A.3 Text mining on the Web

- **Google Trends**

  Google Trends shows visual statistics about how often keywords have been searched on Google over time. Google Trends also shows how frequently topics have appeared in Google News stories, and in which geographic regions people have searched for them most.

  `http://www.google.com/trends`

- **Google Flu Trends**

  Google Flu Trends uses aggregated Google search data to estimate flu activity. Data available for download as well.

  `http://www.google.org/flutrends/`

- **The Observatorium**

  The Observatorium project focuses on complex network dynamics in the Internet, proposing to monitor its evolution in real-time, with the general objective of better understanding the processes of knowledge generation and opinion dynamics.

  `http://www.theobservatorium.eu/`

- **We Feel Fine**

  A database of several million human feelings, harvested from blogs and social pages in the Web. Using a series of playful interfaces, the feelings can be searched and sorted across a number of demographic slices. Web api available as well.

  `http://www.wefeelfine.org/`

- **CyberEmotions**

  The CyberEmotions project focuses on the role of collective emotions in creating, forming and breaking-up ecommunities. It makes available for download three datasets containing news and comments from the BBC News forum, Digg and MySpace, only for academic research and only after the submission of an application form.

  `http://www.cyberemotions.eu/data.html`

A.4 Social data sharing

- **Linked Data**

  Linked Data is about using the Web to connect related data that was not previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.

  `http://linkeddata.org`

- **Dataverse Network Project**

  The Dataverse Network is an application to publish, share, reference, extract and analyze research data. It facilitates making data available to others, and allows to replicate others work. Researchers and data authors get credit, publishers and distributors get credit, affiliated institutions get credit.

  `http://thedata.org/`

- **Data360**

  Data360 is an open-source, collaborative and free Web site. The site hosts a common and shared database, which any person or organization, committed to neutrality and non-partisanship (meaning "let the data speak"), can use for presentations and visualizations.

  `http://www.data360.org/`

- **Swivel**

  Swivel is a Web site where people share reports of charts and numbers. It is free for public data, and charges a monthly fee to people who want to use it in private.

  `http://www.swivel.com/`

- **Many Eyes**

  A IBM initiative that allows users to upload their datasets and use a collection of tools to obtain meaningful visualizations from them. Each visualization is publicly stored on a dedicated page, where users can comment, rate and tag it. Reuse of the data is possible and encouraged.

  `http://manyeyes.alphaworks.ibm.com/manyeyes/`

A.5 Conflict data

- **CSCW Data on Armed Conflict**

  CSCW and Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University, have collaborated in the production of a dataset of armed conflicts, both internal and external, in the period 1946 to the present. Currently, probably the most extensive dataset repository available, in particular for historic data.

  `http://www.prio.no/CSCW/Datasets/Armed-Conflict/`

- **WarViews**

  The aim of the WarViews project is to create an easy-to-use front-end for the exploration of GIS data on conflict. It can run on a Web browser or it can be displayed using Google Earth.

  `http://www.icr.ethz.ch/research/warviews/`

  The following are civil war specific datasets with additional empirical information:

  Ethnic group location dataset: `http://www.icr.ethz.ch/research/greg`

  Ethnic power balances dataset: `http://www.icr.ethz.ch/research/geoepr`

- **UCDP Datasets**

  Collection of updated datasets and codebooks from the Uppsala Conflict Data Program (UCDP).

  `http://www.pcr.uu.se/research/UCDP/data_and_publications/datasets.htm`

- **ACLED**

  Partially contained in the PRIO dataset, ACLED (Armed Conflict Location and Events Dataset) is designed for disaggregated conflict analysis and crisis mapping. This dataset codes the location of all reported conflict events in 50 countries in the developing world. Data are currently being coded from 1997 to 2009 and the project continues to backdate conflict information for African states to the year of independence.

  `http://www.acleddata.com/`

- **CERAC**

  The Conflict Analysis Resource Center hosts several cross country conflict data sets and a few datasets of particular countries. Repositories also have datasets of political instability and conflict.

  `http://www.cerac.org.co/datasets.htm`

- **The Cross-National Time-Series Data Archive**

  The Cross-National Time-Series Data Archive provides annual data for a range of countries from 1815 to the present. Frequently cited, it is one of the "leading datasets on political violence", according to Robert Bates at Harvard University. It is "possibly the most widely used event dataset" according to Henrik Urdal, International Peace Research Institute, Oslo (PRIO).

  `http://www.databanksinternational.com/`

- **Country specific repositories**

  - Iraq: `http://www.iraqbodycount.organdhttp://www.icasualties.org/`
  - Afghanistan: `http://www.icasualties.org/`

- **Terrorism**

  Collection of datasets of terrorist acts.

  `http://people.haverford.edu/bmendels/terror_attacks`

A.6 Data in economics and finance

- **Bloomberg**

  International real-time data provider for decision makers in finance, business and government.

  `http://www.bloomberg.com/`

- **Maddison Data**

  Historical statistics about GDP and population data.

  `http://www.ggdc.net/maddison/`

- **UNCTAD Statistics**

  UNCTAD offers the following databases on-line:

  - The UNCTAD Handbook of Statistics on-line provides time series of economic data and development indicators, in some cases going back as far as 1950.
  - The Commodity Price Statistics Online Database.
  - The UNCTAD-TRAINS on the Internet (Trade Analysis and Information System) for trade control measures as well as import flows by origin for over 130 countries.
  - The Foreign Direct Investment database (FDI).

  `http://www.unctad.org/Templates/Page.asp?intItemID=2364&lang=1`

- **OECD Statistics Portal**

  Large collection of datasets covering economics, demographics. Extractions are freely available, full access requires subscription.

  `http://www.oecd.org/statsportal/0,3352,en_2825_293564_1_1_1_1_1,00.html`

- **EUROSTAT**

  Detailed statistics on the EU and candidate countries, and various statistical publications for sale.

  `http://ec.europa.eu/eurostat/`

- **Where's George?**

  Spatial tracking system for U.S. and Canadian dollars.

  `http://www.wheresgeorge.com`

- **Eurobilltracker**

  Spatial tracking system for Euro banknotes.

  `http://en.eurobilltracker.com`

- **EPO Worldwide Patent Statistical Database**

  Snapshot of the EPO master documentation database (DOCDB) with worldwide coverage, containing 20 tables including bibliographic data, citations and family links.

  `http://www.epo.org/patents/patent-information/raw-data/test/product-14-24.html`

- **World Bank**

  The World Bank Data Catalog provides download access to over 2,000 socio-economic indicators from World Bank data sets.

  `http://data.worldbank.org/`

- **Penn World Tables**

  The Penn World Table provides purchasing power parity and national income accounts converted to international prices for 188 countries for some or all of the years 1950-2004.

  `http://pwt.econ.upenn.edu/`

- **World Input Output Database**

  Exposes data about the effects of increasing globalization on trade patterns, environmental degradation and economic development that uncovers the global interrelatedness of production and its socio-economic and environmental effects. More in detail, data are available for the period from 1995 to 2006, and for some major countries back to 1980 27 EU countries and 13 other major countries in the world More than 30 industries and at least 60 products

  `http://www.wiod.org/database/index.htm`

- **International Monetary Found**

  The IMF publishes a range of time series data on IMF lending, exchange rates and other economic and financial indicators.

  `http://www.imf.org/external/data.htm`

A.7 Scientific collaboration data

- **ISI Web of Knowledge**

  Comprehensive source of information in the sciences, social sciences, arts, and humanities. It encompasses several datasets, among which the following are maybe the most noteworthy:

  - *Journal Citation Reports*. It allows one to evaluate and compare journals using citation data drawn from over 7,500 scholarly and technical journals.
  - *Web of Science*. It consists of seven databases containing information gathered from thousands of scholarly journals, books, book series, reports, conferences, and more.

  `http://isiknowledge.com`

- **Google Scholar**

  Google Scholar is search engine specialized in scholarly literature. It indexes different sources (articles, books, abstract, thesis, etc.) from several disciplines and sort them according to number of citations, author and journal impact factor.

  `scholar.google.com`

- **Scholarometer**

  Scholarometer is a social tool to facilitate citation analysis and help evaluate the impact of an author's publications. It works as a software plug-in for the Firefox browser.

  `http://scholarometer.indiana.edu`

- **Scopus**

  Scopus is a very large abstract and citation database of research literature. It is available only for registered users.

  `http://www.scopus.com`

- **Living Science**

  Living Science is a real time global science observatory based on publications submitted to arXiv.org. It covers real time (daily) submissions of publications in areas as diverse as Physics, Astronomy, Computer Science, Mathematics and Quantitative Biology. Currently, contents are dynamically updated each day. Living Science is a powerful analysis tool to identify the magnitude and impact of scientific work worldwide.

  `http://www.livingscience.ethz.ch/`

- **PubMed**

  PubMed comprises more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books.

  `http://www.ncbi.nlm.nih.gov/pubmed`

A.8 Social sciences

- **ICPSR of the University of Michigan**

  ICPSR offers more than 500,000 digital files containing social science research data. Disciplines represented include political science, sociology, demography, economics, history, gerontology, criminal justice, public health, foreign policy, terrorism, health and medical care, early education, education, racial and ethnic minorities, psychology, law, substance abuse and mental health, and more.

  `http://www.icpsr.umich.edu/icpsrweb/ICPSR/`

- **UK Data Center of the University of Essex**

  The UK's largest collection of digital research data in the social sciences and humanities.

  `http://www.data-archive.ac.uk/`

- **Berkeley's UC DATA Archive**

  UC DATA's data holdings are primarily in the areas of Political, Social and Health Sciences.

  `http://ucdata.berkeley.edu/data_record.php?recid=6`

- **The Economic and Social Data Service (ESDS)**

  The Economic and Social Data Service (ESDS) is a national data service providing access and support for an extensive range of key economic and social data, both quantitative and qualitative, spanning many disciplines and themes. It contains a map of additional datasets from several European countries.

  - `http://www.esds.ac.uk/`
  - `http://www.esds.ac.uk/findingData/map.asp`

- **CESSDA**

  Wide data collections including sociological surveys, election studies, longitudinal studies, opinion polls, and census data. Among the materials are international and European data such as the European Social Survey, the Eurobarometers, and the International Social Survey Programme.

  `http://www.cessda.org/`

- **Gapminder Data**

  Gapminder is a popular technology and Web application for cross-visualisation of trends in time series of data. It also opens an archive of multiple datasets on diverse socio-economic indicators.

  `http://www.gapminder.org/data/`

- **World Value Survey**

  The World Value Survey provides data about values and cultural changes in societies all over the world.

  `http://www.worldvaluessurvey.org/`

A.9 Urban data

- **Global Urban Observatory database**

  The Global Urban Observatory (GUO) offers policy-oriented urban indicators, statistics and other urban information.

  `http://www.devinfo.info/urbaninfo/`

- **Urban Observatory**

  U.S. based datasets about wealth, innovation and crime across cities.

  `http://santafe.edu/urban_observatory/`

- **Urban Audit**

  Urban Audit contains a collection of comparable statistics and indicators for European cities. Data for most recent years is missing at the time of writing.

  `http://www.urbanaudit.org/`

- **Globalization and World Cities Research Network**

  The Globalization and World Cities Research Network (GaWC) promotes himself as the leading academic thinktank on cities in globalization. Several datasets are available for large cities networks.

  `http://lboro.ac.uk/gawc/data.html`

A.10 Traffic data

- **NGSIM**

  The Next Generation Simulation (NGSIM) program was initiated by the United States Department of Transportation (US DOT). The program developed a core of open behavioral algorithms in support of traffic simulation, and collected high-quality primary traffic and trajectory data intended to support the research and testing of the new algorithms.

  `http://ngsim-community.org/`

- **Swiss Federal Roads Office FEDRO**

  The Swiss Federal Roads Office offers a comprehensive overview on traffic flows in Switzerland. Data are collected by permanent automatic traffic counting stations and complemented by regular manual checking since 1961.

  `http://www.astra.admin.ch/verkehrsdaten/00297/index.html`

- **TrafficData**

  The aim of the International Traffic Database (ITDb) project is to provide traffic data to various groups (researchers, practitioners, public entities) in a format according to their particular needs, ranging from raw measurement data to statistical analysis. ITDb promotes a flexible traffic data provision format based on user needs and standard habits.

  `http://www.trafficdata.info/`

- **Clearing House for Transport Data**

  The Clearing House for Transport Data in the German Aerospace Center is the first point of contact for a quick overview of the available data. It is targeted at both organizations who gather transport-relevant data and those who wish to use the results of such research. The information offered includes the preparation of detailed metadata on the data sets, as well as notes on possible uses and sources.

  `http://www.dlr.de/cs/en/desktopdefault.aspx/tabid?669/1177\_read?2160/`

- **Desweiteren das Regiolab Delft**

  The regiolab-delft initiative started just after 2000 as a joint project led by TU Delft in association with the Municipality of Delft, the TRAIL research school, the Province of South Holland, the Ministry of Transport and several industrial partners. The archived dataset consists of over 6 years of 1 minute averaged speed and aggregate flow data from densely spaced inductive loops on the freeway network in the province of south Holland and other data from intersection controllers, license plate detection camera's and much more.

  `http://www.regiolab-delft.nl`

- **RITA** The Research and Innovative Technology Administration (RITA) of the U.S. Department of Transportation offers several datasets about maritime, freights, airline, passengers, etc. traffic statistics.

  `http://www.bts.gov/data_and_statistics/`

- **ETH Travel Data Archive (ETHTDA)**

  The ETH Travel Data Archive (ETHTDA) is a virtual platform allowing end users to browse the archived travel data over the Web and enabling simple statistical analysis.

  `http://www.ivt.ethz.ch/vpl/publications/ethtda`

- **Metropolitan Travel Survey Archive**

  The Metropolitan Travel Survey Archive to store, preserve, and make publicly available, via the Internet, travel surveys conducted by metropolitan areas, states and localities.

  `http://www.surveyarchive.org/`

- **Infoblu**

  Infoblu is a private company providing real-time traffic monitoring services for Italy. All services are available for a fee.

  `http://www.infoblu.it`

- **ENAC**

  Our Air Transport database comprises rich and detailed information on airlines, airports and traffic flow. In order to increase its scope and its reliability, ENAC also carries out annual surveys of airlines and airports.

  `http://www.enac.fr/recherche/leea/databaseA.htm`

A.11 Open maps

- **Google Maps**

  World-famous map service. It offers several additional services such as: Street View, user-uploaded content (photos, comments and ratings) and personalized overlays through service apis.

  `http://maps.google.com`

- **OpenStreetMap**

  OpenStreetMap (by UCL) is a free editable map of the whole world. OpenStreetMap allows you to view, edit and use geographical data in a collaborative way from anywhere on Earth.

  `http://www.openstreetmap.org/`

- **Tracksource Brasil**

  Tracksource is a collaborative project aimed at creating and distributing for free maps of Brasil.

  `http://www.tracksource.org.br`

A.12 Logistics data

- **National Household Travel Survey**

  The National Household Travel Survey (NHTS) collect data on both long-distance and local travel by the American public. The joint survey gathers trip-related data such as mode of transportation, duration, distance and purpose of trip. It also gathers demographic, geographic, and economic data for analysis purposes. It is part of RITA (A.10).

  `http://www.bts.gov/programs/national_household_travel_survey/`

- **Commodity Flow Survey**

  The Commodity Flow Survey (CFS) is the primary source of national and state-level data on domestic freight shipments by American establishments in mining, manufacturing, wholesale, auxiliaries, and selected retail industries. Data are provided on the types, origins and destinations, values, weights, modes of transport, distance shipped, and ton-miles of commodities shipped. It is part of RITA (A.10) and it is conducted every five years (last sampling on 2007).

  `http://www.bts.gov/publications/commodity_flow_survey/`

A.13 Health Data

- **World Health Organization**

  The World Health Organization publishes on line several statistics and supply direct access to four rich databases:

  - Global Health Observatory - WHO Global InfoBase - Global Health Atlas - Regional statistics

  `http://www.who.int/research/en/index.html`

A.14 Climate and Environmental data

- **Jülich**

  Climate data from Jülich Research Center. **I did not find data available for download!**

  `http://www.fz-juelich.de`

- **Google.org**

  Google introduces its data-driven philanthropic projects, among which two environmental satellite observatories:
  - the Earth Engine: for monitoring trends in world deforestation;
  - the Crisis Response: for monitoring the oil spill from the Deep Horizon sank platform.

  `http://www.google.org/`

- **Footprint Network**

  Ecological Footprint and the biocapacity results for more than 100 nations, based upon data from 2007, the most recent year for which source data are available. The tables reflect the calculations from the 2010 National Footprint Accounts.

  `http://www.footprintnetwork.org`

- **PSD Climate and Weather Data**

  PSD archives a wide range of data ranging from gridded climate datasets extending hundreds of years to real-time wind profiler data at a single location. The data or products derived from this data, organized by type, are available to scientists and the general public at the links below.

  `http://esrl.noaa.gov/psd/data/`

- **EPA DataFinder**

  The Environmental Protection Agency Data Finder is a single place to find EPA's numerical data sources so that people can access and understand environmental information. All of the data sources are available on the Internet and have been organized by topics such as air, water, and chemicals.

  `http://www.epa.gov/datafinder/`

A.15 Energy

- **International Energy Agency**

  Vast repository of statistics about supply and consumption of energy sources. Some datasets available only for sale.

  `http://iea.org/stats/index.asp`

A.16 Governance, Trade and Settlements Data

- **Govindicators**

  The Worldwide Governance Indicators (WGI) project reports aggregate and individual governance indicators for 213 economies over the period 19962009, for six dimensions of governance:

  1. Voice and Accountability
  2. Political Stability and Absence of Violence
  3. Government Effectiveness
  4. Regulatory Quality
  5. Rule of Law
  6. Control of Corruption

  http://info.worldbank.org/governance/wgi/index.asp

- **WTO International trade and tariff data**

  The World Trade Organization offers an updated and comprehensive outlook over trade policy and multilateral trading systems.

  http://www.wto.org/english/res_e/statis_e/statis_e.htm

A.17 Reality mining

- **Reality Mining**

  Behavioral data collected from 100 mobile phones over 9 months. Includes both proximity and phone usage statistics. Two anonymized datasets available: single user (MySQL) and global (Matlab).

  http://reality.media.mit.edu/

A.18 Other open data initiatives

- **Data.gov**

  Wide collection of public US datasets available for research.

  http://www.data.gov

- **Data.gov.uk**

  Wide collection of public UK datasets available for research.

  http://data.gov.uk/

- **Digging Into Data**

  Launched by the National Science Foundation (NSF), it offers a collection of diverse data repositories.

  http://www.diggingintodata.org/

- **Guardian Data Blog**

  Data journalism initiative that posts public interest (primarily UK relevant) datasets together with their analysis. A few collaborations with data visualization artists are present as well.

  http://www.guardian.co.uk/news/datablog