# W01-1: Getting and organizing data

For a warm up, this worksheet guides you in getting some data from online data basis. Your challenge is to select the right data set and organize it on your hard disk. The downloaded data set will be the basis for following sessions.

After completing this worksheet you should have made up your mind on how to structure your data on your hard disk.

## Things you need for this worksheet

• Internet browser — Any browser will be fine.

## Online data basis

There are countless online data sources available. Some are serious, some are dubious and you have to decide which one to trust (if you are not sure about that, have a look on our media literacy module).

In general, it is a good idea to start an online search when you are looking for data. However, you might notice after some time that e.g. USGS or NASA is a good source for satellite images and the UN or World Bank Group is a good source for some world wide statistical indicators. During the module, you will get to know some of such data bases and it might be a good idea to make yourself a collection of these data bases e.g. using your browser's bookmark functions.

Let's get some data.

## Learning log assignments

😎 As always, please add these entries to your today's learning log at teachwiki:

• Favorite aspect of the session (if any)
• Superfluous aspect of the session (if any)
• Eureka effect (if any)
• Links to what I've learned so far (if any)
• Questions (if any)

For more information see this short howto.

**As today's special, please complete the following assignment:**

As basis for the next sessions, you should download five data sets from three different sources and store them in a folder-based but organized way on your storage medium.

🙂 Please get information on the

- gross national income (GNI) per capita converted to U.S. dollars using the World Bank Atlas method
- carbon dioxide emissions (absolute in kt)

for the period since 1960 from the World Bank Group data base.

🙂 Please get information on the

- impact of natural disasters including information on the location, disaster groups to subtypes and helth impacts

for the period since 1900 from the EM-DAT data base maintained by the Université Catholique de Louvain, Brussels (Belgium). Since one can access the EM-DAT data in two ways and one of them requires a registration, please use the advanced search interface to build your data.

🙂 Please get information on the

- chronology of Ebola outbreaks since 1976 including information on the location, death rates and situation etc.
- chronology of the latest Ebola outbreaks in 2014

from the of the Centers for Disease Control and Prevention. In contrast to the other data sources, the CDC does neither offer search nor export functions on their site so you have to extract and store the information directly from the respective web sites which can be found here and here.

Feel free to download/store the data in a format of your choice and organize it on your storage media. Please think of additional structuring components which might be feasible for further data processing using self-developed R scripts.

🙂 Please use your learning log to present how you organized your data and explain why you did it that way and why you choose the individual data formats with a maximum of 160 words.