**Analysis Data Codebook**

Data is provided by The Uppsala Conflict Data Program. Data from this source is widely used in academic research, and has been prepared for that use, both in its definitions and in its reliability. Data set we are looking at shows one violent incident per entry with over 40 properties for that incident. Data runs till 2015.

The focus of the dataset is conflict dynamics and the effects of armed violence, in the form of deaths.

The dataset is constructed in such a way as maximise the comparability and consistency across time and space, and provide a globally consistent image of the phenomenon of organised violence. It is a tool for global understanding of subnational conflict patterns and trends.

Current data represents period between 1989 and 2015 and does not include data for Syria.

*Data file exploration*

Using powershell it becomes clear that there are 127264 lines in the file. We know that every line represents one entry, and therefore the dataset contains 127.264 entries of conflict events.

First entry sample:   
country of events (country): Turkey  
best estimate of number of fatalities (best\_est or best): 33  
type of violence (type\_of\_violence): 1 = state-based conflict (2 = non-state conflict; 3 = one-sided violence)  
year (year): 1997

*Formulating Research Question*

Japan: 17

China: 183

South-Africa: 0

Iran: 1710

Colombia: 4495, 2004: 689

USA: 2028, 2010: 178

Bosnia: 1946, 1993: 671

*Research Question: What was the total number of fatalities in Bosnia in the year 1993?*