Recommended Datasets for Mini Projects

Hamed Ahmadinia, Ph.D.

Hamed.Ahmadinia@metropolia.fi

≡ Table of Contents

Quick Jump to Sections

- > Agriculture
- > Energy
- > Neuroscience

- > Architecture
- > Entertainment
- > Physics

- > Biology
- > Finance
- ➤ ProstateCancer

- > Chemistry > Climate
- Datasets
- > Finnish/European > Psychology
- ➤ ComplexNetworks ➤ General Cross-
- > PublicDomains

- > ComputerNetworks
- Domain
- > SearchEngines

- > Cryptocurrency
- > GIS > Government
- > SocialNetworks

- Datasets
- > SocialSciences

- > CyberSecurity ➤ DataChallenges
- > Healthcare
- > Software > Sports

- > EarthScience
- **➤** ImageProcessing > MachineLearning > TimeSeries

- > Economics
- > Museums
- > Transportation

- **>** Education
- ➤ NaturalLanguage ➤ eSports

Agriculture

The global dataset of historical yields for major crops (1981-2016)

Link

Historical yields data for major crops around the world.

Hyperspectral benchmark dataset on soil moisture Link

Five-day measurement dataset for soil moisture analysis using hyperspectral imaging.

Lemons quality control dataset

Prepared to investigate the quality control of lemons for research and analysis.

Optimized Soil Adjusted Vegetation Index

Remote sensing tool for analyzing soil-vegetation interaction and reflectance.

U.S. Department of Agriculture's Nutrient Database

Extensive database of nutrient composition for a wide range of foods.

U.S. Department of Agriculture's PLANTS Database

Contains a nearly comprehensive PLANTS Checklist with data on U.S. flora.

Architecture

Swiss Apartment Models

Link

Contains detailed data on 42,207 apartments (242,257 records).

ArchNet

link

An open-access digital library focused on architecture, urbanism, environmental and landscape design, visual culture, and conservation issues related to the Muslim world.

ArchINFORM

Link

An international architecture database documenting architectural projects from the 20th century to contemporary works.

OpenStreetMap (OSM)

A collaborative project creating a free editable map of the world, valuable for architectural studies, urban planning, and infrastructure development.

CityGML

Link

An open standardized data model and exchange format to store digital 3D models of cities and landscapes, aiding urban planning and architecture.

NOAA Observing System Architecture (NOSA)

A collection of over 100 environmental datasets by NOAA, relevant to architectural projects concerning environmental impact and sustainability.

▼ Biology

1000 Genomes

The 1000 Genomes Project ran between 2008 and 2015, creating the largest [...]

ANHIR

Automatic Non-rigid Histological Image Registration (AN-HIR) consists of 2D [...]

American Gut (Microbiome Project)

The American Gut project is the largest crowdsourced [...]

BCNB

Link

There are WSIs of 1058 patients; part of tumor regions are annotated in WSIs. Except [...]

Broad Bioimage Benchmark Collection (BBBC)

Link

The Broad Bioimage Benchmark Collection (BBBC) [...]

Broad Cancer Cell Line Encyclopedia (CCLE)

link

Data for the Broad Cancer Cell Line Encyclopedia (CCLE).

CIMA

Link

CIMA dataset includes images of 2D histological microscopy tissue slices.

Cell Image Library

Link

This library is a public and easily accessible resource database of [...]

Complete Genomics Public Data

link

A diverse data set of whole human genomes is freely [...]

${\bf CytoImageNet}$

Link

A large-scale dataset of microscopy images. Contains 890,737 total grayscale [...]

EBI ArrayExpress

Link

Array Express Archive of Functional Genomics Data stores data from high- [...]

EBI Protein Data Bank in Europe

Link

The Electron Microscopy Data Bank (EMDB) is a public [...]

ENCODE project

Link

The Encyclopedia of DNA Elements (ENCODE) Consortium is an ongoing $[\ldots]$

Electron Microscopy Pilot Image Archive (EMPIAR)

EMPIAR, the Electron Microscopy Public [...]

Ensembl Genomes

Link

Gene Expression Omnibus (GEO)

Link

GEO is a public functional genomics data repository [...]

Gene Ontology (GO)

Link

Global Biotic Interactions (GloBI)

Link

Harvard Medical School (HMS) LINCS Project

Link

The Harvard Medical School (HMS) LINCS Center is [...]

Human Genome Diversity Project

Link

A group of scientists at Stanford University have [...]

Human Microbiome Project (HMP)

Link

The HMP sequenced over 2000 reference genomes isolated from $[\ldots]$

ICOS PSP Benchmark

Link

The ICOS PSP benchmarks repository contains an adjustable real-world [...]

International HapMap Project

Link

Journal of Cell Biology DataViewer

Link

KEGG

Link

KEGG is a database resource for understanding high-level functions and utilities [...]

NCBI Proteins

Link

NCBI Taxonomy

Link

The NCBI Taxonomy database is a curated set of names and classifications for [...]

NCI Genomic Data Commons

Link

The GDC Data Portal is a robust data-driven platform that allows $[\ldots]$

NIH Microarray data

Link

OpenSNP genotypes data

Link

open SNP allows customers of direct-to-customer genetic tests to $[\ldots]$

Palmer Penguins

Link

The goal of palmer penguins is to provide a great dataset for data $[\ldots]$

Pathguid - Protein-Protein Interactions Catalog

Link

Protein Data Bank

Link

This resource is powered by the Protein Data Bank archive-information $[\ldots]$

Psychiatric Genomics Consortium

Link

The purpose of the Psychiatric Genomics Consortium (PGC) is [...]

PubChem Project

link

PubChem is the world's largest collection of freely accessible chemical [...]

PubGene (now Coremine Medical)

Link

COREMINE $^{\!\!\top\!\!\!M}$ is a family of tools developed by the Norwegian $[\ldots]$

Sanger Catalogue of Somatic Mutations in Cancer (COSMIC)

Link

COSMIC, the Catalogue Of Somatic [...]

Sanger Genomics of Drug Sensitivity in Cancer Project (GDSC)

Link

Sequence Read Archive (SRA)

Link

The Sequence Read Archive (SRA) stores raw sequence data from [...]

Serratus

Link

Analysis of 7.1 million RNA/DNA sequencing datasets to discover the total [...]

Stanford Microarray Data (Retired NOW)

Link

Stowers Institute Original Data Repository

Link

Systems Science of Biological Dynamics (SSBD) Database

Link

Systems Science of Biological [...]

The Cancer Genome Atlas (TCGA), available via Broad GDAC

Link

The Catalogue of Life

Link

The Catalogue of Life is a quality-assured checklist of more than 1.8 $[\ldots]$

The Personal Genome Project

Link

UCSC Public Data

Link

UniGene

Link

Universal Protein Resource (UniProt)

Link

The Universal Protein Resource (UniProt) is a [...]

Rfam

Link

The Rfam database is a collection of RNA families, each represented by multiple [...]

■ Chemistry

Ionic Liquids Database - ILThermo

Link

Data on properties of ionic liquids.

🔼 Climate

Actuaries Climate Index

Link

Australian Weather

Link

Aviation Weather Center

Link

Consistent, timely, and accurate weather information for the world.

Brazilian Weather (Historical data, in Portuguese)

Link

Data related to climate and weather in Brazil.

Several Climate Datasets (C3S Climate Data Store)

Link

The C3S Climate Data Store (CDS) is a one-stop shop for [...].

Canadian Meteorological Centre

Link

Caravan (Large-sample Hydrology)

Link

Caravan is an open community dataset of [...].

Climate Data from UEA (updated monthly)

Dutch Weather

Link

The KNMI Data Center (KDC) portal provides access to KNMI data on weather [...].

Washington Post Climate Change

To analyze warming temperatures in the United States, The Washington Post [...].

European Climate Assessment & Dataset

WorldClim - Global Climate Data

Link

German Climate Data Center

Complex Networks

A Miner Citation Network Dataset

Link

Global Climate Data Since 1929

Link

CrossRef DOI URLs

Link

Charting the Global Climate Change News Narrative (2009-2020)

Link

DBLP Citation dataset

link

NASA Global Imagery Browse Services

Link

DIMACS Road Networks Collection

Link

NOAA Bering Sea Climate

NBER Patent Citations

Link

NOAA Climate Datasets

Link

NIST complex networks data collection

Link

NOAA Realtime Weather Models

Link

Network Repository (Interactive Exploratory Analysis Tools)

Link

NOAA SURFRAD Meteorology and Radiation **Datasets**

Link

Protein-protein interaction network

Link

Open-Meteo (Open-Source Weather API)

Open-source weather API with free access for non-commercial

PyPI and Maven Dependency Network

The World Bank Open Data Resources for Climate

Scopus Citation Database

Link

Change

Link

Small Network Data

Link

UEA Climatic Research Unit

Link

Stanford GraphBase

Link

WU Historical Weather Worldwide

Link

Stanford Large Network Dataset Collection Link

Stanford Longitudinal Network Data Sources

ink

Criteo click-through data

Link

The Koblenz Network Collection

Link

Internet-Wide Scan Data Repository

Link

The Laboratory for Web Algorithmics (UNIMI)

Link

MIRAGE-2019

Link

UCI Network Data Repository

Link

Merklemap DNS records Dataset

Link

Contains 4B+ DNS records across 700 million unique [...].

UFL sparse matrix collection

Link

OONI: Open Observatory of Network Interference

Link

Internet censorship data.

WSU Graph Database

Link

Open Mobile Data by MobiPerf

Link

Community Resource for Archiving Wireless Data At Dartmouth

Link

The Peer-to-Peer Trace Archive

Link

Real-world measurements play a key role in studying the [...].

Computer Networks

3.5B Web Pages from CommonCrawl 2012

Link

Rapid7 Sonar Internet Scans

Link

53.5B Web clicks of 100K users in Indiana Univ. Link

UCSD Network Telescope, IPv4 /8 net Link

CAIDA Internet Datasets

Link

B Cryptocurrency Datasets

CRAWDAD Wireless datasets from Dartmouth Univ. Link

ClueWeb09 - 1B web pages

Link

 ${\bf Crypto Data Download}$

Link

Provides free historical cryptocurrency data in CSV format, organized by exchange, symbol, and timeframe. Data includes spot, futures, and options assets across various timeframes.

ClueWeb12 - 733M web pages

Link

Kaggle - Top 3000+ Cryptocurrency Dataset

Link

A comprehensive dataset covering various cryptocurrencies, available for download in CSV format.

CommonCrawl Web Data over 7 years

Link

Kaggle - Cryptocurrency Historical Prices

Link

Historical price data for multiple cryptocurrencies, with each cryptocurrency's data provided in a separate CSV file.

Shopper Intent Prediction from Clickstream E-Commerce Data

Link

with Minimal Browsing Information.

Cryptocurrency Alerting

Link

Offers CSV and JSON datasets of historical exchange listings dating back to 2018, along with curated coin and token metadata.

Mendeley Data - Cryptocurrency Historical Data

Provides integrated cryptocurrency historical data for 80 cryptocurrencies, including Bitcoin, Ethereum, and Binance Coin, from January 2012 to October 2021.

GitHub - Bitcoin Dataset

Link

A dataset available on GitHub containing Bitcoin data suitable for supervised machine learning tasks, provided in CSV format.

Cyber Security

CCCS-CIC-AndMal-2020

Link

The dataset includes 200K benign and 200K malware samples $\left[\ldots \right]$

Traffic and Log Data Captured During a Cyber Defense Exercise

Link

This dataset was acquired during a Cyber Defense Exercise $[\ldots]$

Data Challenges

AIcrowd Competitions

Link

Bruteforce Database

Link

Challenges in Machine Learning

Link

CrowdANALYTIX dataX

Link

D4D Challenge of Orange

Link

DrivenData Competitions for Social Good

Link

ICWSM Data Challenge (since 2009)

Link

KDD Cup by Tencent 2012

Link

Kaggle Competition Data

Link

Localytics Data Visualization Challenge

Link

Netflix Prize

Link

Space Apps Challenge

امنا

Telecom Italia Big Data Challenge

Link

${\bf TravisTorrent~Dataset~-~MSR'2017~Mining~Challenge}$

Link

TunedIT

Link

Data mining & machine learning datasets, algorithms, and challenges. $\,$

Yelp Dataset Challenge

Link

The Yelp dataset is a subset of businesses, reviews, and user $[\ldots]$

Earth Science

38-Cloud (Cloud Detection)

Lin

Contains 38 Landsat 8 scene images and their manually extracted $[\ldots]$

$\mathbf{AQUASTAT}$ - Global water resources and uses Link

BODC - marine data of ~22K vars

Link

EOSDIS - NASA's earth observing system data

Link

Earth Models

Link

Global Wind Atlas

Link

A free, web-based application developed to help [...].

Integrated Marine Observing System (IMOS)

Link

Roughly 30TB of ocean measurements.

Marinexplore - Open Oceanographic Data

Link

Alabama Real-Time Coastal Observing System

Lin

National Estuarine Research Reserves System-Wide Monitoring Program

link

Long-term estuarine observations.

Oil and Gas Authority Open Data

Link

Covers 12,500 offshore wellbores, 5,000 seismic [...].

Smithsonian Institution Global Volcano and Eruption Database

Link

USGS Earthquake Archives

link

Wellhead Protection Area (WHPA) prediction using breakthrough curves

Link

This dataset [...].

Economics

Asian Productivity Organization (APO)

Link

The AEPM provides a graphic dashboard view of [...].

ASEAN Stats

Link

The ASEANstatsDataPortal was first launched in June 2018.

American Economic Association (AEA)

Link

Asian KLEMS

Link

Asia KLEMS is an Asian regional research consortium to promote building [...].

Harvard Atlas of Economic Complexity

Link

A database for exploring global trade flows [...].

BIS Financial Database

Link

Contains the same data as in the BIS Statistics Explorer [...].

Barro-Lee Education Attainment

Link

Barro-Lee Educational Attainment Data from 1950 to 2010.

CEPII Database

Link

A database of the world economy, through its country and region profiles [...].

EUKLEMS

Link

An industry-level growth and productivity research project.

Economic Freedom of the World Data

Link

Historical National Accounts

Link

Comparative Historical National Accounts [...].

Historical MacroEconomic Statistics

Link

INFORUM - Interindustry Forecasting at the University of Maryland

Link

DBnomics - The World's Economic Database

Link

Aggregates hundreds of millions of time series [...].

International Trade Statistics

Link

Internet Product Code Database

Link

Joint External Debt Data Hub

Link

Jon Haveman International Trade Data Links

Link

Latin America KLEMS

Link

A technical cooperation project financed by the Inter-[...].

Long-Term Productivity Database

Link

Created as a [...].

Maddison Project Database

Link

Provides information on comparative [...].

National Transfer Accounts (NTA)

Link

The goal of the National Transfer Accounts (NTA) project is [...].

OpenCorporates Database of Companies in the World Link

Our World in Data

Link

Penn World Table

Link

PWT version 10.0 is a database with information on relative levels of $[\ldots]$.

SciencesPo World Trade Gravity Datasets

Link

The Atlas of Economic Complexity

The Center for International Data

Link

The Observatory of Economic Complexity

Link

UN Commodity Trade Statistics

Link

UN Human Development Reports

Link

World Input-Output Database

Link

Covers 43 [...] with underlying data.

World KLEMS

Link

Analytical KLEMS-type data sets for a broad set of countries.

🞓 Education

College Scorecard Data

Link

New York State Education Department Data

Link

The New York State Education Department (NYSED) is [...].

Program for International Student Assessment (PISA)

Link

Contains 15-year-old students' [...].

Student Data from Free Code Camp

Link

Energy

AMPds - The Almanac of Minutely Power dataset

Link

BLUEd - Building-Level Fully Labeled Electricity

Disaggregation dataset

Link

COMBED

Link

DBFC - Direct Borohydride Fuel Cell (DBFC)

Dataset

Link

DEL - Domestic Electrical Load study datasets for South Africa (1994 - 2014)

Link

ECO

Link

A comprehensive dataset for non-intrusive load monitoring [...].

EIA

Link

Global Power Plant Database

Link

A comprehensive, open source database of power plants [...].

HES - Household Electricity Study, UK

Link

HFED

Link

MORED: Moroccan Buildings' Electricity Consumption Dataset

Link

Data gathered since spring of 2019.

Marktstammdatenregister (MaStR)

Link

A German database of all energy generation plants [...].

PEM1 - Proton Exchange Membrane Fuel Cell Dataset

Link

PLAID - The Plug Load Appliance Identification Dataset

Link

The Public Utility Data Liberation Project (PUDL)

Link

Makes US energy data easier to [...].

REDD

Link

SYND - A synthetic energy dataset for non-intrusive load monitoring

Link

We present a synthetic energy dataset, SynD, for [...].

Smart Meter Data Portal

Link

A part of the National Science [...].

Tracebase

Link

OANDA

Link

Ukraine Energy Centre Datasets

Link

OSU Financial data

Link

UK-DALE - UK Domestic Appliance-Level Electricity

Quandl

link

WHITED

link

SEC EDGAR

Link

EDGAR is the Electronic Data Gathering, Analysis, and

Retrieval system [...].

iAWE Link

St Louis Federal

Link

Entertainment

Top Streamers on Twitch

Contains data of the Top 1000 Streamers from the past year.

Yahoo Finance

link

Finance

BIS Statistics

BIS statistics, compiled in cooperation with central banks and

Finnish/European Datasets

Finnish National Election Studies (FNES)

link

Voter behavior, political attitudes, and democracy surveys.

Blockmodo Coin Registry

A registry of JSON formatted information files primarily [...].

Kansalaismielipide Project

Citizen opinions on governance and public services in Finnish municipalities.

CBOE Futures Exchange

Link

European Values Study (EVS)

Cultural, moral, and political values across Europe.

Complete FAANG Stock data

This dataset contains all the stock data of FAANG companies from [...].

Eurobarometer

EU-wide public opinion on policies and societal issues.

Google Finance

Link

European Social Survey (ESS)

Longitudinal data on attitudes and behavior across Europe.

Google Trends

Link

European Data Portal

Provides access to over 1.6 million datasets from various European countries, including Finland, covering diverse sectors such as agriculture, education, and transportation.

NASDAQ

Link

Europeana

A digital platform providing access to millions of digitized items from European museums, galleries, libraries, and archives, including Finnish cultural heritage materials.

NYSE Market Data

Publications Office of the European Union

Link

Provides access to EU law, publications, open data, research results, procurement notices, and other official information, encompassing datasets relevant to Finland.

Finnish Social Science Data Archive (FSD)

link

Offers a wide range of digital research data for teaching and research purposes, including surveys, interviews, and statistical data related to Finnish society.

Finnish Meteorological Institute (FMI) Open Data

Provides open access to weather, sea, and climate data in Finland, including real-time observations and historical data.

Natural Resources Institute Finland (Luke) Open Data

Link

Offers datasets related to agriculture, forestry, fisheries, and the environment in Finland.

Finnish Transport Infrastructure Agency (Väylä) Open Data

Link

Provides datasets related to Finland's transport infrastructure, including roads, railways, and waterways.

Finnish Environment Institute (SYKE) Open Data

Offers environmental datasets, including information on water resources, biodiversity, and land use in Finland.

Avoindata.fi

Link

Finland's central open data portal (transport, environment, etc.).

Statistics Finland (Tilastokeskus)

Link

Demographics, economy, education, and social trends.

Finnish Institute for Health and Welfare (THL)

Link

Health surveys and infectious disease data.

Finnish Patent and Registration Office (PRH)

Link

Company registrations and financial statements.

Bank of Finland

Link

Financial market statistics and macroeconomic indicators.

♠ General Cross-Domain Sources

Kaggle

Link

Large variety of datasets spanning finance, healthcare, marketing, and more.

Google Dataset Search

Link

Meta-search engine that locates datasets across multiple disciplines.

UCI Machine Learning Repository

Link

Classic and widely used machine learning datasets for benchmarking algorithms.

Zenodo

Link

Multidisciplinary research data (often from EU-funded projects).

GIS

Awesome 3D Semantic City Models

Link

Collection of open 3D semantic city and region models.

ArcGIS Open Data portal

Link

Cambridge, MA, US, GIS data on GitHub

Link

Database of all continents, countries, States/Subdivisions/Provinces and Cities

Link

Database [...] in JSON format.

Factual Global Location Data

Link

IEEE Geoscience and Remote Sensing Society DASE Website

Link

Geo Maps

Link

High Quality GeoJSON maps generated programmatically.

Geo Spatial Data from ASU

Link

Geo Wiki Project - Citizen-driven Environmental Monitoring

Link

GeoFabrik

Link

OSM data extracted to a variety of formats and areas.

GeoNames Worldwide TwoFishes - Foursquare's coarse geocoder Link Link Global Administrative Areas Database (GADM) **UN Environmental Data** Link Geospatial data organized by country, includes multiple levels of subdivisions. World boundaries from the U.S. Department of State Homeland Infrastructure Foundation-Level Data Link World countries in multiple formats Link Landsat 8 on AWS Link **H** Government List of all countries in all languages Alberta, Province of Canada Link Link National Weather Service GIS Data Portal Antwerp, Belgium Link Natural Earth Argentina (non official) link Link Vectors and rasters of the world. **Datos Argentina** OpenAddresses Link Portal de datos abiertos de la República Argentina. OpenStreetMap (OSM) Austin, TX, US Link Pleiades - Gazetteer and graph of ancient places Australia (abs.gov.au) Link Link Reverse Geocoder using OSM data Australia (data.gov.au) link Link Robin Wilson - Free GIS Datasets Austria (data.gv.at) Link Link Shadow Accrual Maps Baton Rouge, LA, US Link Accumulated shadow information for New York [...]. Beersheba, Israel - Open Data Portal TIGER/Line - U.S. boundaries and roads Link Link

TZ Timezones shapefile

Link

Belgium

City of Berkeley Open Data ${\bf EuroStat}$ Link Link Brazil **EveryPolitician** Link Collating and sharing data on every politician. Buenos Aires, Argentina Link Federal Committee on Statistical Methodology (FCSM) Link Formerly FedStats. Calgary, AB, Canada Finland Cambridge, MA, US Link Link France Canada Link Link Fredericton, NB, Canada Chicago Link Gatineau, QC, Canada Chile Link Germany China Link Link Ghent, Belgium Dallas Open Data Link DataBC - Province of British Columbia Glasgow, Scotland, UK Link Link Debt to the Penny Greece Link Information about total public debt outstanding. Guardian world governments Denver Open Data Link Link Halifax, NS, Canada Durham, NC Open Data Link Link Helsinki Region, Finland Edmonton, AB, Canada Link Link **England LGInform** Hong Kong, China Link Link

Houston, TX, US Luxembourg Link Luxembourgish Open Data Portal. **Indian Government Data** Link MassGIS, Massachusetts, U.S. Indonesian Data Portal Link Metropolitan Transportation Commission (MTC), California, US Iowa Link Explore data about Iowa. Mexico Link Ireland's Open Data Portal Link Mississauga, ON, Canada Link Israel's Open Data Portal Link Moldova Link Istanbul Municipality Open Data Portal Link Moncton, NB, Canada Link Italy Link Montreal, QC, Canada Il Portale dati.gov.it è il catalogo nazionale dei metadati. Link Jail deaths in America Mountain View, California, US (GIS) Link The U.S. government does not release jail-by-jail mortality data. NYC Open Data Link Japan Link NYC betanyc Link Laval, QC, Canada Link Netherlands Link Lexington, KY Link New York Department of Sanitation Monthly Tonnage Link London Datastore, UK DSNY Monthly Tonnage Data provides [...]. Link New Zealand London, ON, Canada Link Link **OECD** Los Angeles Open Data Link Link

Oakland, California, US Rio de Janeiro, Brazil Link Link Romania Oklahoma Link Link Russia Open Data for Africa Link Link San Diego, CA Open Government Data (OGD) Platform India Link San Antonio, TX - Community Information Now OpenDataSoft's list of 1,600 open data A nonprofit serving Bexar (San Antonio) [...]. Link San Francisco Data sets Oregon Link Link San Jose, California, US Ottawa, ON, Canada Link Link San Mateo County, California, US Link Palo Alto, California, US Link Saskatchewan, Province of Canada Link OpenDataPhilly A catalog of open data in the Philadelphia region. Seattle Link Portland, Oregon Link Singapore Government Data Portugal - Pordata organization Link South Africa Trade Statistics Link Puerto Rico Government South Africa Link Link Quebec City, QC, Canada State of Utah, US Link Link Quebec Province of Canada Switzerland Link Link Regina SK, Canada Taiwan gov Link Link

U.S. Open Government Link Link UK 2011 Census Open Atlas Project Tel-Aviv Open Data Link Link Texas Open Data **US** Counties Link A repository of various data, broken down by US county. The World Bank Link U.S. Patent and Trademark Office (USPTO) Bulk **Data Products** link Toronto, ON, Canada Link Uganda Bureau of Statistics Link Tunisia Link Ukraine Link U.K. Government Data link **United Nations** link U.S. American Community Survey Link Uruguay Link U.S. CDC Public Health datasets Link Valley Transportation Authority (VTA), California, USLink U.S. Census Bureau Link Vancouver, BC Open Data Catalog Link U.S. Department of Housing and Urban Development (HUD) Link Victoria, BC, Canada Link U.S. Federal Government Agencies Link Vienna, Austria Link U.S. Federal Government Data Catalog Link Statistics from the General Statistics Office of Vietnam Link U.S. Food and Drug Administration (FDA) Data in different categories are [...]. Link

Link

U.S. National Center for Education Statistics (NCES)

Taiwan

Link

U.S. Congressional Research Service (CRS) Reports

🔁 Healthcare

AWS COVID-19 Datasets

Working with organizations to provide COVID-19-related data [...].

COVID-19 Case Surveillance Public Use Data

Link

The COVID-19 case surveillance system database [...].

Covid-19 non-processed data of Ecuador

Link

Non-processed datasets and analysis for COVID-19 in Ecuador.

2019 Novel Coronavirus COVID-19 Data by Johns Hopkins CSSE

This is the data repository for the 2019 Novel Coronavirus [...].

Coronavirus (Covid-19) Data in the United States

The New York Times is releasing a series of COVID-19 data [...].

COVID-19 Reported Patient Impact and Hospital Capacity by Facility

Link

Composition of Foods (USDA National Nutrient Database)

Link

Raw, processed, prepared USDA National Nutrient Database $[\ldots].$

The COVID Tracking Project

Link

Collects and publishes the most complete testing data [...].

EHDP Large Health Data Sets

Link

GDC - Cancer Genome Programs (CCG, TCGA, TARGET)

Supports several cancer genome programs.

Gapminder World Demographic Databases

link

MeSH (vocabulary thesaurus for PubMed)

Used for indexing articles for PubMed.

MeDAL - Medical Abbreviation Disambiguation Dataset

Link

A large medical text dataset curated for abbreviation disambiguation.

Medicare Coverage Database (MCD), U.S.

link

Medicare Data Engine of medicare.gov

Link

Medicare Data File

Link

Nightingale Open Science

Link

Number of Ebola Cases and Deaths in Affected Countries (2014)

Link

Open-ODS (Structure of the UK NHS)

OpenPaymentsData (Healthcare financial relationships)

Link

PhysioBank Databases

Link

A large archive of physiological data.

The Cancer Imaging Archive (TCIA)

Link

The Cancer Genome Atlas project (TCGA)

World Health Organization Global Health Observatory

Link

Yahoo Knowledge Graph COVID-19 Datasets

The Yahoo Knowledge Graph team at Verizon Media is [...].

i2b2 (Informatics for Integrating Biology and the Bedside)

10k US Adult Faces Database

Link

2GB of Photos of Cats

Link

Audience Unfiltered faces for gender and age classification

Link

Affective Image Classification

Link

Airborne Object Detection and Tracking

Link

The AOT dataset is for airborne object tracking [...].

Animals with attributes

Link

CADDY Underwater Stereo-Vision Dataset (Hand Gestures)

Link

Contains 10K stereo pair images of divers' hand gestures [...].

Cytology Dataset - CCAgT

Link

9339 images of Cervical Cells with AgNOR Stain Technique.

Caltech Pedestrian Detection Benchmark

Link

Chars74K dataset

Link

Character Recognition in Natural Images (English & Kannada).

Cube++

Link

4890 raw 18-megapixel images, each containing a SpyderCube color target $[\ldots]$.

Densely Annotated Video Driving Data Set (DAVID)

Link

28 video sequences of driving scenes with detailed annotations.

Danbooru Tagged Anime Illustration Dataset

Link

Large-scale anime image database with 3.33m+ images, extensive tags.

DukeMTMC Data Set

Link

Multi-target multi-camera tracking data.

ETH Entomological Collection (ETHEC) Fine Grained Butterfly Images

Link

Fine-grained Lepidoptera images.

Face Recognition Benchmark

Link

Flickr: 32 Class Brand Logos

Link

GDXray

Link

X-ray images for testing and Computer Vision.

HumanEva Dataset

Link

Video sequences with calibrated grayscale and color images for

3D pose.

ImageNet (in WordNet hierarchy)

Link

Indoor Scene Recognition

Link

International Affective Picture System (UFL)

Link

KITTI Vision Benchmark Suite

Link

LILA-BC (Labeled Information Library of Alexandria

- Biology & Conservation)

Link

Over 10 million images for biology and conservation research.

MNIST database of handwritten digits

Link

Near 1 million examples.

Multi-View Region of Interest Prediction Dataset (Autonomous Driving)

Link

16 driving sequences for multi-view ROI prediction.

Massive Visual Memory Stimuli, MIT

Link

Newspaper Navigator

Link

Extracted visual content from 16,358,041 newspaper pages.

Open Images from Google

Link

Pictures with segmentation masks for 2.8M object instances.

RuFa - Arabic font dataset

Link

Images of text written in Ruqaa & Nastaliq fonts.

SUN database, MIT

SVIRO (Synthetic Vehicle Interior Rear Seat Occupancy)

Link

25,000 synthetic sceneries across ten vehicle types.

Several Shape-from-Silhouette Datasets

Link

Stanford Dogs Dataset

Link

The Action Similarity Labeling (ASLAN) Challenge

Link

The Oxford-IIIT Pet Dataset

Link

Violent-Flows

Link

Crowd Violence / Non-violence Database and benchmark.

Visual Genome

Link

YouTube Faces Database

Link

All-Age-Faces Dataset

Link

Contains 13,322 Asian face images across ages 2-69.

Audi Autonomous Driving Dataset

Link

Contains sensor data from real road traffic for autonomous driving.

B3FD - Facial age (and gender) estimation dataset

Link

375k images, "Biometrically Filtered Famous Figures" dataset.

Context-aware datasets from five domains

Link

Delve Datasets for classification and regression

Link

Discogs Monthly Data

Link

Fluorescent Neuronal Cells

Link

A new testbed for image-based neuron detection.

Free Music Archive

Link

IMDb Database

Link

Iranis (Farsi/Arabic License Plate Characters)

Link

A large-scale dataset for license plate character recognition.

Keel Repository for classification, regression & time series

Link

LLVIP

Link

Contains 30,976 images (15,488 pairs) for low-light visible-infrared person detection.

Labeled Faces in the Wild (LFW)

Link

Lending Club Loan Data

Link

Machine Learning Data Set Repository

Link

Million Song Dataset

Link

More Song Datasets

Link

MovieLens Data Sets

Link

New Yorker caption contest ratings

Link

RDataMining - "R and Data Mining" ebook data

Link

Registered Meteorites on Earth

Link

Restaurants Health Score Data in San Francisco

TikTok Dataset

300+ dance videos capturing single-person dance performance.

The Big Bad NLP Database

UCI Machine Learning Repository

Link

Blizzard Challenge Speech

Speech + text data from professional audiobooks.

Yahoo! Ratings and Classification Data

Link

Blogger Corpus

Link

YouTube-BoundingBoxes

CLiPS Stylometry Investigation Corpus

Link

Youtube 8m

Link

ClueWeb09 FACC

Link

eBay Online Auctions (2012)

Link

ClueWeb12 FACC

Link

Museums

Canada Science and Technology Museums Corporation's Open Data

Link

DBpedia - Structured data from Wikipedia

Link

Cooper-Hewitt's Collection Database

Link

Dirty Words

Link

Metropolitan Museum of Art Collection API

Link

Flickr Personal Taxonomies

Link

Minneapolis Institute of Arts metadata

Link

Link

German Political Speeches Corpus

Freebase of people, places, and things

Collection of political speeches from the German parliament.

Natural History Museum (London) Data Portal Link

Rijksmuseum Historical Art Collection

Link

Google Books Ngrams (2.2TB)

Link

Tate Collection metadata

Link

Google MC-AFP

Generated from public Gigaword dataset using Paragraph Vectors.

The Getty vocabularies

Link

Google Web 5gram (1TB, 2006)

Link

Natural Language

Automatic Keyphrase Extraction

Link

Gutenberg eBooks List

Link

© 2025 Hamed Ahmadinia / CC BY-NC 4.0

Hansards text chunks of Canadian Parliament

SMS Spam Collection in English Link

LJ Speech

Link

13,100 short audio clips of a single speaker reading.

SaudiNewsNet (Arabic, 30K articles)

Link

Collection of Saudi Newspaper Articles.

M-AILabs Speech

Link

Stanford Question Answering Dataset (SQuAD)

Link

Microsoft MAchine Reading COmprehension (MS MARCO)

Link

USENET postings corpus of 2005–2011

ink

Machine Comprehension Test (MCTest) from Microsoft Research

Link

Universal Dependencies

Link

Machine Translation of European languages

Link

Webhose - News/Blogs in multiple languages

Making Sense of Microposts 2013 - Concept Extrac-

Link

Wikidata - Wikipedia databases

Link

Making Sense of Microposts 2016 - NER and Linking

Link

Wikipedia Links data

Link

40 million entities in context.

Multi-Domain Sentiment Dataset (version 2.0)

Link

WordNet databases and tools

Link

No Language Left Behind (NLLB - 200vo)

Link

Based on Meta's metadata for mined bitext [...].

Wordbank

Link

An open, de-identified database of vocabulary development from 84,138 children.

Noisy speech database for speech enhancement / TTS models

Link

Contains clean and noisy speech data.

WorldTree Corpus of Explanation Graphs (Elementary Science Questions)

Link

A corpus of explanation graphs for QA tasks.

Open Multilingual Wordnet

Link

Neuroscience

Allen Institute Datasets

Link

POS/NER/Chunk annotated data

Link

Brain Catalogue

Link

Personae Corpus

Link

Brainomics

CodeNeuro Datasets

Link

Crystallography Open Database

Link

Collaborative Research in Computational Neuroscience (CRCNS)

link

IceCube - South Pole Neutrino Observatory

Link

FCP-INDI

Link

LIGO Open Science Center (LOSC)

Gravitational wave data from the LIGO Hanford and Living-

stone Observatories.

Human Connectome Project

Link

NASA Exoplanet Archive

Link

NDAR

Link

NSSDC (NASA) data of 550 space spacecraft

Link

NIMH Data Archive

Link

well

NeuroData

Link

Link Data from quantum simulations as referenced in Phys. Rev. A.

Quantum simulations of an electron in a 2D potential

NeuroMorpho

A curated inventory of digitally reconstructed neurons.

Sloan Digital Sky Survey (SDSS)

Link

Mapping the Universe through extensive sky surveys.

Neuroelectro

Link

Prostate Cancer

OASIS

Link

EOPC-DE-Early-Onset-Prostate-Cancer-Germany Link

OpenNEURO

Link

GENIE

Link Data from the Genomics Evidence Neoplasia Information Exchange (GENIE) project.

OpenfMRI

Link

GENE Link

Comprehensive genomic profiling of 477 prostate cancer samples.

Genomic-Hallmarks-Prostate-Adenocarcinoma-CPC-

Study Forrest

Link

MSK-IMPACT-Clinical-Sequencing-Cohort-MSKCC-Prostate-Cancer

Targeted sequencing of clinical metastatic prostate cancer.

The Nencki-Symfonia EEG/ERP dataset

A high-density EEG dataset with event-related potentials.

3 Physics

CERN Open Data Portal

Link

Metastatic-Prostate-Adenocarcinoma-MCTP

Link

Comprehensive profiling of 61 prostate cancer samples.

Metastatic-Prostate-Cancer-SU2CPCF-Dream-Team

Comprehensive analysis of 150 metastatic prostate cancer samples.

NPCR-2001-2015

link

Database from CDC's National Program of Cancer Registries.

PRAD-UK-Prostate-Adenocarcinoma-United-Kingdom

Link

Link

Prostate Adenocarcinoma - United Kingdom (ICGC Project).

PRAD-FR-Prostate-Adenocarcinoma-France

Prostate Adenocarcinoma - France (ICGC Project).

NPCR-2005-2015

Link

Database from CDC's National Program of Cancer Registries.

PROSTATEx-Challenge

link

Retrospective set of prostate MR studies for classification challenges.

NaF-Prostate

Link

A collection of F-18 NaF PET/CT prostate cancer scans.

Prostate-3T

Link

Data for an ISBI segmentation challenge.

Neuroendocrine-Prostate-Cancer

Whole-exome and RNA-Seq of castration resistant prostate cancer.

Prostate-Adenocarcinoma-Broad-Cornell-2012

Link

Comprehensive profiling of 112 prostate cancer samples.

PLCO-Prostate-Diagnostic-Procedures

Link

Prostate Diagnostic Procedures dataset (95,837 records).

Prostate-Adenocarcinoma-Broad-Cornell-2013

Link

Comprehensive profiling of 57 prostate cancer samples.

PLCO-Prostate-Medical-Complications

Link

Prostate Medical Complications dataset (3,350 records).

Prostate-Adenocarcinoma-CNA-study-MSKCC

Copy-number profiling of 103 primary prostate tumors.

PLCO-Prostate-Screening-Abnormalities

Link

Prostate Screening Abnormalities dataset (10,527 records).

Prostate-Adenocarcinoma-Fred-Hutchinson-CRC

link

Comprehensive profiling of prostate cancer.

PLCO-Prostate-Screening

35,875 subjects, 177,315 screening records.

Prostate Adenocarcinoma (MSKCC/DFCI)

Whole-exome sequencing of 1013 prostate cancer samples.

PLCO-Prostate-Treatments

link

Prostate Treatments dataset (13,409 records, 7,614 subjects).

Prostate-Adenocarcinoma-MSKCC

Link

MSKCC Prostate Oncogenome Project (181 primary, 37 metastatic).

PLCO-Prostate

A comprehensive dataset that contains nearly all PLCO Prostate data.

Prostate-Adenocarcinoma-Organoids-MSKCC

Exome profiling of prostate cancer and matched organoids.

PRAD-CA-Prostate-Adenocarcinoma-Canada

Prostate Adenocarcinoma - Canada (ICGC Project).

Prostate-Adenocarcinoma-Sun-Lab

Link

Whole-genome and transcriptome sequencing of 65 prostate tumors.

Prostate-Adenocarcinoma-TCGA-PanCancer-Atlas

Comprehensive TCGA PanCanAtlas data from 11k samples.

TCGA-PRAD-US

Link

TCGA Prostate Adenocarcinoma (499 samples).

Prostate-Adenocarcinoma-TCGA

Integrated profiling of 333 primary prostate adenocarcinoma samples.

Psychology

OSU Cognitive Modeling Repository Datasets

Prostate-Diagnosis

PCa T1- and T2-weighted MRIs acquired for diagnosis.

Open Cognitive Science Data

Publicly available behavioral datasets from across cognitive science.

Prostate-Fused-MRI-Pathology

Fused MRI/Pathology dataset for prostate.

m Public Domains

Ably Open Realtime Data

Link

Prostate-MRI

link

Prostate MRI collection for detection and segmentation.

Amazon

Link

Prostate-R

Dataset from Stamey et al, included in R'ElemStatLearn'.

Archive.org Datasets

Link

QIN-PROSTATE-Repeatability

A dataset with repeated prostate MR scans for consistency checks.

Archive-it from Internet Archive

link

CMU JASA data archive

Link

QIN-PROSTATE

Link

Collection of the Quantitative Imaging Network (QIN).

CMU StatLab collections

Link

SEER-YR1973 $_{2}$ 015.SEER9

SEERNovember 2017 Research Data from nine SEER registries (19) and the second second-2015).

Data.World

Link

SEER-YR1992₂015. $SJ_LA_RG_AK$

Link

SEERNovember 2017 Research Data from San Jose, Los Angeles, e

Data360

Link

Enigma Public

Link

SEER-YR2000₂015. $CA_KY_LO_NJ_GA$

SEERNovember 2017 Research Data from Greater California, Ke

Google

Link

 $\mathbf{SEER}\textbf{-}\mathbf{YR2005}.\mathbf{LO}_{2}ND_{H}ALF$

July--December 2005 diagnoses for Louisian a from local registry.

Grand Comics Database

The GCD is a nonprofit, internet-based database of comics.

Infochimps Wikileaks 911 pager intercepts Link **KDNuggets Data Collections** Yahoo Webscope Link Link **Q** Search Engines Microsoft Azure Data Market Free DataSets Academic Torrents of data sharing from UMB Microsoft Data Science for Research Link Base dos Dados (Data Basis) Open Data Repository for Brazil. Microsoft Research Open Data Link Datahub.io Link Open Library Data Dumps **Domains Project** Link Sorted list of Internet domains. Reddit Datasets Link Harvard Dataverse Network of scientific data Link RevolutionAnalytics Collection ICPSR (UMICH) Link Sample R data sets Link Institute of Education Sciences Stack Overflow Annual Developer Survey Link Annual developer surveys from 2011 onward. National Technical Reports Library Link StatSci.org Link

Stats4Stem R data sets (archived)

The Washington Post List Link

UCLA SOCR data collection Link

UFO Reports Link

Link

Open Data Certificates (beta)

Link

OpenDataNetwork

A search engine of Socrata powered data portals.

Statista.com - statistics and Studies Link

Zenodo

An open dependable home for the long-tail of science.

Social Networks

2021 Portuguese Elections Twitter Dataset

link

 $57\mathrm{M}+$ tweets, $1\mathrm{M}+$ users related to the 2021 Portuguese Elections.

72 hours gamergate Twitter Scrape

Link

CMU Enron Email of 150 users

Link

Cheng-Caverlee-Lee September 2009 - January 2010 Twitter Scrape

Link

China Biographical Database

Link

A freely accessible relational database of biographical info for Chinese history.

Clubhouse Dataset

Link

A Twitter Dataset of 40+ million tweets related to COVID-19

Link

43k+ Donald Trump Twitter Screenshots

Link

Screenshots of 43,475 Donald Trump tweets.

EDRM Enron EMail of 151 users, hosted on S3

Link

Facebook Data Scrape (2005)

Link

Facebook Social Connectedness Index

Link

An anonymized snapshot of active Facebook relationships.

Facebook Social Networks from LAW (since 2007)

Link

Foursquare from UMN/Sarwat (2013)

Link

GitHub Collaboration Archive

Link

Google Scholar citation relations

Link

High-Resolution Contact Networks from Wearable Sensors

Link

Indie Map

Link

Social graph and crawl of top IndieWeb sites.

Mobile Social Networks from UMASS

link

Network Twitter Data

Link

Reddit Comments

Link

Skytrax' Air Travel Reviews Dataset

Link

Social Twitter Data

Link

SourceForge.net Research Data

Link

The Reddit COVID dataset

Link

Attempts to capture the full extent of COVID-19 topics on

Reddit.

Twitch Top Streamer's Data

Link

Twitter Data for Online Reputation Management

Link

Twitter Data for Sentiment Analysis

Link

Twitter Graph of entire Twitter site

Link

Twitter Scrape Calufa May 2011

GDELT Global Events Database UNIMI/LAW Social Network Datasets Link Link United States Congress Twitter Data General Social Survey (GSS) since 1972 Tweets from 1100+ accounts of U.S. Congress members/staff, daily. German Social Survey Link Yahoo! Graph and Social Data Link Global Religious Futures Project Link Youtube Video Social Graph in 2007,2008 Link Gun Violence Data Link Social Sciences Records of over 260k gun violence incidents. ACLED (Armed Conflict Location Event Data Project) **Humanitarian Data Exchange** Link Link Authoritarian Ruling Elites Database **INFORM Index for Risk Management** Link The Authoritarian Ruling Elites Database (ARED). Link Institute for Demographic Studies Canadian Legal Information Institute Link Link Inter-American Development Bank Open Data Center for Systemic Peace Datasets Link Link Conflict Trends, Polities, State Fragility, etc. **International Networks Archive** Correlates of War Project Link Link International Social Survey Program (ISSP) Cryptome Conspiracy Theory Items Link Link International Studies Compendium Project

Datacards Link

European Social Survey Link

FBI Hate Crime 2013 - aggregated data Link

Fragile States Index Link

Link

James McGuire Cross National Data Link

MIT Reality Mining Dataset Link

MacroData Guide by Norsk samfunnsvitenskapelig datatjeneste Link

Mass Mobilization Data Project

Link

Data on mass mobilizations (protests, etc.).

Microsoft Academic Knowledge Graph

Link

A large RDF knowledge graph of academic publications.

Minnesota Population Center

Link

Notre Dame Global Adaptation Index (ND-GAIN)

Link

Open Crime and Policing Data in England, Wales and Northern Ireland

Link

OpenSanctions

Link

A global database of persons and companies of political, criminal, or economic interest.

Paul Hensel General International Data Page

Link

PewResearch Internet Survey Project

Link

PewResearch Society Data Collection

Link

Political Polarity Data

Link

StackExchange Data Explorer

Link

Terrorism Research and Analysis Consortium

Link

Texas Inmates Executed Since 1984

Link

Titanic Survival Data Set

Link

UCB's Archive of Social Science Data (D-Lab)

Link

UCLA Social Sciences Data Archive

Link

UN Civil Society Database

Link

UPJOHN for Labor Employment Research

Link

Universities Worldwide

Link

Uppsala Conflict Data Program

Link

World Bank Open Data

Link

World Inequality Database

Link

Provides open data on the wealth and income distribution

across countries.

WorldPop project

Link

Worldwide human population distributions.

</>
Software

FLOSSmole data about free, libre, and open source software development

Link

GHTorrent

Link

Scalable, queryable, offline mirror of data offered through the

GitHub REST API.

Libraries.io Open Source Repository and Dependency Metadata

Link

Public Git Archive

Link

Dataset of 182,014 top-bookmarked Git repositories.

Code duplicates

Link

 $2k\ Java\ files\ +\ 600\ Java\ function\ pairs\ labeled\ as\ similar\ or\ different.$

Commit messages

link

1.3 billion GitHub commit messages till March 2019.

Tennis database of rankings, results, and stats for $\ensuremath{\mathsf{ATP}}$

Link

Pull Request review comments

Link

25.3 million GitHub PR review comments since January 2015.

Tennis database of rankings, results, and stats for WTA

Link

Source Code Identifiers

Link

 $41.7\ \mathrm{million}$ distinct splittable identifiers from $182{,}014\ \mathrm{repositories}.$

Transfermarkt Datasets

Link

Clean, structured, and automatically updated football data.

Sports

American Ninja Warrior Obstacles

Link

USA Soccer Teams and Locations

Link

MLS, NWSL, and USL soccer teams data.

Betfair Historical Exchange Data

Link

Time Series

3W dataset Link

The first realistic, public dataset capturing rare undesirable real events in oil wells.

Cricsheet Matches (cricket)

Link

Databanks International Cross National Time Series Data Archive

Link

Equity in Athletics

Link

Brought to you by the Office of Postsecondary Education of the U.S. Department of Education.

Hard Drive Failure Rates

Link

Ergast Formula 1, from 1950 up to date (API) Link

Football/Soccer resources (data and APIs)

Link

Heart Rate Time Series from MIT

Link

Lahman's Baseball Database Link Time Series Data Library (TSDL) from MU $_{\mbox{\scriptsize Link}}$

NFL play-by-play data

Link

Sourced from an aggregator of NFL game logs.

Turing Change Point Dataset

Link

Contains 42 annotated time series for change point detection research.

Pinhooker: Thoroughbred Bloodstock Sale Data

Link

UC Riverside Time Series Dataset

Link

Pro Kabadi season 1 to 7

Link

Professional-level kabaddi league in India.

Transportation

Airlines OD Data 1987-2008 Link

Retrosheet Baseball Statistics

Link

Ford GoBike Data (formerly Bay Area Bike Share Data)

RITA Airline On-Time Performance data Bike Share Systems (BSS) collection Link **Dutch Traffic Information** RITA/BTS transport data collection (TranStat) Link Link GeoLife GPS Trajectory from Microsoft Research Renfe (Spanish National Railway Network) dataset German train system by Deutsche Bahn Toronto Bike Share Stations (JSON and GBFS files) Linkl Hubway Million Rides in MA Transport for London (TFL) Link Melbourne Pedestrian Counting Travel Tracker Survey (TTS) for Chicago Link Link Hourly pedestrian counts since 2009. U.S. Bureau of Transportation Statistics (BTS) Montreal BIXI Bike Share Link Link U.S. Domestic Flights 1990 to 2009 NYC Taxi Trip Data 2009-Link U.S. Freight Analysis Framework since 2007 NYC Taxi Trip Data 2013 (FOIA/FOILed) Link Link U.S. National Highway Traffic Safety Administration - Fatalities since 1975 NYC Uber trip data April 2014 to September 2014 Link/ Link Contains CSV data about fatal traffic accidents. Open Traffic collection eSports Link CS:GO Competitive Matchmaking Data OpenFlights - airport, airline and route data Data about CSGO matchmaking (damage, stats, etc.). Link FIFA-2021 Complete Player Dataset Philadelphia Bike Share Stations (JSON) Link Link OpenDota data dump Plane Crash Database, since 1920 Contains Dota 2 match data from OpenDota. Link

Compiled and organised by Hamed Ahmadinia. Website: ahmadinia.fi. © 2025 Hamed Ahmadinia. All rights reserved.