Bachelor in Politics And International Studies
Bachelor in Sociology

Digital Data, Data Collection (QS102)

Week 3
The role of data in the social sciences
before and since the digital times
+
Surveys in social and political research
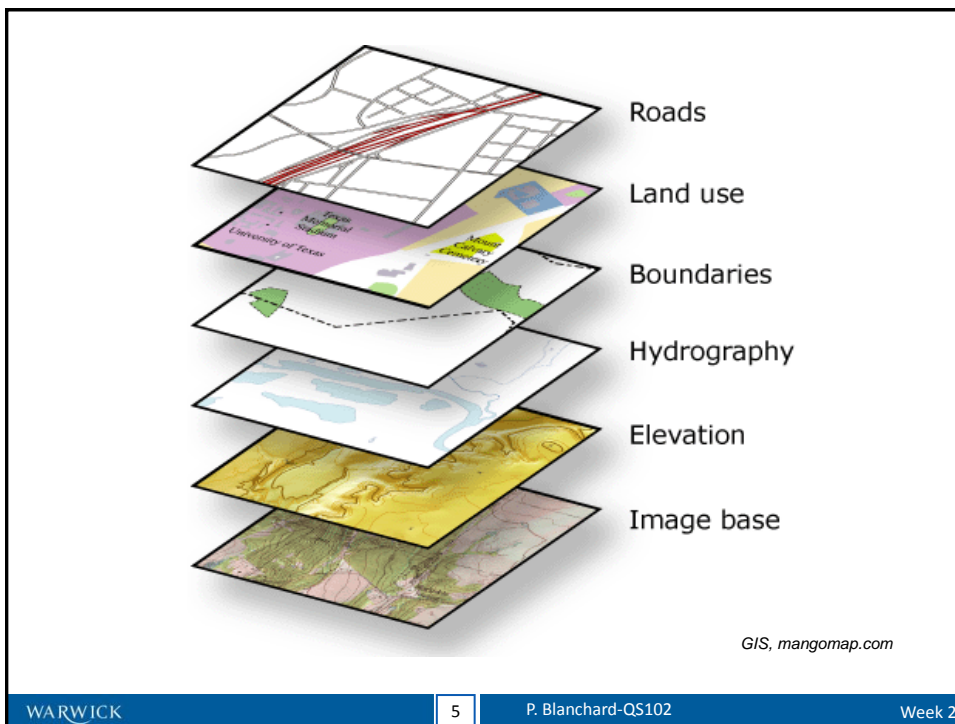
WARWICK | 1 | P. Blanchard-QS102 | Week 2

Summary of week 1

- Data are rising, including in the social sciences

- Data are part of a philosophical and epistemological debate

WARWICK | 2 | P. Blanchard-QS102 | Week 1

29/01/2015

"The School of Athens", Raffaello Sanzio da Urbino 1510

WARWICK | 3 | P. Blanchard-QS102 | Week 2

Summary of week 1

- Data are rising, including in the social sciences

- Data are part of a philosophical and epistemological debate

- There any hardly any pure/raw data, rather layers of transformation

WARWICK | 4 | P. Blanchard-QS102 | Week 2

GIS, mangomap.com

Summary of week 1

- Data are rising, including in the social sciences

- Data are part of a philosophical and epistemological debate

- There any hardly any pure/raw data, rather layers of transformation

- Data can be of many kinds and formats, but need standardisation

Dataset definition

column properties
(checks, limitation etc)

dataset (table)

columns definition

dataset
properties
(some table
properties)

dataset row

dataset value

abantecart.com

British Social Attitudes 2013

What about data for social and political sciences?

- Data can be of any kind: words, raw numbers, percentages, images, hyperlinks…

- Data can describe many objects: events, opinions, ideological positions, attitudes, statuses, actions, relationships, etc.

- Data are related to a unit of analysis: individuals, groups, regions, countries, organisations, families, events, actions, records, products, etc.

- Most common: individuals x variables

⟶ Data are **formatted** (standardised) in order to be used, reused, shared, replicated, combined, merged, compared, etc.

What about data for social and political sciences?

⟶ Data are **formatted** (standardised) in order to be used, reused, shared, replicated, combined, merged, compared.

⟶ Formatting, standardisation, reuse, comparisons, etc. require **infrastructures**: data services, data software, data specialists, data servers, data conferences, data networks, data providers...

⟶ Data norms, data procedures, data skills.

Simple examples of social and political sciences data

Counts

Victims of police violence

http://www.innocentdown.org/

Regressin.deadspin.orgg.deadspin.org

Fatal encounters

Migrants to Europe

The migrants files

Automated measures

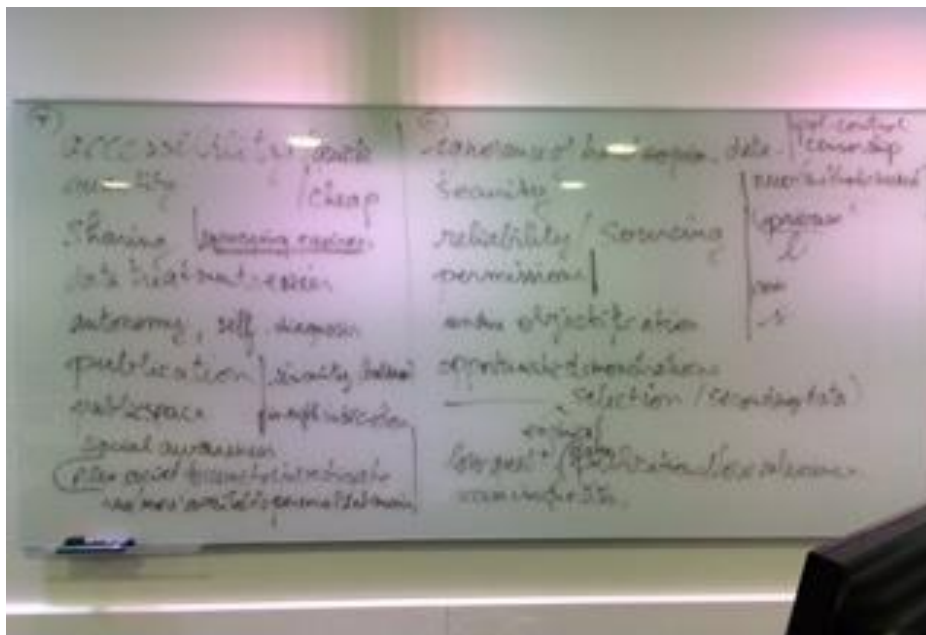What the birth of a son can change to your life… ("Quantified self")

What are good data (for social and political sciences)?

Criteria:

- Availability (technical, political, financial…)
- Relevance
- Cost

- Reliability
- Validity
- Richness
- Usability

$\longrightarrow$ **Data quality is variable. Seldom an easy work.**

What does Internet change to the role of data?

WARWICK 13 P. Blanchard-QS102 Week 2

What does Internet change to the role of data?

- Opportunities
  - Standard medium, permanent, everywhere, mainly free, most people connected
  - Standardised and rationalised information
  - End of organisational gate-keeping
  - New activities (transnational and trans-class interactions, self-exhibition…), new social settings, new borders between private and public…

- Challenges
  - Profusion ('big data') → 'data bubble' (Golder and Macy *in G*itelman (ed) 2013, chap. 7)
  - Distortions and fakes
  - Overlaps and contradictions
  - Sourcing ommitted or wrong, theft
  - Quality
  - Prenotions, routines, seduction, conditionning, manipulation
  - Illusion of knowledge (see Kitchin 2014)

WARWICK 14 P. Blanchard-QS102 Week 2

What does Internet change to the role of data?

- Opportunities

- Challenges

- Internet is not so ideal
  - Without borders and beyond domestic cultures?
  - Anonymous?
  - A public space?
  - Exempt from geopolitical and geoeconomic constraints?
  - Free space for expression?
  - Easy technique?
  - Homogeneous space?

→ **A bundle of political and technical limitations!**

WARWICK   15   P. Blanchard-QS102   Week 2

What kinds of data on the Internet?

- Traces *or* experiments
  - Archives, automatic records, blogs and other webpages
    versus interviews, emails, experiments, participatory observation

- At individual *or* collective level
  - CVs, smartphones, blogs
    versus forums, social media, administrations

- Small *or* big
  - Targetted survey, family blog
    versus FaceBook, health insurances/administrations/laboratories, Amazon

→ **Already as diverse as pre-Internet data.**

WARWICK   16   P. Blanchard-QS102   Week 2

How do big data challenge social sciences?

- Kitchin 2014
  - Why is data mining opposed to theorising? Do you think the 'new empiricism' can work in the social sciences?
  - What are the 'digital humanities'? Can you give examples of 'big data' in history, archeology or literature, or another social science or humanities?