

(Extracted from <https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/>)

Feb 12, 2016

Big Data: 33 Brilliant And Free Data Sources Anyone Can Use



[Bernard Marr](#) Contributor
[Enterprise Tech](#)

Data is ubiquitous — but sometimes it can be hard to see the forest for the trees, as it were. Many companies of various sizes believe they have to collect their own data to see benefits from big data analytics, but it's simply not true.

There are hundreds (if not thousands) of free data sets available, ready to be used and analyzed by anyone willing to look for them. Below is a list of 35 of the most globally interesting I've come across, but there are many, many more in many different niches.

1. Data.gov <http://data.gov> The US Government pledged last year to make all government data available freely online. This site is the first stage and acts as a portal to all sorts of amazing information on everything from climate to crime.
2. US Census Bureau <http://www.census.gov/data.html> A wealth of information on the lives of US citizens covering population data, geographic data and education.
3. [Socrata](#) is another interesting place to explore government-related data, with some visualisation tools built-in.
4. European Union Open Data Portal <http://open-data.europa.eu/en/data/> As the above, but based on data from European Union institutions.
5. Data.gov.uk <http://data.gov.uk/> Data from the UK Government, including the British National Bibliography – metadata on all UK books and publications since 1950.
6. [Canada Open Data](#) is a pilot project with many government and geospatial datasets.
7. [Datacatalogs.org](#) offers open government data from US, EU, Canada, CKAN, and more.
8. The CIA World Factbook <https://www.cia.gov/library/publications/the-world-factbook/> Information on history, population, economy, government, infrastructure and military of 267 countries.
9. Healthdata.gov <https://www.healthdata.gov/> 125 years of US healthcare data including claim-level Medicare data, epidemiology and population statistics.

10. NHS [Health](http://www.hscic.gov.uk/home) and Social Care Information Centre <http://www.hscic.gov.uk/home> Health data sets from the UK National Health Service.
11. [UNICEF](#) offers statistics on the situation of women and children worldwide.
12. [World Health Organization](#) offers world hunger, health, and disease statistics.
13. Amazon Web Services public datasets <http://aws.amazon.com/datasets> Huge resource of public data, including the 1000 Genome Project, an attempt to build the most comprehensive database of human genetic information and [NASA](#) 's database of satellite imagery of Earth.
14. [Facebook](#) Graph <https://developers.facebook.com/docs/graph-api> Although much of the information on users' Facebook profile is private, a lot isn't – Facebook provide the Graph API as a way of querying the huge amount of information that its users are happy to share with the world (or can't hide because they haven't worked out how the privacy settings work).
15. [Face.com](#): A fascinating tool for facial recognition data.
16. [UCLA](#) makes some of the data from its courses public.
17. [Data Market](#) is a place to check out data related to economics, healthcare, food and agriculture, and the automotive industry.
18. [Google Public data explorer](#) includes data from world development indicators, OECD, and human development indicators, mostly related to economics data and the world.
19. [Junar](#) is a data scraping service that also includes data feeds.
20. [Buzzdata](#) is a social data sharing service that allows you to upload your own data and connect with others who are uploading their data.
21. Gapminder <http://www.gapminder.org/data/> Compilation of data from sources including the World Health Organization and World Bank covering economic, medical and social statistics from around the world.
22. [Google Trends](#) <http://www.google.com/trends/explore> Statistics on search volume (as a proportion of total search) for any given term, since 2004.
23. Google Finance <https://www.google.com/finance> 40 years' worth of stock market data, updated in real time.
24. Google Books Ngrams <http://storage.googleapis.com/books/ngrams/books/datasetv2.html> Search and analyze the full text of any of the millions of books digitised as part of the Google Books project.
25. National Climatic Data Center <http://www.ncdc.noaa.gov/data-access/quick-links#loc-clim> Huge collection of environmental, meteorological and climate data sets from the US National Climatic Data Center. The world's largest archive of weather data.
26. DBPedia <http://wiki.dbpedia.org> Wikipedia is comprised of millions of pieces of data, structured and unstructured on every subject under the sun. DBPedia is an ambitious project to catalogue and create a public, freely distributable database allowing anyone to analyze this data.
27. [New York Times](#) <http://developer.nytimes.com/docs> Searchable, indexed archive of news articles going back to 1851.
28. Freebase <http://www.freebase.com/> A community-compiled database of structured data about people, places and things, with over 45 million entries.
29. Million Song Data Set <http://aws.amazon.com/datasets/6468931156960467> Metadata on over a million songs and pieces of music. Part of Amazon Web Services.

30. [UCI Machine Learning Repository](#) is a dataset specifically pre-processed for machine learning.
31. [Financial Data Finder at OSU](#) offers a large catalog of financial data sets.
32. [Pew Research Center](#) offers its raw data from its fascinating research into American life.
33. [The BROAD Institute](#) offers a number of cancer-related datasets.