Volunteered GI, Big Data and Social Media: Opportunities for Health and Wellbeing

Prof Clive Sabel, Aarhus University, Denmark
Scope of talk

- Big Data

- Volunteered Geographic Information
  - Citizens as sensors

- Social Media Data Mining
  - Social media mining is a process of representing, analyzing, and extracting actionable patterns from social media data such as twitter.
Underpinned by (Spatial) Data Science

Spatial Big-Data
Geographical Information Science
exploratory data mining
data visualisation
pattern detection
use of citizen volunteered, social media

Smart sensing of individuals
GPS
Environmental
Development of apps and micro-sensors to capture response to physical and social environments
My working definition of **Big Data**

- Big data not just about using large data sets, but critically, the COMBINATION of (huge) datasets.

- But its not just about data, but also:
  - Extremely large data sets that may be *analysed* computationally to reveal patterns, trends, and associations,
  - The term has been used simply to mean the use of *predictive analytics* to extract value from data, particularly the *added value* from integrating disparate datasets to reveal a sum greater than the individual parts.
Volunteered Geographic Information (VGI) is the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals (Goodchild, 2007).

- VGI=Crowdsourced
- Passive/Active contributions
  - Eg. Apps polling geographically referenced personal activity data
- Can be problematic, ref quality
- Power in the quantity, not quality of data
Understanding mobility

Paths of African American (purple) and Asian American (blue) women in Portland, Oregon, over the course of a typical day. The vertical dimension is time. Mei-Po Kwan, Department of Geography, Illinois University.
Mining Social Media – to understand social processes

- Twitter, facebook etc
- Aggregate data
- Search terms, transactions, text...
- Examples
  - Google flu trends
- Characteristics
  - Big data
  - Unsolicited
  - Geographical precision variable
  - Demographically skewed by youth, income, gender...
  - Quality variable – low signal to noise ratio

United States Flu Activity

Google flu searches (Standard Deviations)
Quantified Self: essentially, collecting data on yourself

So what?
Get a smart device, get the data,
display data, analyse the data, .... ?
• Economic necessity to adopt inclusive definition of health:

more than simply the absence of disease, but rather ‘… a state of complete physical, mental and social well-being’
WHO (1946)

• Healthcare costs are spiralling upwards
  – 95% of healthcare costs are in treating disease
  – Only 5% are in preventative healthcare

• Can investment in Wellbeing promotion reduce healthcare costs?
Don’t forget the social dimensions of health...

• Sugest that social ‘exposures’ could be as important as physical environmental exposures.

• How does our wellbeing respond to differing social environments?

• Health related behaviours such as smoking, alcohol consumption and diet are partially explained by the social contexts in which people live

• Burden falls particularly in vulnerable populations young, elderly, low-SES, ethnic minorities, sick, females …
‘New’ Social Media Epidemiology

Twitter feeds [microblog] of ‘happy’
In Summary

• Big Data presents significant opportunities for research and science in Health & Wellbeing.
• Big Data is a Disruptive Technology for many applications.
• Big Data is not just about “Big”, better termed ‘Rich Data’
• We should exploit all obtainable data.
• There are still many challenges.