# Big Data And AI: 30 Amazing (And Free) Public Data Sources For 2018

**Bernard Marr** Contributor ⓘ
Enterprise Tech

Machine learning, artificial intelligence, blockchains, predictive analytics – all amazing technologies which have promised to revolutionize business and society.



Adobe Stock    ADOBE STOCK

They are useless, however, without data. Fortunately for businesses and organizatio which don't have the resources to methodically collect every piece of useful information, they will need themselves, a huge (and growing) amount is available freely online.

Two years ago I wrote an article listing 33 sources of Big Data available for free online. Of course, in business technology terms that was a lifetime ago, so here's an update with thirty new entries:

**Today In:** Tech                                                                          ⌄

1. [World Bank Open Data](#) Datasets covering population demographics and a huge number of economic and development indicators from across the world
2. [IMF Data](#) The International Monetary Fund publishes data on international finances, debt rates, foreign exchange reserves, commodity prices and investments.
3. [The US National Center for Education Statistics](#) Data on educational institutions and education demographics from the US and around the world.
4. [The UK Data Centre](#) The UK's largest collection of social, economic and population data.
5. [FiveThirtyEight](#) A large number of polls providing data on public opinion of political and sporting issues.
6. [FBI Uniform Crime Reporting](#) The FBI is responsible for compiling and publishing national crime statistics, with free data available at national, state and county level.
7. [Bureau of Justice](#) Here you can find data on law enforcement agencies, jails, parole and probation agencies and courts.
8. [Qlick Data Market](#) Offers a free package with access to datasets covering world population, currencies, development indicators and weather data.
9. [NASA Exoplanet Archive](#) Public datasets covering planets and stars gathered by NASA's space exploration missions.
10. [UN Comtrade Database](#) Statistics compiled and published by the United Nations on international trade. Includes Comtrade Lab which is a showcase of how cutting edge analytics and tools are used to extract value from the data.
11. [Financial Times Market Data](#) Up to date information on financial markets from around the world, including stock price indexes, commodities and foreig exchange.
12. [Google Trends](#) Examine and analyze data on internet search activity and trending news stories around the world.
13. [Twitter](#) The advantage Twitter has over the others are that most conversation are public. This means that huge amounts of data is available through their API on who is talking about what, where, when and why.
14. [Google Scholar](#) Entire texts of academic papers, journals, books and legal case law.
15. [Instagram](#) As with Twitter, Instagram posts and conversations are public by default. Their APIs allow likes, mentions and business details to be analyzed.
16. [OpenCorporates](#) The world's largest open database of companies.

17. [Glassdoor API](#) Information about job vacancies, candidates, salaries and employee satisfaction is available through their developer API.
18. [IMDB Datasets](#) Datasets in a number of formats drawn from the web's largest resource on movies, television and people working in those industries.
19. [OpenLibrary Data Dumps](#) Datasets on books including catalogs from libraries around the world
20. [Labelled Faces in the Wild](#) 13,000 collated and labeled images of human faces for use in developing applications involving facial recognition.
21. [Microsoft Marco](#) Microsoft's open machine learning datasets for training systems in reading comprehension and question answering.
22. [Machine Learning Dataset Repository](#) Collection of open datasets contributed by data scientists involved in machine learning projects.
23. [eBay Market Data Insights](#) Data on millions of online sales and auctions from eBay
24. [Natural History Museum Data Portal](#) Information on nearly 4 million historical specimens in the London museum's collection, as well as scientific sound recordings of the natural world.
25. [CERN Open Data](#) More than one petabyte of data from particle physics experiments carried out by CERN.
26. [One Million Audio Cover Images](#) Dataset hosted at archive.org covering music released around the world, for use in image processing research
27. [Complete Public Reddit Comments Corpus](#) Over one billion public comments posted to Reddit between 2007 and 2015, for training language algorithms
28. [Microsoft Azure Data Markets Free Datasets](#) Freely available datasets covering everything from agriculture to weather
29. [Irish Electric Vehicle Charge Point Status](#) Collates data from the body which oversees the network of EV charge points across the Republic of Ireland and Northern Ireland.
30. [LondonAir](#) Pollution and air quality data from across London

I hope these sources are useful and as new ones become available every day I will be updating this list on a regular basis, so stay connected.

*Follow me on [Twitter](#) or [LinkedIn](#). Check out my [website](#).*

**Bernard Marr**

Bernard Marr is an internationally best-selling author, popular keynote speaker, futurist, and a strategic business & technology advisor to governments and companies... **Read More**

Bernard Marr is an internationally best-selling author, popular keynote speaker, futurist, and a strategic business & technology advisor to governments and companies... **Read More**