1 Sports Dataset Repository

URL: http://www.opensourcesports.com/
Contents: The repository provides nearly up-to-date openly accessed data of various sports for the last 100+ years in case of hockey, last 40+ years in case of baseball or last 10+ years in case of basketball for men and women. Repository’s content is constantly enriched and expanded and authors are seeking contributions. Datasets are structured as relations, where every dataset has a master table with main information about the players and their IDs used by supplementary tables that contain statistics about managers, coaches, standings, salaries, teams, results, scoring, etc.

2 Time Series Dataset Repository

URL: https://podaac.jpl.nasa.gov/datasetlist
Contents: Data in PO.DAAC repository consists of successive satellite measurements of ocean and climate, focusing on topography, temperature, circulations, gravity, pressure, etc. Datasets are grouped among others by geographical location and sensors used obtaining data. Datasets are represented by a collection of measurements over the time period of the project. Every measurement consists of the date and time the measurement had been taken, the location with which the measurement is associated, followed by the administrative information regarding a particular measurement. In addition, every measurement may be associated with a graphics file that supports that measurement.

3 Anything Else Dataset Repository

URL: http://dumps.wikimedia.org/other/pagecounts-raw/
Contents: Wikipedia page-view dataset contains information about non-unique visits per every page requested by a user. The dataset lists all the requested pages line by line together with the amount of visits on that page. Dataset uses pre-defined prefixes in the dataset to differentiate between the projects within the Wikimedia resource (i.e. wikipedia, wiktionary, etc.). Overall dataset is divided by years, months and is sampled for every single hour during a day.