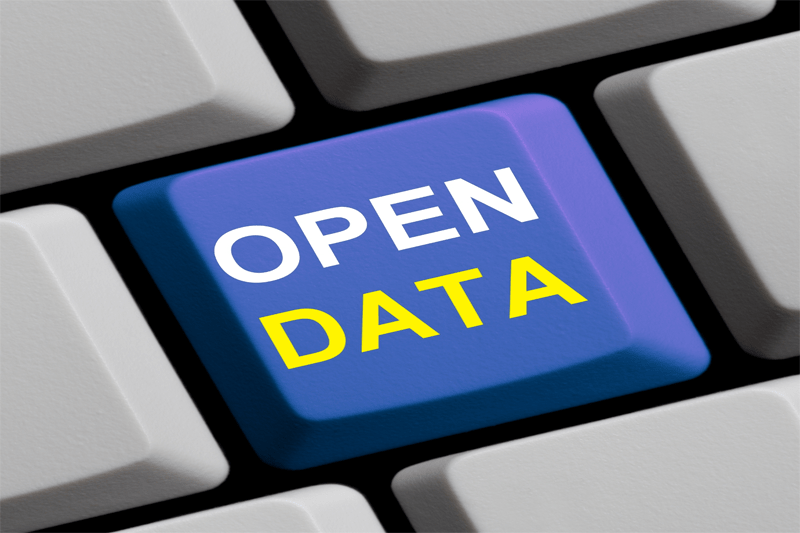
# O Poder do Open Data

David Matos

12 fev. 2020

Fonte: https://www.cienciaedados.com/o-poder-do-open-data/



Open data (também chamado public datasets) faz parte de uma tendência em direção a um conceito mais amplo sobre o que é propriedade intelectual, uma tendência que tem ganhado enorme popularidade ao longo da última década. Dados abertos são dados que foram disponibilizadas ao público e que podem ser utilizados, reutilizados, analisados e compartilhados com os outros. O open data faz parte de um movimento que inclui software open-source, hardware livre, trabalho criativo de conteúdo aberto, acesso aberto a publicações científicas e a ciência livre – todos comprometidos com a noção de que o conteúdo (incluindo dados brutos obtidos a partir de experimentos) devem ser compartilhados livremente.

A característica que distingue licenças de uso livre, são que elas tenham [Copyleft](https://copyleft.org/), ao invés de direitos autorais. Com [Copyleft](https://copyleft.org/), a única restrição é que a fonte do trabalho deve ser identificada, por vezes, com a ressalva de que trabalhos derivados não podem ser protegidos por direitos autorais com uma licença mais restritiva do que o original. Desta forma, comercializar com sucesso o trabalho de origem [Copyleft](https://copyleft.org/), torna-se difícil, embora seja possível encontrar formas criativas e indiretas de comercialização. Esteja ciente de que, por vezes, mesmo um trabalho rotulado como aberto pode realmente não se encaixar na definição acima. Você é responsável por verificar os direitos de licenciamento e restrições de dados abertos que você utilizar.

As pessoas muitas vezes confundem licenças de uso livre com licenças [Creative Commons](http://creativecommons.org/). [Creative Commons](http://creativecommons.org/) é uma organização sem fins lucrativos, dedicada a incentivar e difundir obras criativas, oferecendo um quadro jurídico através do qual as permissões de uso podem ser concedidas e obtidas, de modo que os autores e as pessoas ou empresas que utilizaram os dados, estejam a salvo de riscos legais provenientes do conhecimento presente no material compartilhado livremente. Algumas licenças [Creative Commons](http://creativecommons.org/) estão abertas e alguns explicitamente proíbem obras e / ou comercialização de derivados.

Como parte das iniciativas de open data, surgiu o conceito de *governo aberto*. Os governos de todo o mundo começaram a liberar dados livres para consulta e utilização. Os governos geralmente fornecem esses dados de modo que ele possa ser usado por analistas e [cientistas de dados](https://www.cienciaedados.com/cientista-de-dados-parte-1/), além de programadores que trabalham em colaboração para construir soluções open-source que utilizam dados abertos para resolver problemas sociais – em um esforço para beneficiar a sociedade em geral. Em 2013, o G8 (França, Estados Unidos, Reino Unido, Rússia, Alemanha, Japão, Itália e Canadá), assinou uma carta comprometendo-se a abrir os dados, priorizando as áreas de estatísticas nacionais, resultados eleitorais, orçamentos governamentais e mapas nacionais. O movimento de um governo aberto promove a transparência e a prestação de contas do governo, nutre um eleitorado bem informado e incentiva a participação do público. Para colocá-lo em termos de computação, um governo aberto facilita uma relação de leitura / gravação entre um governo e seus cidadãos.

Os dados abertos têm outra imensa vantagem. Eles podem ser usados para validar modelos de análise e calibrar algoritmos de [Machine Learning](https://www.cienciaedados.com/fascinante-machine-learning/). Trabalhando com dados reais, os [Cientistas de Dados](https://www.cienciaedados.com/cientista-de-dados-x-engenheiro-de-dados/) podem construir seus modelos e validar os resultados, antes de liberar suas aplicações para os usuários finais. Os dados abertos também podem ser usados para complementar os dados obtidos internamente dentro de uma empresa. Os dados internos podem ser usados em conjunto com fontes externas para se obter um resultado final mais completo possível e que mostre a influência do mundo externo nas decisões da empresa e vice-versa. A [Open Data Foundation](http://www.odaf.org/) reúne empresas e pessoas interessadas em trabalhar com dados abertos.

Abaixo, algumas fontes de open data bem interessantes:

Dados do Governo do Brasil: <http://dados.gov.br/>

IPEA: <http://www.ipeadata.gov.br>

Banco Central do Brasil: [BCB](https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries)

Dados do Governo dos EUA: <http://data.gov>

Dados sobre as cidades americanas: <http://datasf.org>

Dados do Governo do Canadá (em inglês e francês): <http://open.canada.ca>

Dados do Governo do Reino Unido: <https://data.gov.uk>

Dados da União Europeia: <http://open-data.europa.eu/en/data>

Dados do Censo dos EUA (dados da população americana e mundial): <http://www.census.gov>

Dados da NASA: <https://data.nasa.gov>

Dados do Banco Mundial: <http://data.worldbank.org>

Dados sobre a saúde: <http://www.healthdata.gov>

Dados públicos da Amazon: <http://aws.amazon.com/datasets>

Dados sobre diversos países (incluindo o Brasil): <http://knoema.com>

Dados sobre diversas áreas de negócio e finanças: <https://www.quandl.com>

Google Trends: <https://www.google.com/trends>

Google Finance: <https://www.google.com/finance>

Gapminder: <http://www.gapminder.org/data>

Dados sobre milhões de músicas: <https://aws.amazon.com/datasets/million-song-dataset>

Dados sobre os mais diversos assuntos: <http://www.freebase.com>

DBpedia: <http://wiki.dbpedia.org/>

Open Data Monitor: <http://opendatamonitor.eu>

Open Data Network: <http://www.opendatanetwork.com>

R Datasets: <https://www.key2stats.com/data-sets/listing>

R Datasets packages: [R Dataset packages](http://stat.ethz.ch/R-manual/R-patched/library/datasets/html/00Index.html)

Datasets: <http://www.statsci.org/datasets.html>

Portal de Estatística: <http://www.statista.com>

Data 360: <http://www.data360.org>

Reconhecimento de Faces: <http://www.face-rec.org/databases>

Stanford Large Network Dataset Collection: <http://snap.stanford.edu/data>

Datahub: <http://datahub.io/dataset>

Esta é apenas uma pequena lista de dados abertos. O volume de dados não pára de aumentar e atualmente é possível encontrar dados livres sobre praticamente tudo que se possa imaginar.