

Social Science Data Analytics
together with the
Political Science Department Workshop on Conflict

present a talk on

“Principled Data Processing: How We Do
Data Analysis that is Scalable, Auditable,
Reproducible and Transparent”

by

Patrick Ball

Executive Director of the Human Rights Data Analysis
Group

When an analysis depends on statistics, tables, and graphs, how can we assure that the published results are the result of all the prior processing and computational steps? Some analyses require hundreds of transformations between the original data import and the final results, including data cleaning, table joins, aggregation, record linkage, estimation, smoothing, and visualization, among others. This talk will describe the data management practices developed by the Human Rights Data Analysis Group over the last fifteen years. From unix pipelines to self-documenting code, literate programming, and reproducible research, we will review ideas and tools that provide an audit trail backward from every statistic in a report, through each transformation all the way back to the data import.

October 23, 2015, 10:00 am
104 South Kedzie Hall