BIG DATA = Volume, Velocity and Variety

Digital Inclusion

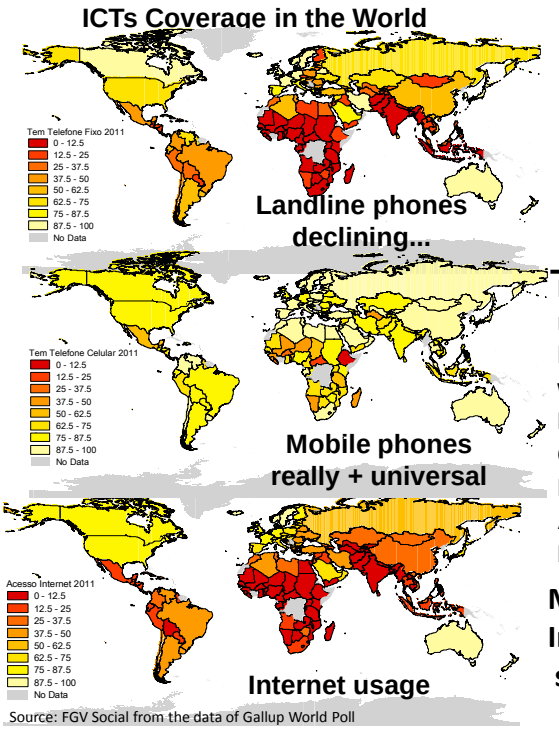
Coverage, Consistency, Convergency, Conectivity

Who? | How much? | How? | Where?

Content and Capabilities:

For what?

The harmonious sum of these vectors keep the promise of reducing costs and broadening the possibilities on our everyday activities (including the design and application of urban policies) but **sets challenges to the BIGDATA sample statistical representation.**

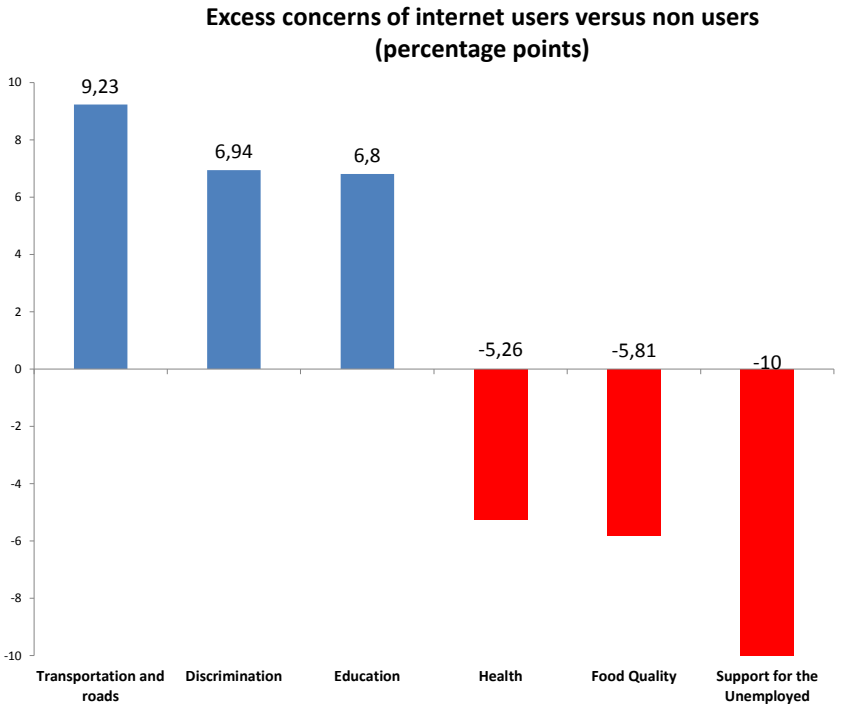


Why not?

Reasons for the e-xclusion

Didn't think it was necessary	**33,14%**
Didn't know how to use web	**31,45%**
No computer or device	**29,79%**
Computer cost was too high	1,76%
Another reason	1,52%
Internet cost too high	0,40%

Main Reasons not to use the Internet are from the Demand side



Data for Public Policies

BIGDATA

Information and action in real time!

Usage of the Volume of idle Data Available (but with unkown quality variability)

Flexible Approach (think outside of the box)

Does it reflect the whole population? (e.g.: Young people with high income and education are overrepresented in the web)

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2978465 **(Critical View)**

PERCEPTION DATA

Low representativeness of population

Technical Flexibilities

Aspirations adaptation is a problem, cultural diversity as well

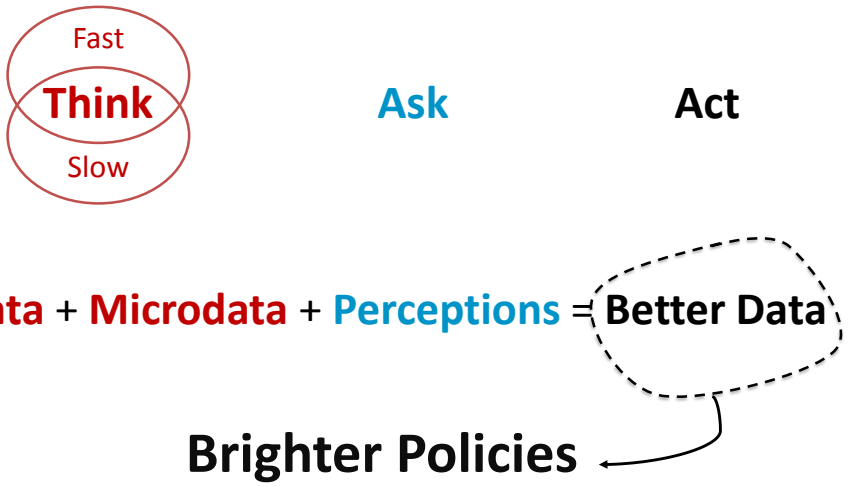
MICRODATA

Representative data of all the population

Structured approach – allows to scientifically test hypothesis of significance and try to identify causality (RCTs)

Incorporate the basis to the database (allows to minimize representativeness distortions of BIGDATA's)

RCTs is key to verify causality https://www.povertyactionlab.org/handbook-field-experiments **but is not everything (see last slide)**



****Understanding and misunderstanding randomized controlled trials (RCTs)**

Angus Deaton and Nancy Cartwright (critical view) http://www.nber.org/papers/w2259

RCTs are valuable tools whose use is spreading in economics and in other social sciences. They are seen as desirable aids in scientific discovery and for generating evidence for policy. Yet some of the enthusiasm for RCTs appears to be based on misunderstandings: that randomization provides a fair test by equalizing everything but the treatment and so allows a precise estimate of the treatment alone; that randomization is required to solve selection problems; that lack of blinding does little to compromise inference; and that statistical inference in RCTs is straightforward, because it requires only the comparison of two means. None of these statements is true. RCTs do indeed require minimal assumptions and can operate with little prior knowledge, an advantage when persuading distrustful audiences, but a crucial disadvantage for cumulative scientific progress, where randomization adds noise and undermines precision. The lack of connection between RCTs and other scientific knowledge makes it hard to use them outside of the exact context in which they are conducted.

Yet, once they are seen as part of a cumulative program, they can play a role in building general knowledge and useful predictions, provided they are combined with other methods, including conceptual and theoretical development, to discover not "what works," but why things work. Unless we are prepared to make assumptions, and to stand on what we know, making statements that will be incredible to some, all the credibility of RCTs is for naught. ***Video on Barnajee x Deaton at NYU

http://int.search.myway.com/search/video.jhtml?n=7839a0aa&p2=%5EBYU%5Exdm119%5ELMPTBR%5Ebr&pg=video&pn=1&ptb=57353D8E-1A15-43BE-A271-