

# 【Open Data】国外开放数据中心及政府数据开放平台汇总

发布日期：2019-08-20，最近更新：9 天前

作者：探码科技，原文链接：<http://www.tanmer.com/blog/537>

---

2015年国际科学理事会发布《Open Data in a Big Data World》协议。该协议是“2015年国际科学”的成果，是国际科学四大顶级代表（国际科学理事会 - ICSU，国际合作伙伴关系 - IAP，世界科学院 - TWAS和国际社会科学理事会 - ISSC）系列年度会议的第一次。旨在代表全球科学界在科学领域的国际政策。该协议将数据革命的机遇和挑战确定为当今全球科学政策的主要问题。该协议认为“大数据”已经成为科学发现的重要机会，而“开放数据”将提高公共研究企业的效率，生产力和创造力，并

参考资料：

- Open Data in a Big Data World ([https://council.science/cms/2017/04/open-data-in-big-data-world\\_long.pdf](https://council.science/cms/2017/04/open-data-in-big-data-world_long.pdf))
- Annual Open Data Plan for 2018 ([https://www.ogcio.gov.hk/en/about\\_us/annual\\_open\\_data\\_plans/doc/Annual-Open-Data-Plan-2018-OGCIO-en.pdf](https://www.ogcio.gov.hk/en/about_us/annual_open_data_plans/doc/Annual-Open-Data-Plan-2018-OGCIO-en.pdf))
- International Open Data Charter (<https://opendatacharter.net/principles/>)

## 纽约政府开放数据平台

<https://opendata.cityofnewyork.us/> (<https://opendata.cityofnewyork.us/>)

## 美国官网数据超市

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

提供230,256个数据集、14个数据目录

## 新加坡政府开放数据平台

<https://data.gov.sg/> (<https://data.gov.sg/>)

提供1700个数据集、9个数据目录

## 休斯顿市开放数据门户网站

<http://data.houstontx.gov/> (<http://data.houstontx.gov/>)

247个数据集、12个数据目录

## Academic Torrents

<http://academictorrents.com/> (<http://academictorrents.com/>)

共享大量数据集的分布式系统，提供445.96TB的研究数据

## hadoopilluminated.com

[http://hadoopilluminated.com/hadoop\\_illuminated/Public\\_Bigdata\\_Sets.html](http://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html)  
([http://hadoopilluminated.com/hadoop\\_illuminated/Public\\_Bigdata\\_Sets.html](http://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html))

提供国外开放数据网站相关信息，目前已集合35个查询途径

## United States Census Bureau

<https://www.census.gov/> (<https://www.census.gov/>)

美国人口普查局

## usgovxml.com

<http://usgovxml.com/> (<http://usgovxml.com/>)

USGovXML.com是美国政府提供的公共Web服务和XML数据源的索引。USGovXML.com索引来自 美国政府 (<http://www.usa.gov/Agencies/federal.shtml>)所有3个分支机构以及董事会，委员会，公司和独立机构的数据来源。

## enigma.com

<https://www.enigma.com/> (<https://www.enigma.com/>)

快速搜索和分析政府、公司和组织发布的数十亿份公共记录。

## datahub.io

<https://datahub.io/> (<https://datahub.io/>)

发现和分享高质量数据集，与他人联系和分享知识。

## aws.amazon.com/datasets

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

帮助人们发现和共享通过AWS资源提供的数据集。

## databib.org

<http://databib.org/> (<http://databib.org/>)

开放数据网站导航

## quandl.com

<https://www.quandl.com/> (<https://www.quandl.com/>)

金融，经济和替代数据集的主要来源，为投资专业人士提供服务。Quandl的平台被超过40万人使用，其中包括来自世界顶级对冲基金，资产管理公司和投资银行的分析师。

## figshare.com

<https://figshare.com/> (<https://figshare.com/>)

研究论文上传网站，已有2600万+浏览量、750万+下载、800,000+上传、200万+文章

## GeoLite Legacy Downloadable Databases

<https://dev.maxmind.com/geoip/geoip2/geolite2/>  
(<https://dev.maxmind.com/geoip/geoip2/geolite2/>)

IP地理定位数据库

## Quora's Big Datasets Answer

<https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>  
(<https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>)

公共开放数据集汇总

## Kaggle Datasets

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

数据文档，拥有20394个数据集

## A Deep Catalog of Human Genetic Variation

<https://www.internationalgenome.org/data> (<https://www.internationalgenome.org/data>)

国际基因组样本资源

## Google Public Data

<https://www.google.com/publicdata/directory>

(<https://www.google.com/publicdata/directory>)

谷歌公开数据搜索网站

## World Bank Data

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

世界银行开放数据搜索网站

## NYC Taxi data

<http://chriswhong.github.io/nyctaxi/> (<http://chriswhong.github.io/nyctaxi/>)

纽约出租车数据开放平台

## Open Data Philly

<https://www.opendataphilly.org/> (<https://www.opendataphilly.org/>)

费城开放数据平台、16个数据目录、354个数据集

## grouplens.org

<https://grouplens.org/datasets/> (<https://grouplens.org/datasets/>)

提供9个数据集，关于书籍、电源、wiki数据集

## UC Irvine Machine Learning Repository

<http://archive.ics.uci.edu/ml/index.php> (<http://archive.ics.uci.edu/ml/index.php>)

加州大学欧文机器学习库，提供481个数据集

## research-quality data sets by Hilary Mason

<http://web.archive.org/web/20150320022752/https://bitly.com/bundles/hmason/1>

(<http://web.archive.org/web/20150320022752/https://bitly.com/bundles/hmason/1>)

公共数据集汇总

## National Climatic Data Center - NOAA

<https://www.ncdc.noaa.gov/> (<https://www.ncdc.noaa.gov/>)

美国国家环境信息中心，监测，评估和提供国家气候和历史天气数据和信息

## ClimateData.us

<http://www.climatedata.us/> (<http://www.climatedata.us/>)

美国宇航局公布的美国气候数据

## r/datasets

<https://www.reddit.com/r/datasets/> (<https://www.reddit.com/r/datasets/>)

开放数据集汇总网站

## MapLight

<https://maplight.org/data/> (<https://maplight.org/data/>)

关于货币的数据集

## GHDx

<http://ghdx.healthdata.org/> (<http://ghdx.healthdata.org/>)

健康指标和评估研究所 - 来自世界各地的健康和人口统计数据集目录，包括IHME结果

## St. Louis Federal Reserve Economic Data - FRED

<https://fred.stlouisfed.org/> (<https://fred.stlouisfed.org/>)

圣路易斯联邦储备银行数据开放网站，该网站提供丰富的经济数据和信息，以促进经济教育和加强经济研究。

## New Zealand Institute of Economic Research – Data1850

<https://data1850.nz/> (<https://data1850.nz/>)

新西兰经济研究所，可在该网站下载自1850年以来的相关经济数据。

## Dept. of Politics @ New York University

[http://www.nyu.edu/projects/politicsdatalab/datasupp\\_datasources.html](http://www.nyu.edu/projects/politicsdatalab/datasupp_datasources.html)  
([http://www.nyu.edu/projects/politicsdatalab/datasupp\\_datasources.html](http://www.nyu.edu/projects/politicsdatalab/datasupp_datasources.html))

纽约大学政治数据中心

## Open Data Sources

<https://github.com/datasciencemasters/data>  
(<https://github.com/datasciencemasters/data>)

Github网站上的开放数据源总结

## UNICEF Statistics and Monitoring

[https://www.unicef.org/statistics/index\\_24287.html](https://www.unicef.org/statistics/index_24287.html)

([https://www.unicef.org/statistics/index\\_24287.html](https://www.unicef.org/statistics/index_24287.html))

联合国儿童基金会官网，开放世界各国、地区的儿童状况报告

## undata

<http://data.un.org/Default.aspx> (<http://data.un.org/Default.aspx>)

联合国国际统计数据库，包含6,000多万个数据点，涵盖广泛的统计主题，包括农业，犯罪，通信，发展援助，教育，能源，环境，金融，性别，健康，劳动力市场，制造业，国民核算，人口与移民，科学技术，旅游，运输和贸易。

## NASA SocioEconomic Data and Applications Center - SEDAC

<https://sedac.ciesin.columbia.edu/> (<https://sedac.ciesin.columbia.edu/>)

社会经济数据和应用中心，是美国国家航空航天局地球观测系统数据和信息系统（EOSDIS）中的分布式主动档案中心（DAAC）之一。

## The GDELT Project

<https://www.gdeltproject.org/#intro> (<https://www.gdeltproject.org/#intro>)

GDELT博客是世界上最大的人类社会开放研究平台的最新新闻，公告，信息和应用程序的官方一站式存储库。

## Sweden, Statistics

<https://www.scb.se/en/> (<https://www.scb.se/en/>)

瑞典统计局，提供瑞典国家统计数据，包含26个数据集。

## Github free data source list

<https://www.datasciencecentral.com/profiles/blogs/great-github-list-of-public-data-sets>  
(<https://www.datasciencecentral.com/profiles/blogs/great-github-list-of-public-data-sets>)

Github公共数据集

## StackExchange Data Explorer

<https://data.stackexchange.com/> (<https://data.stackexchange.com/>)

一个开源工具，用于对来自StackExchange网络的公共数据进行任意查询。

## San Fransisco Government Open Data

<https://datasf.org/opendata/> (<https://datasf.org/opendata/>)

旧金山政府开发数据网站

## IBM Blog about open data

<https://www.datasciencecentral.com/profiles/blogs/the-free-big-data-sources-everyone-should-know> (<https://www.datasciencecentral.com/profiles/blogs/the-free-big-data-sources-everyone-should-know>)

数据科学中心

## Liver Tumor Segmentation Challenge Dataset

<https://competitions.codalab.org/competitions/17094>  
(<https://competitions.codalab.org/competitions/17094>)

## Public Git Archive

<https://github.com/src-d/datasets/tree/master/PublicGitArchive> (<https://github.com/src-d/datasets/tree/master/PublicGitArchive>)

Git Hub开放数据平台汇总

## GHTorrent

<http://ghtorrent.org/> (<http://ghtorrent.org/>)

## Microsoft Research Open Data

<https://msropendata.com/> (<https://msropendata.com/>)

来自Microsoft Research的免费数据集，以推进自然语言处理，计算机视觉和特定领域科学等领域的最新研究。

## Open Government Data Platform India

<https://data.gov.in/> (<https://data.gov.in/>)

印度开放政府数据（OGD）平台-data.gov.in-是一个用于支持印度政府开放数据倡议的平台。

## Google Dataset Search (beta)

<https://toolbox.google.com/datasetsearch> (<https://toolbox.google.com/datasetsearch>)

谷歌数据集搜索门户