# Information Visualization

# CHECKPOINT II: Data cleaning and processing

G01-A

**1. Initial Dataset**

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

We have another table with each country, including their 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 (from Google Developers) to know the coordinates of each country, which has its 2-letter ISO code, latitude and longitude, and its name.

**2. Selected/Derived Data**

Description of:

* Which were selected from your dataset?
* Which derived measures did you consider and why (based on your tasks and questions)?
  + - 1. **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.