Open Data and the Quest of Meaning

Constantin Vica

University of Bucharest

normev.ro

Paris, 6th July 2012

A critical approach

Hypothesis:

open data only under some explicit conditions fosters political **understanding**, **innovation** and **participation**.

Open data are not so linked

From data corpus we are not able to extract **the topology of the relational network** between:

MPs

lobby groups and MPs

NGOs and MPs

MPs and ministers

MPs and state institutions

We need **a good mapping of** those **unofficial relations** between political actors {the hidden part of politics} to help corruption detection.

I assume this network of relations is a **scale-free network**; it has a *small world* topology.

Open data and science

Wired, 2008: Big Data is the end of science.

<p style="text-align:center;">**Wrong!**</p>

2 problems:

- data in itself can't offer answers, you need **representative samples** for analysis.
- you need a scientific methodology {hypothesis, model, testing} for control and **replication** possibility.

Open legislative data and prediction

- Under the current ways of presenting open legislative data, **it is possible to predict future behavior** of MPs and political groups?

- **It is possible to predict future evolutions of the law** in different policy areas?

NO.

Open data movement lacks a *mature scientific methodology* (hypothesis, model, testing) in interpreting results.

Scientific virtues: *openness* {implies possibility of replication}; the interplay between *the reductionist and the holistic* view.

A **reliable predictive model** is a condition for open legislative data to generate political understanding and innovation.

Open data and knowledge

Knowledge is acquired through **true factual well-formed data** (Floridi 2009).

It is impossible to jump from data to knowledge without *conceptualization*, *representation*, and *structure*.

This implies:

(1) a semantic level of data aggregation (e.g. the web application where the data makes sense for the viewer)

(2) a test of logical validity.

(1) Data aggregation can be problematic

Data is a form of reductionism applied to a complex reality, i.e. "the political life" of the Parliament.

From data aggregation emerges *a meaning unseen before* of the complex reality or *new properties* of it? - the classical debate about **emergence**.

+ the problem of relevance and usefulness of data.

(2) What kind of epistemological object is an open database?

It is built under the **open world assumption**:

1. when we don't know something, we don't attribute by default the value "false",
2. the system is always inconsistent,
3. it is impossible to have complete knowledge.

Conclusion

From the conjunction of (1) & (2):
**"Not always open legislative data
will generate knowledge."**

The first task is to find the situations *when knowledge is possible* and focus on them.
The second one is to build *predictive models*.

Thank you!

normev.ro

The Emergence and Evolution of Social Norms