Open Data

by Santiago Mota

2025-04-07

Recopilación de información sobre Open Data. Links, libros, blogs y otra información interesante.

Este fichero es copia de uno alojado en Github, en este repositorio y que se actualiza periódicamente.

Se ha incluido otra copia en Kaggle.

Y se aloja en las webs Github y Netlify.

https://github.com/santiagomota/Open_Data

https://santiagomota.github.io/Open_Data/

https://open-data-pages.netlify.app/

```r
# Creo el link con:
# usethis::create_download_url("https://github.com/santiagomota/Open_Data")

# Si me quiero bajar el repositorio completo en el directorio actual
usethis::use_course(
    "https://github.com/santiagomota/Open_Data/zipball/HEAD", destdir = ".")
```

Fuentes de datos abiertos y APIs

- 20 Awesome Websites For Collecting Big Data
- 25 Open Datasets for Deep Learning Every Data Scientist Must Work With
- 25 Satellite Maps To See Earth in New Ways
- 30 Amazing (And Free) Big Data And AI Public Data Sources For 2018
- 46 museos y bibliotecas que han digitalizado todo su conocimiento y lo ofrecen gratis en internet
- AENA - Estadísticas de tráfico aéreo
- Agencia Tributaria. Estadísticas
- AI for Copernicus - a data repository by CALLISTO
- AI4SmallFarms: A Data Set for Crop Field Delineation in Southeast Asian Smallholder Farms
- AID: A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification
- Alaska Satellite Facility
- Amazon product data 2014
- Amazon product data 2018
- Análisis de 1.100 millones de trayectos de taxis y uber en NYC
- API de Facebook
- API de GitHub
- Argo Floats - Global ocean observations of temperature, salinity, and pressure.
- API TomTom. Tráfico en ciudades
- Armed Conflict Location & Event Data Project (ACLED)
- ASTER Global DEM (GDEM) - ASTER Global Digital Elevation Model 1 arc second
- ArcticDEM - High-resolution DEM for the Arctic region
- Awesome Geospatial
- Awesome Public Datasets 1

- Awesome Public Datasets 2
- Awesome Sentinel. Copernicus Sentinel Satellites resources
- awesome-gee-community-datasets
- AWS Data Exchange
- AWS Datasets
- AWS Open Data Geo
- AWS Open Data
- Ayuntamiento de Madrid. Censo de locales, sus actividades y terrazas de hostelería y restauración
- Berkeley Earth - Global land temperature and air pollution datasets.
- Blog. 100 recursos sobre Big Data y Data Science
- British Ordnance Survey Data Hub
- BUILDING OUTLINE EXTRACTION OF ENSCHEDE, THE NETHERLANDS USING AERIAL IMAGES AND DIGITAL SURFACE MODELS
- CaixaBank Research
- CGIAR-CSI SRTM - SRTM 90m Digital Elevation Database v4.1
- Canada Open Government Portal
- Center for Applied Internet Data Analysis
- Center for Disease Control
- CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations - High-resolution precipitation data.
- CIS. Centro de Investigaciones Sociológicas
- Climate Data Online
- Climate Change Knowledge Porta - Country-specific climate risks, data, and projections.
- Climate TRACE
- Cómo los datos abiertos pueden ayudar en la crisis de los refugiados
- Copernicus Atmosphere Monitoring Service (CAMS) Global Near-Real-Time
- Copernicus Open Access Hub
- Copernicus DEM - European Digital Elevation Model (EU-DEM)
- Copernicus Marine Environment Monitoring Service (CMEMS) - Ocean monitoring for sea surface temperature, sea level, and salinity.
- CRAN Task View OpenData
- Crimen en UK
- DANS Data Station Physical and Technical Sciences
- Data Derived from OpenStreetMap for Download
- Data Kicks
- Data on CO2 and Greenhouse Gas Emissions by Our World in Data
- Data World
- Datasets de ejemplo de IBM Watson Analytics
- Datasets de Quandl
- Dataset4EO
- Datos abiertos Ayuntamiento de Valencia
- Datos abiertos de la Generalitat de Cataluña
- Datos abiertos de la Unión Europea
- Datos abiertos de Santander
- Datos abiertos del Ayuntamiento de Madrid
- Datos Abiertos del Consorcio Regional de Transportes de Madrid
- Datos abiertos del gobierno de España
- Datos abiertos Junta de Andalucía
- Datos de la Eurocopa 2024
- Datos de todos los vuelos en USA entre 1987 y 2008 (datos originales)
- Datos de todos los vuelos en USA entre 1987 y 2008 (otra fuente y ejemplos de uso en H2O). 120G
- Datos estadísticos DGT
- Datosclima. Base de datos meteo
- DH Network

- Digital Earth Africa (DE Africa) Map
- Dirección General de Tráfico (DGT)
- Dynamic World V1 Land Use
- EarthEnv-DEM90 digital elevation model - Global DEM created from multiple datasets
- EarthView dataset
- ECMWF ERA5 - Hourly reanalysis climate data (temperature, precipitation, wind, etc.).
- EM-DAT - The international disaster database
- EDGAR - Emissions Database for Global Atmospheric Research
- EnMAP. The German Spaceborne Imaging Spectrometer Mission
- El planeta Tierra en AWS
- ERA DATASET. Dataset and Deep Learning Benchmark for Event Recognition in Aerial Videos
- ERA5 Daily Aggregates - Latest Climate Reanalysis Produced by ECMWF / Copernicus Climate Change Service
- ESA OpenSR - Robust, accountable super-resolution for Sentinel-2 and beyond
- ESA Third Party Missions (TPM)
- ESA WorldCover 2021. Global land cover product at 10 m for 2021 based on Sentinel-1 and 2 data
- España. Estadísticas de mercado de trabajo
- España. Inmigración. Estadísticas
- España. Seguridad Social. Estadísticas
- Esri Open Data Hub
- European Banking Authority (EBA)
- European Data Portal
- European Forest Fire Information System (EFFIS)
- FAO Map Catalog
- FAO's Global Information System on Water and Agriculture
- FBREF - Estadísticas e Historia del Fútbol
- Fields of The World (FTW)
- Fivethirtyeight
- FLUXNET - Data from flux towers for carbon, water, and energy exchange monitoring.
- Fondo Monetario Internacional
- Free GIS Data
- Freshwater Ecoregions of the World
- Fuentes de datos espaciales (Diva-GIS)
- Functional Map of the World (fMoW) Dataset
- Gapminder
- gee-community-catalog
- geoBoundaries
- geodata.state.gov
- GEBCO (General Bathymetric Chart of the Oceans) - Bathymetric DEM for ocean floors
- Geonames Cities with population > 5000
- Geoportal Registradores
- Geospatial Data Catalogs
- Geospatial Data Abstraction Library (GDAL) links - Provides links to raster datasets from various organizations.
- GHSL - Global Human Settlement Layer
- Global Forest Change 2000-2023
- Global Flood Database v1 (2000-2018)
- Global Health Observatory (GHO) API
- GLOPOP-S. A global dataset of 7 billion individuals with socio-economic characteristics (sintetic) Data Github Paper
- Global Historical Climatology Network (GHCN) - Weather station data for precipitation, temperature, and more.
- Global Land Cover Facility - Land cover and vegetation datasets.
- Global Wildfire Information System (GWIS)

- Gobierno Estados Unidos
- Google Books Ngram Viewe
- Google Cloud Vision API
- Google Datset Search
- Google Earth Engine Catalog
- Google finanzas
- Google Open Buildings
- Google Patents Public Data
- Google Public Data
- Google-Microsoft-OSM Open Buildings - combined by VIDA
- Helsinki Open Data
- Hugging Face Datasets
- HydroRIVERS
- Idealista ux&tech
- idealista18 - 2018 real estate listings in Spain. 3 cities
- ImageNet database
- Infraestructura de Datos Espaciales de España
- Infraestructura de Datos Espaciales de la Comunidad de Madrid
- IPUMS GIS Boundary Files
- ISCGM Global Map
- ISIMIP3b bias-adjusted atmospheric climate input data
- JAXA's Global ALOS 3D World (AW3D30) - ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)"
- Kaggle datasets
- Kaggle Weekly Kernels Award Winner Announcements
- Land Information New Zealand (LINZ) Data Service
- Legacy Aircraft Noise and Performance (ANP) data
- LinkedIn - Data for Impact
- Lista de algunos datatsets dentro de paquetes de R
- M3LEO: A Multi-Modal Multi-Label Earth Observation Dataset
- Mapas de Open Street Maps
- Marine Regions
- Marine Cadastre (AIS)
- Mendeley Data
- Microsoft - A Planetary Computer for a Sustainable Future
- Microsoft Cognitive Services
- Microsoft Research Open Data
- More datasets for teaching data science: The expanded dslabs package
- Multi-Temporal Crop Classification with HLS Imagery across CONUS
- Multimodal Remote Sensing Benchmark Datasets for Land Cover Classification
- Naciones Unidas. Datos detallados de comercio global
- NAIP: National Agriculture Imagery Program
- NASA Common Metadata Repository (CMR) SpatioTemporal Asset Catalog (STAC)
- NASA Earth Observations (NEO)
- NASA
- NASA Fire Information for Resource Management System (FIRMS) Link1 Link2 - Near real-time data on wildfires from MODIS and VIIRS satellites.
- NASA Earthdata - Shuttle Radar Topography Mission (SRTM)
- NASA POWER (Prediction of Worldwide Energy Resources) - Provides global weather and solar radiation data for energy, agriculture, and environmental sectors.
- NASDAQ
- National Historical Geographic Information System (NHGIS)
- National Map (USGS) - National Elevation Dataset (NED), LiDAR, and more
- Natural Earth Data - Raster data for relief and shaded relief imagery.

- Natural Earth
- Nature Scientific Data
- NHS Digital
- NHSR datasets
- NLP Datasets
- NOAA Daily Global Historical Climatology Network - Kaggle dataset
- NOAA. Agencia de meteo. USA.
- NOAA Global Forecast System (GFS) - Weather forecasts for temperature, precipitation, and wind.
- OCDE Data
- One versus One - European football statistics
- Openaerialmap - Aerial imagery collected by individuals and organizations.
- Open Africa dataset
- Open Data Barometer
- Open data EMT
- Open Data Inception. 1.600 portales abiertos
- Open Data Renfe
- Open Data Sources Database
- Open High-Resolution Satellite Imagery: The WorldStrat Dataset – With Application to Super-Resolution
- Open Topography - Various high-resolution DEM datasets from LiDAR and other sources
- Open Trade Statistics
- openaddresses
- OpenCelliD - Open Database of Cell Towers
- Opendata del CERN **Error**
- Opendatasoft
- openflights.org/
- OpenGEOS data
- OpenWeatherMap
- OSM Landuse
- OSM-Building-Classification Data Code Paper - Classification of 67,705,475 buildings across the United States into residential and non-residential
- Overture - Fused-partitioned
- Overture Maps
- Paquete de R 'datasets'
- Paquete pasra acceder al API del Instituto de Canario de Estadística
- Pew Research Center
- Planet SkySat Public Ortho Imagery, Multispectral
- Propublica
- RapidAI4EO: A Corpus of Dense Time Series Satellite Imagery
- Rdatasets
- Recopilación de datasets de BigML
- Red Eléctrica Española (REE) - API
- Red Natura 2000
- Reddit datasets
- rspatialdata is a collection of data sources and tutorials on visualising spatial data using R
- SARDet-100K: large-scale multi-class SAR object detection dataset
- Satélite Landsat
- Satellite imagery datasets containing ships
- SEN12MS-CR. 22,218 patch triplets of corresponding Sentinel-1 dual-pol SAR data, Sentinel-2 multi-spectral images, and cloud-covered Sentinel-2 multi-spectral images
- Sen2Like
- SEN2NAIP - Remote sensing dataset designed to support conventional and reference-based SR model training
- Sentinel Hub NoR Sponsored Accounts and Data Collections

- Sentinel Satellite Data
- Sentinel-5P
- Sentinel-2 data set for the delineation of agricultural field boundaries in Flevoland, The Netherlands
- Síntesis de Indicadores e Informes Macroeconómicos
- SkyFi Geospatial Hub
- SkySat missions
- Socioeconomic Data and Applications Center (SEDAC)Link1 y Link2
- Some datasets for teaching data science
- Source Cooperative Featured Repositories
- STAC Index SpatioTemporal Asset Catalog (STAC)
- StatsBomb sports data
- TanDEM-X 90m DEM (DLR) - Global DEM generated from radar data
- Teaching of Statistics in the Health Sciences
- Tematicas.org Recopilación de series e índices
- Terra Populus
- Terraclimate - Monthly climate and hydrology data at a global scale.
- The Big Bad NLP Database
- The Government Finance Database
- The SpaceNet Datasets
- The World Bank Open Knowledge Repository
- The world's economic database
- TidyRainbow
- TidyTuesday
- Tráfico en el Reino Unido
- UC Irvine Machine Learning Repository
- UC Merced Land Use Dataset
- UCI Machine Learning Repository
- UK Data Service
- UK Office for National Statistics
- UK Open Data
- UK Open Geography Portal
- Ultimos datos de Open Street Map. Spain
- Una recopilación de APIs públicas
- Una recopilación de datasets públicos
- Understat
- UNEP Environmental Data Explorer
- United Nations Platform for Space-based Information for Disaster Management and Emergency Response (un-spider.org) data sources
- United Nations World Urbanization Prospects
- Universidad de Harvard
- US Homeland Infrastructure Foundation-Level Data
- USGS 3DEP LiDAR Point Clouds
- USGS Earth Explorer - SRTM, ASTER GDEM, ALOS, and more
- Viewfinder Panoramas - High-quality DEM for remote regions
- WHU-RS19 is a set of satellite images exported from Google Earth
- Wyvern Open Data Program
- World Economic Forum
- WorldCereal open global harmonized reference data repository [Data]](https://zenodo.org/records/7609500) Github
- Worldpop - Open Spatial Demographic Data y Worldpop Hub
- Yelp Dataset
- Zhu Lab - Data Science in Earth Observation
- Amazon AWS: este y este
- EarthNets for Earth Observation Page Github

- Facebook Neural-Code-Search-Evaluation-Dataset [dataset]](https://github.com/facebookresearch/Neural-Code-Search-Evaluation-Dataset) y noticia
- HREA: High Resolution Electricity Access. Universidad de Michigan y Microsoft
- IPUMS provides census and survey data from around the world Web y paquete ipumsr
- Maxar Open Data: Aquí y aquí
- MIT 1 y 2
- Natural Earth Vector. Github y Web
- Open Charge Map. Global Open Data registry of electric vehicle charging locations. Export y Ejemplo
- SSL4EO-S12 dataset. Large-scale multimodal multitemporal dataset for unsupervised/self-supervised pre-training in Earth observation Paper Github
- World Bank Open Data 1 y 2

Otras referencias interesantes

- 100 Active Blogs on Analytics, Big Data, Data Mining, Data Science, Machine Learning
- 100 Free Tutorials for Learning R
- 16 Cursos
- A dive into R Markdown
- A ggplot2 Tutorial for Beautiful Plotting in R
- AiTLAS: Benchmark Arena – Open-source benchmark suite for evaluating deep learning approaches for image classification in Earth Observation (EO)
- An Introduction to Statistical Learning - Web R & Python
- ArcGIS to R spatial cheat sheet
- Awesome Data Science
- Awesome R
- BigEarthNet A Large-Scale Sentinel Benchmark Archive
- Bivariate Choropleth Maps: A How-to Guide
- blogdown: Creating Websites with R Markdown
- Blogs con github y Blogs con github y RStudio
- CAMIS - A PHUSE DVOST Working Group. The repository below provides examples of statistical methodology in different software and languages, along with a comparison of the results obtained and description of any discrepancies.
- Chuleta de expresiones regulares
- Chuleta general de R
- Codificación de caracteres
- Common Probability Distributions: The Data Scientist's Crib Sheet
- Cómo crear una API en Python
- Computer vision
- Computerworld - Paquetes de R interesantes
- Curso Caltech. Learning from data
- Cursos para aprender más sobre R
- Data Science Blogs
- Data Science Cheatsheets
- Data Science Collected Resources
- Data Science Resources
- Data Scientist Roadmap
- Data Viz Catalogue
- Dataviz Project
- Dealing with Regular Expressions
- Ejemplos de Shiny
- Estadística con R
- EUMETSAT science studies
- Feature Engineering for Machine Learning
- Financial-Times / chart-doctor

- Formatos a medida para R Markdown
- Free R Reading Material
- From Data to Viz
- Galerias de graficos
- Ggplot
- GIS and mapping
- GIS formats
- Glosario de Machine Learning de Google
- Google Dataset Search
- Google Rules of Machine Learning: Best Practices for ML Engineering
- Google's best practices in machine learning
- HDRIs Images (HDRIs))
- HOT - Drone Tasking Manager
- htmlwidgets for R - gallery
- IDEAtlas. Developing AI-based methods to map and characterize informal settlements from Earth Observation data
- Información de Rmarkdown en R Studio
- Information is Beautiful Awards
- Information is beautiful
- Information is Beautiful
- Interactive 4D LiDAR Segmentation
- Investigative Journalism with Satellite Images
- Kaggle Winning Solutions
- Microsoft Presidio - Data Protection and De-identification SDK
- Naming files
- Otra lista de recursos variados en Github
- overpass turbo - Herremaienta de filtrado para OSM
- Pandoc User's Guide
- Periodic Table Of Visualization Methods
- Plataforma H2O
- Practical Introduction to Web Scraping in R
- R Code – Best practices
- R Coding Style Guide
- R Data Science Tutorials
- R for Water Resources Data Science
- R Learning Path: From beginner to expert in R in 7 steps
- R Markdown cheatsheet
- R Markdown referencia
- R package primer
- R Universe search
- RDocumentation
- Regular Expression Language - Quick Reference
- Regular Expressions Every R programmer Should Know
- Remote Sensing for OSINT
- Remote sensing image retrieval
- RMarkdown Driven Development (RmdDD)
- rseek.org - rstats search engine
- Rstudio cheatsheets
- Simplifying the ROC and AUC metrics
- Soporte técnico de RStudio
- Study finds 94% of business spreadsheets have critical errors
- Template para documentos científicos con Rmarkdown
- The Chartmaker Directory
- The Data Visualisation Catalogue

- The R Graph Gallery
- The State of Naming Conventions in R
- The TimeViz Browser 2.0
- Tipos de licencias de software
- Tipos de licencias open data (minicurso de data.europa.edu)
- Tutorials for learning R
- UK government using R to modernize reporting of official statistics
- Usar git
- useR! Machine Learning Tutorial
- Using Geospatial Data in R
- Utilizando Sweave y Knitr
- Writing an R package from scratch
- Global Fishing Watch. AI and satellite imagery to reveal the expanding footprint of human activity at sea. Post. Github. Train data. Analysis data and Vessel detection from Sentinel-1 SAR
- Legalidad Web sraping: Is Web Scraping Legal? : The Definitive Guide (2024 update) y Web Scraping: ¿legal o ilegal?
- Pautas para dar formato al código programando en R: Google, Hadley Wickham (RStudio) y Coding Club
- Sistemas de Coordenadas. Aqui y aqui
- Statistical Learning de Stanford with R Curso, Libro, Código y Transparencias

Libros

- 10 Free Must-Read Books for Machine Learning and Data Science
- Advanced R
- Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA
- AI With R
- An Introduction to R
- An Introduction to Spatial Data Analysis and Visualisation in R
- An R companion to Statistics: data analysis and modelling
- Análisis de datos y algoritmos de predicción con R
- Aprendiendo R sin morir en el intento
- Aprendizaje Estadístico con R
- Bayesian inference with INLA
- BBC Visual and Data Journalism cookbook for R graphics
- Big Book of R
- Bioinformática Estadística. Análisis estadístico de datos Ómicos
- Building reproducible analytical pipelines with R
- Command Line Basics for R Users
- Creating APIs in R with Plumber
- Data Analysis and Prediction Algorithms with R
- Data Management in Large-Scale Education Research
- Data Science in Education Using R
- Data Skills for Reproducible Science
- Data Visualization with R
- Databases using R by RStudio
- Dendrometria
- Deep Learning and Scientific Computing with R torch
- Deep Learning
- Econometrics with the Tidyverse
- Efficient R programming
- Efficient Machine Learning with R
- Elegant and informative maps with tmap
- Engineering Production-Grade Shiny Apps

- Estadística básica
- Estilometría, análisis de textos en R para filólogos
- Exploring Complex Survey Data Analysis Using R
- Exploratory Data Analysis with R - Roger D. Peng
- Forecasting: Principles and Practice
- Feature Engineering A-Z
- Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny
- Handbook of Graphs and Networks in People Analytics With Examples in R and Python
- Handbook of Regression Modeling in People Analytics
- Handling Strings with R
- Hands-On Data Visualization
- Hands-On Machine Learning with R
- Hands-On Programming with R
- Happy Git and GitHub for the useR
- Interpretable Machine Learning
- Introducción a R
- Introduction to Econometrics with R
- Introduction to Probability for Data Science
- Introduction to urban accessibility: a practical guide in R
- JavaScript for R
- Large Language Model tools for R
- Learning Statistics with R
- Libro Vivo de Ciencia de Datos
- Linear Algebra for Data Science
- Model to Meaning
- Modern R with the tidyverse
- NASA Earthdata Cloud Cookbook
- Officeverse R & Office
- Open Source Technology in Clinical Data Analysis
- Outstanding User Interfaces with Shiny
- Predictive Soil Mapping with R
- Probabilidad básica
- Quantitative Politics with R
- R Advanced Spatial Lessons
- R for Data Analysis
- R for data science: tidyverse and beyond
- R for everyone
- R for Health Data Science
- R Graphics Cookbook
- R in action
- R intro
- R Markdown Cookbook
- R Markdown: The Definitive Guide
- R Notes for Professionals
- R Packages
- R para principiantes
- R para profesionales de los datos: una introducción
- R Programming for Data Science. Roger D. Peng.
- R Programming for Data Science
- R4JournalismBook
- rstudio4edu
- Simulación Estadística con R
- Spatial Analysis With R
- Spatial Data Science with applications in R

- [Spatial Data Science](#)
- [Spatial Microsimulation with R](#)
- [Spatial Modelling for Data Scientists](#)
- [Statistical Inference via Data Science](#)
- [Supervised Machine Learning for Text Analysis in R](#)
- [Technical Foundations of Informatics](#)
- [Text Mining with R](#)
- [The 20 Best Data Science Books Available online in 2020](#)
- [The Art of Data Science](#)
- [The caret Package](#)
- [The Epidemiologist R Handbook](#)
- [The R Book](#)
- [The Shiny AWS Book](#)
- [Think Bayes 2e](#)
- [Tidy Finance with R](#)
- [Tidy Finance](#)
- [Todos los libros en bookdown](#)
- [Twitter for Scientists](#)
- [What They Forgot to Teach You About R](#)
- [YaRrr! The Pirate's Guide to R](#)
- Applied Statistics with R [Libro](#) y [Código](#)
- Data Science Live Book [Libro](#) y [Código](#)
- Fundamentals of Data Visualization [Libro](#) y [Código](#)
- Geocomputation with R [Libro](#) y [Código](#)
- Introduction to Data Science [Libro](#) y [Código](#)
- Mastering Apache Spark with R [Libro](#) y [Código](#)
- R for Data Science. [Inglés](#) y [Castellano](#)
- R for Statistical Learning [Libro](#) y [Código](#)
- sits: Satellite Image Time Series Analysis on Earth Observation Data Cubes [Libro](#) y [Kaggle](#)

Revisar los links

Dentro del repositorio, se ha creado un archivo `revisar_links.R` para revisar si los links son válidos. Para que sea mas fácil su uso, recopila los links del repositorio público de [Open Data](#) en el fichero [README](#), pero el código se puede modificar con la variable `repo_url`.