

# Knowledge and the Web 2016/2017:

## Exercise session 2

October 19, 2016

### OBJECTIVE

In this exercise session we will address one of the core issues when working the data - data quality.

### TASK:

You are given 4 datasets:

1. *flights.data* : data containing a number of flights from and to Brussels.
2. *payments.data* : data about payments to employees of a company
3. *adult.data* : data about people from different countries
4. *eda.data* : data about patients in different hospitals.

Each of these datasets have some quality issues. your task is to find them. You can use any tool you want to analyse the data. At the end of the session, we will discuss different ways you used to find the problems in data.