# Information Visualization

# CHECKPOINT II: Data cleaning and processing

G01-A

**1. Initial Dataset**

We have a table (All Winners) containing all the editions of the games from 1896 until 2008, and for each there is the city and year they were at, and a set of podium finishes, which include the sport, discipline and event, and the athlete's name and country code (NOC, same as IOC and 3-letter ISO) and the medal they won. We have another table (Total) with the total of medals each country won, with its name, NOC, total medals and how many of gold, silver and bronze.

We have a table (Codes) with each country name, IOC (same as NOC) code and their ISO code.

We have another table (Population) with each country’s name, country code (3-letter ISO), indicator name and code, and a field for their population for each year from 1960 until 2014.

We have a table (Coordinates) to know the coordinates of each country, which has its 2-letter ISO code, latitude and longitude, and its name.

**2. Selected/Derived Data**

* + - 1. All Winners – Edition year, Sport, Medal won and the NOC of the country of the medallist.  
         Total – NOC, Total, Golds, Silvers, Bronzes.  
         Codes – Country, IOC, ISO code.  
         Population – ISO Code, Years from 1960 until 2008.  
         Coordinates – ISO Code, Latitude, Longitude.
      2. Derived measures – We want to compare the medals per capita over the years, so we will count the number of medals each country won in each year and divide it by the population in that year.
      3. **3. Data abstraction**

1. Data abstraction description:

* Description of the dataset type (spatial, table, field, etc.);
* Description of each item and attribute (nominal/ordinal/etc., diverging/sequential scale, etc.);
* Semantics (what does each attribute and item stand for).
  + - 1. **4. Dataset processing**

Description of how the dataset was processed (cleaned, problems found and solutions, how did you fix missing values, cross-referenced different tables/datasets, etc.).

* + - 1. **5. Mapping (Data sample / Questions)**

Some examples that show that with your data sample you will be able to provide the answers to the questions you formulated.