* **Specifically for marchine learning :**
  + **UC Irvine Machine Learning Repository :** <https://archive.ics.uci.edu/ml/index.php> → Currently maintain 559 data sets as a service to the machine learning community
* **Kaggle**: <https://www.kaggle.com/datasets>
* **Deep learning** : <http://deeplearning.net/datasets/>
* Biology
* **Gene expression Omnibus**: <https://www.ncbi.nlm.nih.gov/geo/info/datasets.html> → GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted.
* **Cancer Image Archives** : <https://www.cancerimagingarchive.net/collections/>
* **Temple University EEG Corpus (ajouté par Laura Gagliano):** <https://www.isip.piconepress.com/projects/tuh_eeg/html/overview.shtml>

→A rich archive of over 30,000 clinical EEG recordings collected at Temple University Hospital (TUH) from 2002 - present. Read this [journal paper](https://www.isip.piconepress.com/publications/journals_refereed/2016/frontiers_neuroscience/tuh_eeg/) for a more complete description of the corpus.

* Other general sources:
  + <https://www.dataquest.io/blog/free-datasets-for-projects/>
  + <https://towardsdatascience.com/all-the-datasets-you-need-to-practice-data-science-skills-and-make-a-great-portfolio-857a348883b5>
  + **Transport (ajouté par Laura Gagliano):** <https://mbta-massdot.opendata.arcgis.com>