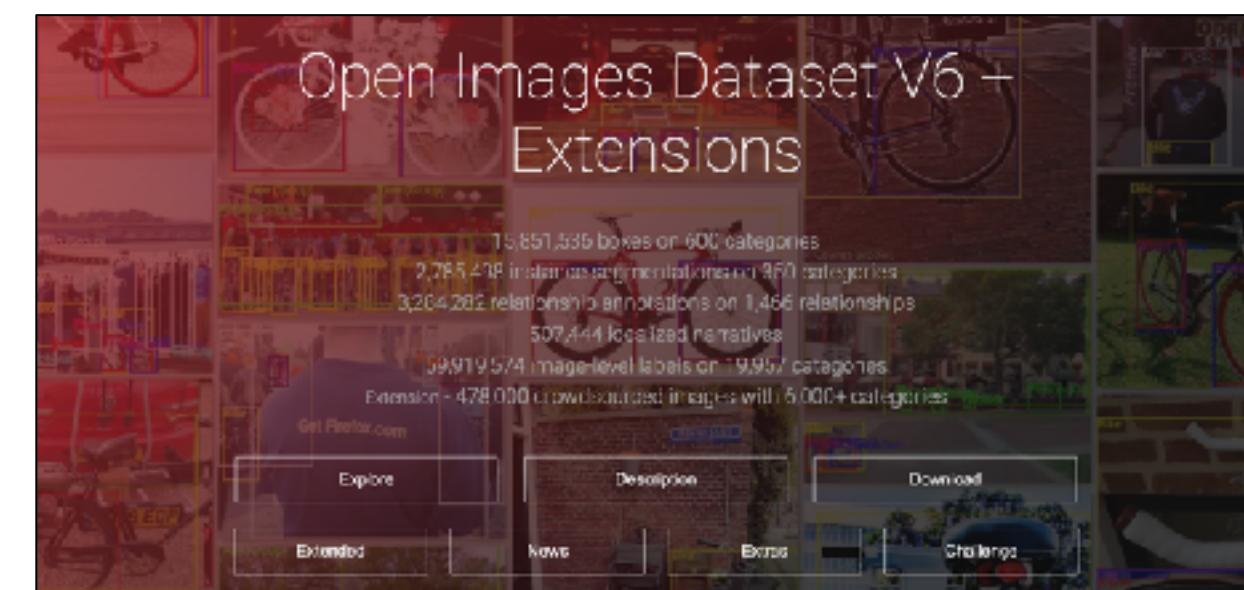
<https://research.google.com/youtube8m/>

Google: AVA, audiovisual annotations of video



<https://research.google.com/ava/>

Google: Open Image



<https://storage.googleapis.com/openimages/web/index.html>

Open Data on Amazon

Registry of Open Data on AWS

About

This registry exists to help people discover and share datasets that are available via AWS resources. [Learn more about sharing data on AWS.](#)

See all usage examples for datasets listed in this registry.

See datasets from [Facebook Data for Good](#), [NASA Space Act Agreement](#), [NIH STRIDES](#), [NOAA Big Data Project](#), [Space Telescope Science Institute](#), and [Amazon Sustainability Data Initiative](#).

Search datasets (currently 184 matching datasets)

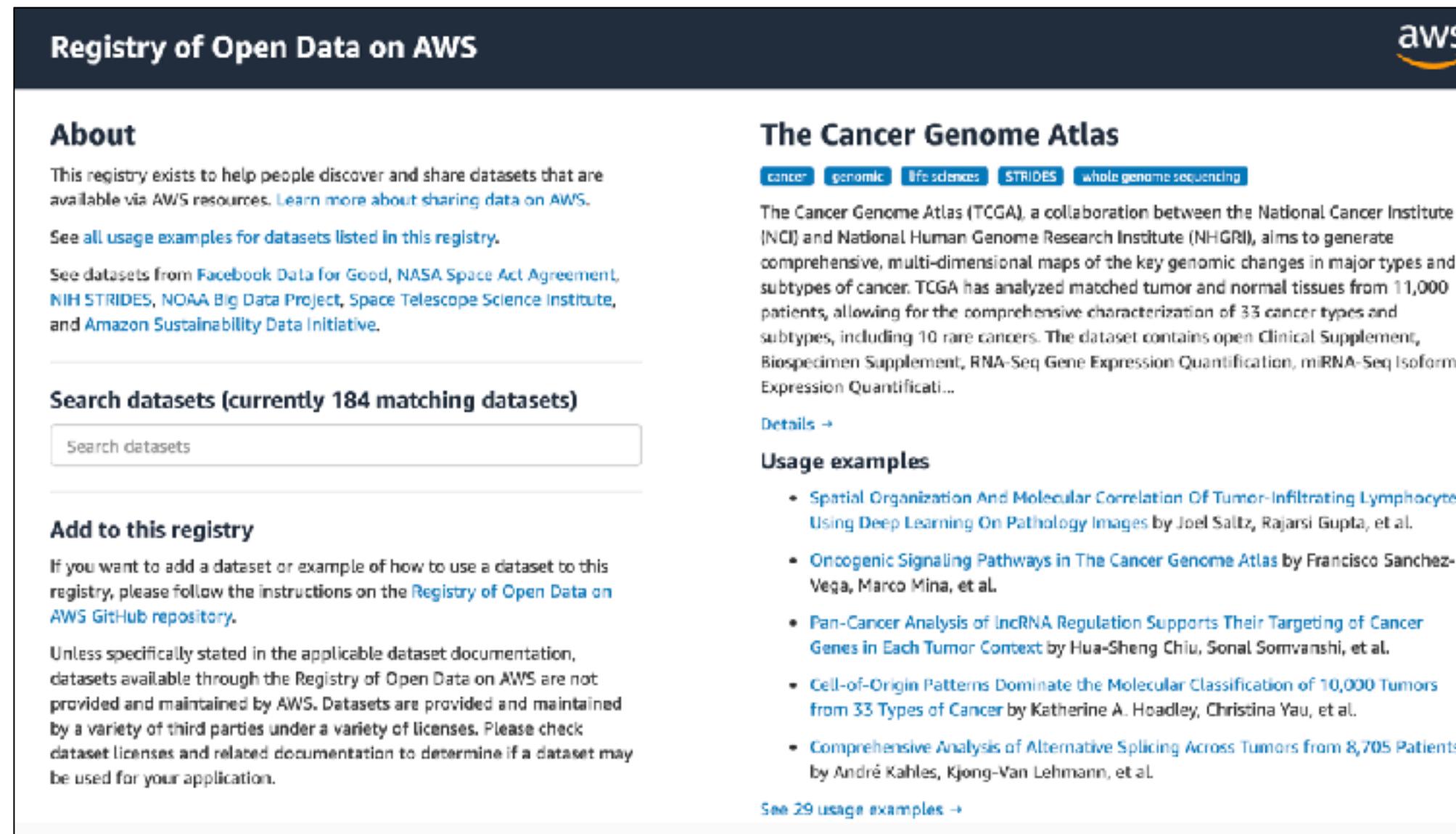
Search datasets

Add to this registry

If you want to add a dataset or example of how to use a dataset to this registry, please follow the instructions on the [Registry of Open Data on AWS GitHub repository](#).

Unless specifically stated in the applicable dataset documentation, datasets available through the Registry of Open Data on AWS are not provided and maintained by AWS. Datasets are provided and maintained by a variety of third parties under a variety of licenses. Please check dataset licenses and related documentation to determine if a dataset may be used for your application.

[See 29 usage examples →](#)



<https://registry.opendata.aws/>

awslabs / open-data-registry

Code Issues 4 Pull requests 11 Actions Security Insights

master 5 branches 0 tags Go to file Add file Code

5 authors Add DANDI dataset to registry (#387) af935b4 3 days ago 1,153 commits

_scripts Add created & last modified dates to yaml (#488) 8 months ago

datasets Add DANDI dataset to registry (#387) 3 days ago

.gitignore fix(isid): Update ISD resources for CSV files 10 months ago

CONTRIBUTING.md Initial commit 3 years ago

LICENSE Initial commit 3 years ago

README.md Fix README link 2 months ago

buildspec.yml Add created & last modified dates to yaml (#488) 8 months ago

ext.py Add services.yaml (#489) 8 months ago

requirements.txt Add check on bucket existance & region (#462) 6 months ago

resources.yaml Update resources.yaml 2 years ago

schemam.yaml Updating URL schema to account for longer TLDs 2 months ago

https://github.com/awslabs/open-data-registry/commit/1aae348271b576014fead623cc30ca53c23a724e 9 days ago

About
A registry of publicly available datasets on AWS

[registry.opendata.aws](#)

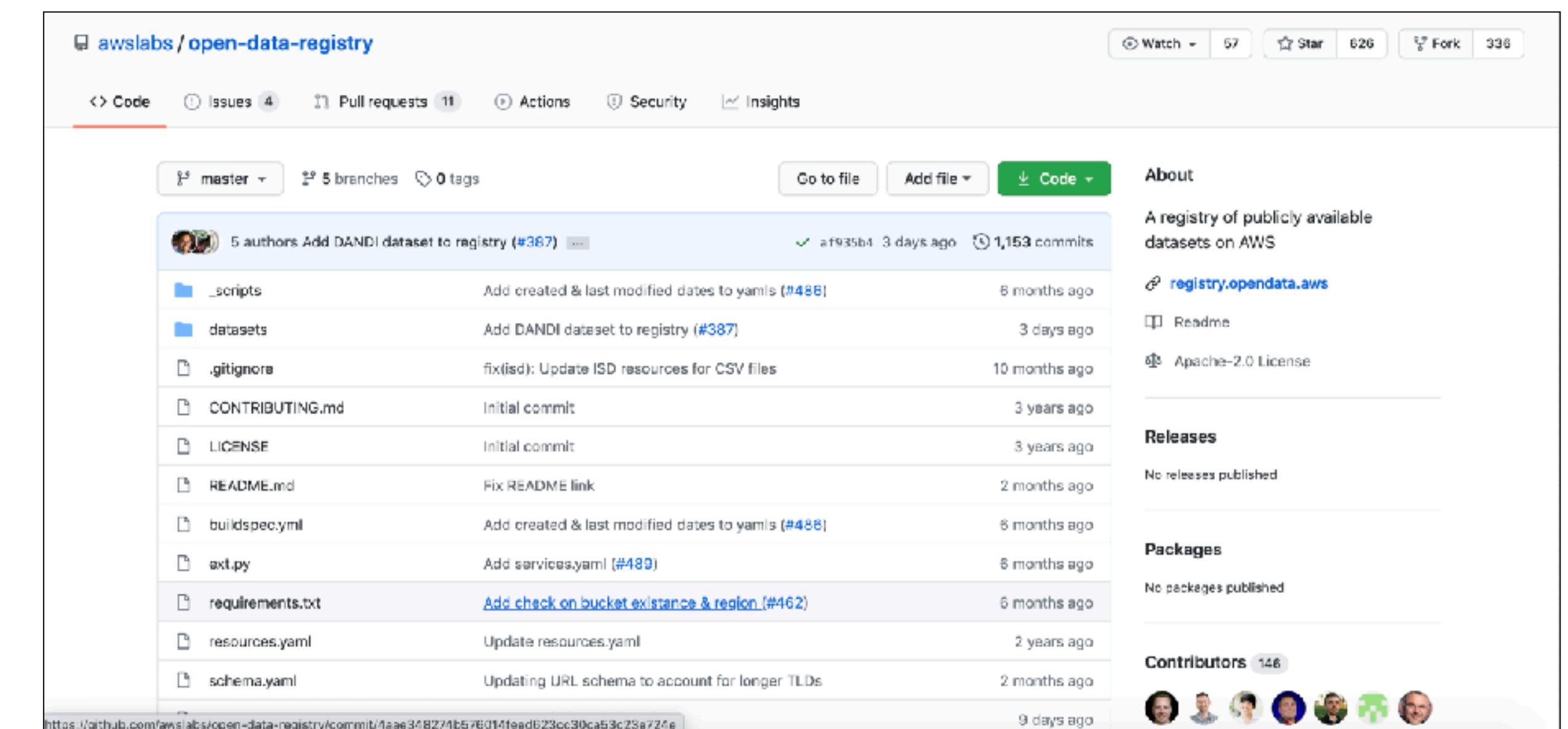
[Readme](#)

[Apache-2.0 License](#)

Releases
No releases published

Packages
No packages published

Contributors 146



[https://aws.amazon.com/ru/opendata/
open-data-sponsorship-program/](https://aws.amazon.com/ru/opendata/open-data-sponsorship-program/)

Microsoft Research Open Data

The screenshot shows the Microsoft Research Open Data homepage. At the top, there's a dark header bar with the Microsoft logo, the site name "Microsoft Research Open Data", and links for "Categories", "About", "FAQs", "Feedback", and "Login". Below the header is a large, dark blue banner featuring a network graph pattern. The central text on the banner reads "Microsoft Research Open Data" in white, with a "BETA" badge to its right. Below this, there's a search bar with the placeholder "Search datasets" and a magnifying glass icon. Two descriptive text blocks are visible: one on the left about datasets for research in areas like natural language processing and computer vision, and one on the right about domain-specific sciences. The main content area below the banner has a white background and features a section titled "Dataset Categories" with two items: "Computer Science" (represented by a keyboard icon) and "Social Science" (represented by a speech bubble icon). Each category has a "VIEW DATASETS >" link.

Microsoft Research Open Data **BETA**

Search datasets

A collection of free datasets from Microsoft Research to advance state-of-the-art research in areas such as natural language processing, computer vision, and

domain specific sciences. Download or copy directly to a cloud-based Data Science Virtual Machine for a seamless development experience.

Dataset Categories

Computer Science
[VIEW DATASETS >](#)

Social Science
[VIEW DATASETS >](#)

<https://msropendata.com/>

Платформы распространения данных

CKAN

The CKAN data management platform is in use by numerous governments, organisations and communities around the world. Being [open source](#), we don't know of all the instances, but here is a showcase of the ones we are aware of. We are proud and excited to see all the different uses being made of CKAN – if you would like to see your CKAN instance here [add it to the CKAN Census!](#)

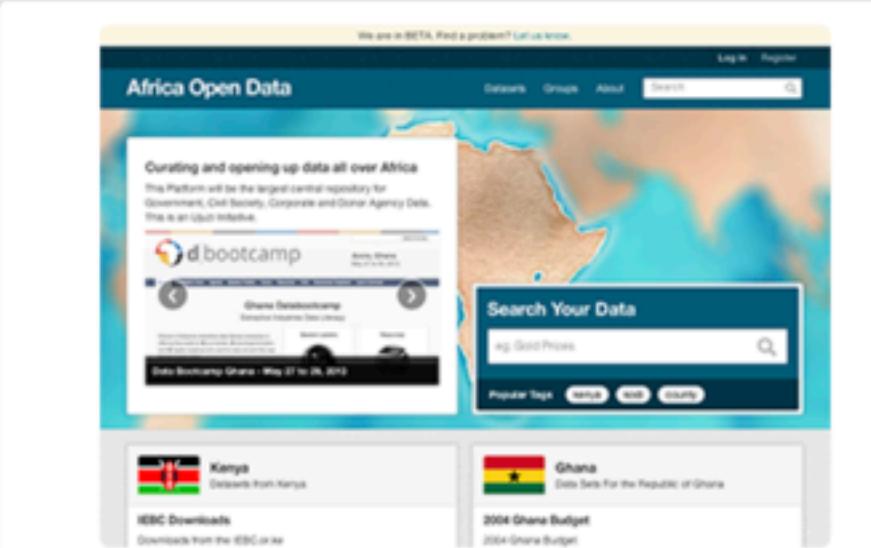
To view instances by region or type, just choose the correct facet below.

Show: [All](#) [Region](#) [Type](#)

199 instances



A.D.E.R.E. - Asociación de Entes Reguladores Eléctricos
A.D.E.R.E. - Asociación de Entes Reguladores



Africa Open Data
The Open Africa Platform initiative aims to be largest repository of Data on the African continent

RECENT POSTS

 [New CKAN 2.9 version released, patch releases for 2.7 and 2.8 available](#)

 [Subscribe to datasets: new CKAN feature explained](#)

 [Getting CKAN 2.9 over the line](#)

TWEETS

We are happy to announce the release of CKAN 2.9! Thanks to all our wonderful contributors. [#opendata #OpenSource](http://t.co/caDv7hRCVA)

<https://ckan.org/about/instances/>

DataHub

DataHub.io by DATAPLAN — We build solutions that unleash the potential of data. Let's start with yours! [Get in touch now >](#)

DATA HUB ABOUT BLOG FIND DATA COLLECTIONS DOCS PRICING TOOLS CHAT • LOGIN JOIN FREE

Search for datasets

There are thousands of datasets from financial market data and population growth to cryptocurrency prices. If you don't find what you are looking for [ask the Data Concierge](#) for a free quote for us to find you the data.

Search

VIX - CBOE Volatility Index updated daily
finance-vix core Files 2 1MB

CBOE Volatility Index (VIX) time-series dataset including daily open, close, high and low. The CBOE Volatility Index (VIX) is a key measure of market expectations of near-term volatility conveyed by S&P 500 stock index option prices introduced in 1993. Data From the VIX FAQ: In 1993, the [Explore Dataset >](#)

Premium Data Offer
Now you can request additional data and/or customized columns!

[Request Quote](#)
or Find out more

<https://datahub.io/search>

Проект Хаб открытых данных

The screenshot shows the homepage of the Hub of Open Data. At the top, there is a navigation bar with links for 'Пакеты данных' (Data Packages), 'Организации' (Organizations), 'Группы' (Groups), 'О проекте' (About the project), and a search bar. A user profile for 'Irina Radchenko' is visible on the right.

Поиск данных

Например, окружающая среда

Популярные теги: статистика, rosstat, емисс

Хаб открытых данных statistics

8,1k наборы данных | 37 организации | 87 группы

Администрация муниципального района "Белгородский район" Белгородской области

budget.gov.ru
Данные с сайта <http://budget.gov.ru>

<https://hubofdata.ru/>

Проект Wiki DATA

Wikidata logo

Main Page Discussion Read View source View history Search Wikidata

Main page Community portal Project chat Create a new Item Recent changes Random item Query Service Nearby Help Donate Lexicographical data Create a new Lexeme Recent changes Tools What links here Related changes Special pages Permanent link Page information Wikidata item In other projects Wikimedia Commons MediaWiki

Welcome to Wikidata

the free knowledge base with 89,530,714 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

Want to help translate? [Translate the missing messages.](#)

open
multilingual
free

collaborative
linked
structured

Welcome!

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.

Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others.

Wikidata also provides support to many other sites and services beyond just Wikimedia projects! The content of Wikidata is available under a free license, exported using standard formats, and can be interlinked to other open data sets on the linked data web.

Learn about data

New to the wonderful world of data? Develop and improve your data literacy through content designed to get you up to speed and feeling comfortable with the fundamentals in no time.



https://www.wikidata.org/wiki/Wikidata:Main_Page

Информационные ресурсы по поиску данных

Коллекция KDnuggets

The screenshot shows the KDnuggets website with a yellow header bar. The header includes the KDnuggets logo, social media links for Twitter, Facebook, and LinkedIn, and a search bar. Below the header is a navigation menu with links to Blog/News, Opinions, Tutorials, Top stories, Companies, Courses, Datasets, Education, Events (online), Jobs, Software, and Webinars. The main content area has a yellow header "Datasets for Data Mining, Data Science, and Machine Learning". It features social sharing buttons (Like 102, Share 102, Tweet, Share, 8.5K) and a "See also" section with links to Government data sites, Data APIs, and Google Dataset Search. A "Data repositories" section lists various datasets like Anacode Chinese Web Datastore, Appen Open Source Datasets, AssetMacro, Awesome Public Datasets, AWS Public Data Sets, and BigML. To the right, there's a "Latest News" sidebar with links to articles about Python projects, Simpson's Paradox, generative models, Coursera's Machine Learning course, consumer insights, and unpopular opinions. At the bottom, there are "Top Stories Last Week" and "Most Popular" sections.

Datasets for Data Mining, Data Science, and Machine Learning

Like 102 Share 102 Tweet Share 8.5K

See also

- Government, State, City, Local, public data sites and portals
- Data APIs, Hubs, Marketplaces, Platforms, and Search Engines.
- Google Dataset Search

Data repositories

- Anacode Chinese Web Datastore: a collection of crawled Chinese news and blogs in JSON format.
- Appen Open Source Datasets.
- AssetMacro, historical data of Macroeconomic Indicators and Market Data.
- Awesome Public Datasets on github, curated by caesar0301.
- AWS (Amazon Web Services) Public Data Sets, provides a centralized repository of public data sets that can be seamlessly integrated into AWS cloud-based applications.
- BigML: big list of public data sources

Latest News

- Automating Every Aspect of Your Python Project
- What is Simpson's Paradox and How to Automatically De...
- The Insiders' Guide to Generative and Discriminative ...
- Coursera's Machine Learning for Everyone Fulfils...
- How to Effectively Obtain Consumer Insights in a Data O...
- Unpopular Opinion – Data Scientists Should Be Mor...

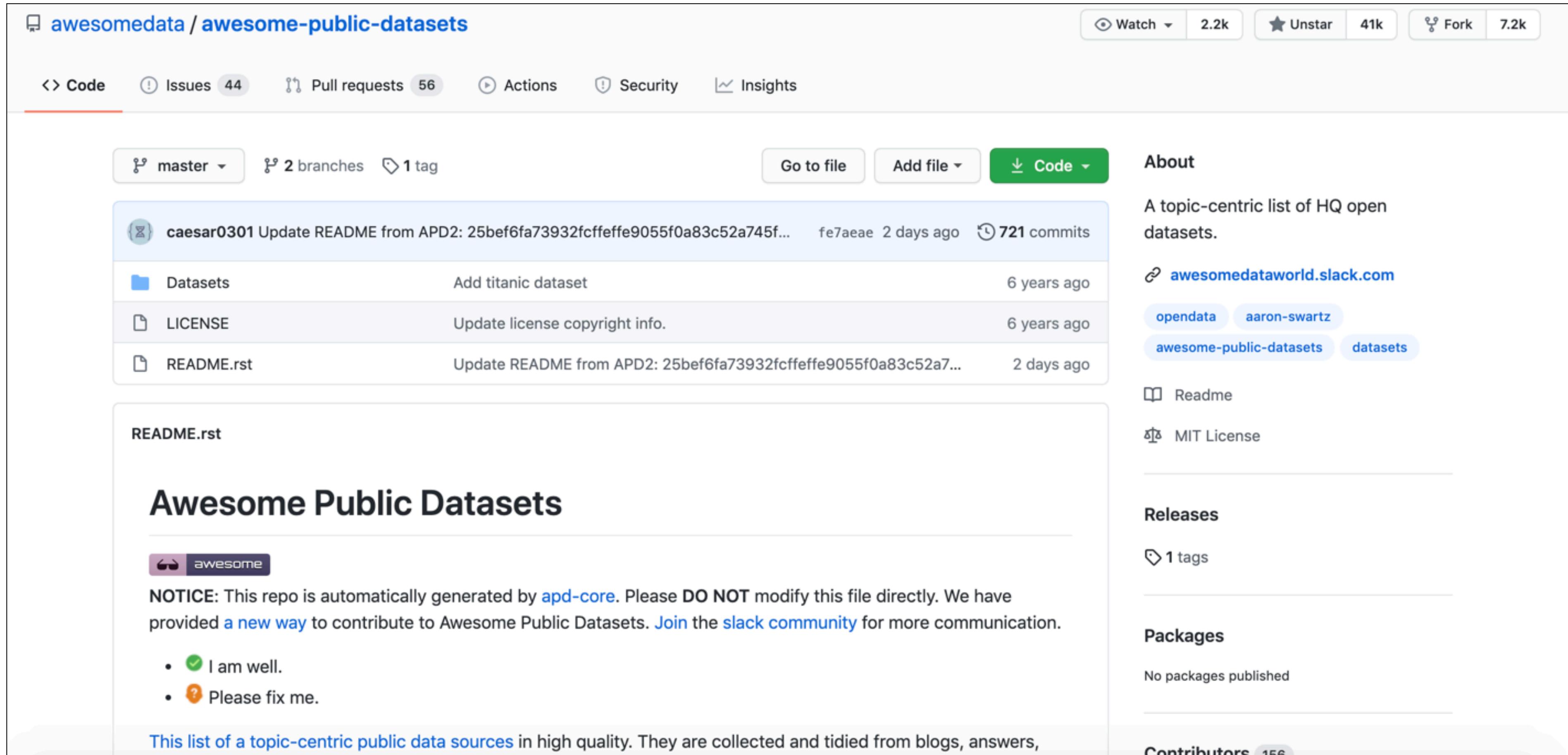
Top Stories Last Week

Most Popular

- Free From MIT: Intro to

<https://www.kdnuggets.com/datasets/index.html>

Awesome Public Datasets

A screenshot of a GitHub repository page for "awesomedata / awesome-public-datasets". The repository has 2.2k stars, 41k forks, and 7.2k issues. The "Code" tab is selected. The repository contains 2 branches and 1 tag. The master branch has 721 commits. Recent commits include adding the titanic dataset, updating license copyright info, and updating the README. The README file contains the "Awesome Public Datasets" list, which is a topic-centric list of HQ open datasets. It includes a badge for "awesome" and a notice that the repo is automatically generated by apd-core. Contributors can join a slack community for communication. The repository also links to awesomedataworld.slack.com and provides MIT License information.

awesomedata / awesome-public-datasets

Code Issues Pull requests Actions Security Insights

master 2 branches 1 tag Go to file Add file Code

caesar0301 Update README from APD2: 25bef6fa73932fcffeffe9055f0a83c52a745f... fe7aeae 2 days ago 721 commits

Datasets Add titanic dataset 6 years ago

LICENSE Update license copyright info. 6 years ago

README.rst Update README from APD2: 25bef6fa73932fcffeffe9055f0a83c52a7... 2 days ago

README.rst

Awesome Public Datasets

awesome

NOTICE: This repo is automatically generated by [apd-core](#). Please **DO NOT** modify this file directly. We have provided [a new way](#) to contribute to Awesome Public Datasets. [Join the slack community](#) for more communication.

- ✓ I am well.
- ? Please fix me.

This list of a topic-centric public data sources in high quality. They are collected and tidied from blogs, answers,

About

A topic-centric list of HQ open datasets.

awesomedataworld.slack.com

opendata aaron-swartz awesome-public-datasets datasets

Readme

MIT License

Releases

1 tags

Packages

No packages published

Contributors 156

<https://github.com/awesomedata/awesome-public-datasets>

Открытые данные на русском языке

[infoculture / awesome-opendata-rus](https://github.com/infoculture/awesome-opendata-rus)

Code Issues Pull requests Actions Projects Wiki Security Insights

master 1 branch 0 tags Go to file Add file Code

ivbeg Merge pull request #8 from definitelymaybe/patch-2 ... 55e8a25 on Jul 27 29 commits

.Rhistory Добавлены региональные порталы открытых данных в раздел Гос... 5 years ago

LICENSE Initial commit 5 years ago

README.md Update README.md 2 years ago

README.md

Russian awesome open data list  awesome

Opendata resources in Russian / Открытые данные на русском языке

Этот список пополняется сообществом, для внесения изменений откройте этот файл [для редактирования](#) и сохраните изменения. Они будут оперативно рассмотрены модератором. Если Вы хотите редактировать файл регулярно - напишите ibegin@infoculture или <https://t.me/begin>

Table of contents / Содержание

About

Opendata resources in Russian /
Открытые данные на русском языке

okf-russia moscow-data-science
opendata opendata-resources
awesome awesome-list data
russian

Readme
CC0-1.0 License

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

<https://github.com/infoculture/awesome-opendata-rus>

BuzzFeed

The screenshot shows the GitHub profile of the BuzzFeed News organization. The repository count is 108, and there are 3 pinned repositories. A search bar, type filter (All), and language filter (All) are visible. A specific repository, `nics-firearm-background-checks`, is highlighted.

BuzzFeed News
Open-source data, analysis, libraries, tools, and guides from BuzzFeed's newsroom.
[https://github.com/buzzfeednews/...](https://github.com/buzzfeednews/)

Repositories 108 Packages People 3 Projects

Pinned repositories

everything
An index of all our open-source data, analysis, libraries, tools, and guides.
922 stars, 99 forks

Find a repository... Type: All Language: All

nics-firearm-background-checks
Monthly data from the FBI's National Instant Criminal Background Check System, converted from PDF to CSV.
Python, MIT, 106 issues, 107 stars, 0 forks, 0 open issues, Updated 19 days ago

Top languages
Jupyter Notebook, HTML, Python, R, Shell

<https://github.com/BuzzFeedNews>

Pew Research Centre

The screenshot shows the Pew Research Center's website for Internet & Technology. At the top, there's a navigation bar with links for 'ABOUT', 'FOLLOW', 'MY ACCOUNT', and 'DONATE'. Below the header, the main navigation menu includes categories like 'HOME', 'U.S. POLITICS', 'MEDIA & NEWS', 'SOCIAL TRENDS', 'RELIGION', 'INTERNET & TECH' (which is currently selected), 'SCIENCE', 'HISPANICS', 'GLOBAL', and 'METHODS'. A secondary navigation bar below features 'PUBLICATIONS', 'TOPICS', 'PRESENTATIONS', 'DATASETS' (selected), 'INTERACTIVES', 'FACT SHEETS', and 'OUR EXPERTS'. The main content area is titled 'Datasets' and displays a message: 'Displaying 1-10 of 143 results'. To the right, there's a 'REFINE YOUR RESULTS' sidebar with an 'Update' button and a 'DATE' section showing options for 'Past 6 Months (3)', 'Past 12 Months (4)', and 'Past 2 Years (7)'. There's also a dropdown for 'Years' and a 'RANGE OF YEARS' input field with an 'Update' button. On the left, there's a 'Log in' form with fields for 'Email Address *' and 'Password *', both marked with red asterisks. A note says 'By signing in to your account, you agree to our [Terms of Use](#)'. A 'LOGIN' button is at the bottom of the form.

<https://www.pewresearch.org/internet/datasets/>

Спасибо за внимание!

Телеграм-канал: @DataPlace

Сайт: <http://iRadche.ru>



@dadaistka



@iRadche



<https://www.facebook.com/iRadche>



<http://www.slideshare.net/iRadche>