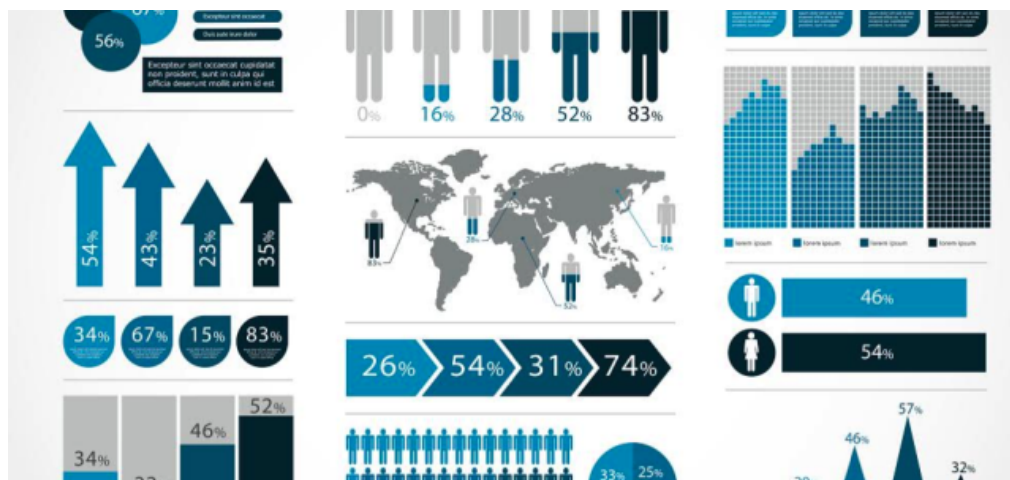


DISCOVER SOCIETY

MEASURED – FACTUAL – CRITICAL

[HOME](#) [FOCUS](#) [VIEWPOINT](#) [ARTICLES](#) [ON THE FRONTLINE](#) [POLICY BRIEFING](#) [CONTACT](#) [ARCHIVE](#)


RAPID RESPONSE



Support International Day for the Elimination of Violence against Women

A POLITICS OF COUNTING - PUTTING PEOPLE BACK INTO BIG DATA

By [discoversociety](#) July 30, 2015 1 Comment Articles, Issue 23

[Email](#)
[Twitter](#)
[Facebook 45](#)
[Tumblr](#)
[Google+](#)
[More](#)

Hamish Robertson and Joanne Travaglia (University of New South Wales)

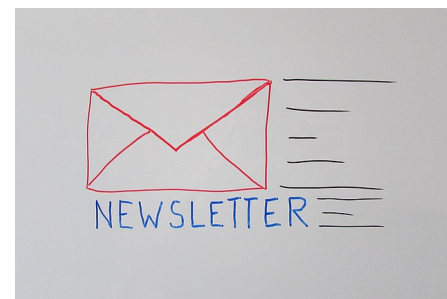
Our growing enthusiasm for the big data concept glosses over a dynamic and expanding politics of the information age. The focus on new tools and methods can distract us from those data, ideas or people who are included or excluded in this new digital world. Who gets included in or excluded from big data environments depends on how we are defined and quantified in the first instance. There is a politics to this because even with ubiquitous technologies, inclusions and exclusions persist.

Historically data collection was built on the premise that people are essentially analogue and that their binary digital data was by definition a substitute for the real thing – not least of all because of the limitations of the systems used to collect and analyse data. Although this process of exclusion/inclusion has a well-known parallel in the policy field (where “no policy as policy” refers to the lack of inclusion of an issue or a group as a process of ignoring or downplaying a problem) it is heightened in relation to big data because for so many the assumption is that big data almost by definition means *all* data.

We are however, entering a period when so much of who we are and what we do begins and ends as digital data – including our online comments and opinions, our preferences of and for everything from our anxieties to our holiday destinations and even our biomedical selves, as online consultations, digital pathology testing and electronic health records become the norm. The risk is that the complexity and nuances of our “analogue” selves will be ignored unless new methods emerge to better capture the embodied experiences of our lives and choices, both individually and collectively.

There is a risk that we are developing an ideological system that insists on the reality of the digital in preference to the analogue. This is partly because big data collection and analysis sits at the core of many surveillance technologies and therefore at the core systems of power and control. The social, political and personal focus on risk **so clearly identified by Beck**, as the

SIGN UP TO RECEIVE A MONTHLY NEWSLETTER WITH THE LATEST ISSUE OF DISCOVER SOCIETY



BECOME A FRIEND OF DISCOVER SOCIETY

Help Support Discover Society

[DONATE to DS](#)

Find out more about donating to us

MOST POPULAR POSTS OF THE DAY

The Media and Child Abuse

VIEWPOINT: Brexit, Class and British 'National' Identity

The Full Monty? Women Customers' Experiences at a Male Strip Show

prevailing organizing principle of the 20th century, has developed along with its enabling technologies into a seemingly endless desire for data.

It has become clear that what is missing from this field is a fully-fledged theory of digital political economy. Many information systems still collect highly abstracted and abbreviated data about people for two reasons. First, the process of reduction by definition has always been central to traditional data collection processes. The seemingly basic question of “in which city were you born?” belies the analogue complexity of the answer “St Petersburg/Leningrad/Petrograd/St Petersburg”.

Secondly, the transition from electro-mechanical information systems to fully digital ones has not yet seen a revision of the way database fields rely on negotiated, abbreviated and contested social concepts. While the mechanisms for collecting, storing and even analyzing data have become infinitely more sophisticated, the philosophical, political and economic debates continue. The persistence in 2015, for example, of the linguistically convoluted and largely meaningless “culturally and linguistic diverse backgrounds” as the nomenclature for people from (primarily) non-British, non-Indigenous origins in Australia speaks to the disjuncture in the debate between technology, language and the humanities. The labels matter because they shape our responses and the data supports the labels.

Broad age categories (rather than specific years) are a useful example. Such categories remain routinely utilized in survey tools. Their origins (much like the QWERTY keyboard) remain partly a function of historical necessity. The recognition of the potential value of demographic data emerged at a time when computers were largely human beings utilizing pen and paper. Computation might still have been guided by emerging mathematical and statistical rules but the actual work was done by and repeatedly checked by people, not machines.

Demography was an emergent and highly political science with limited methods at its disposal. Consequently individual ages got collated into bands and these became a normalized part of commercial survey and social science work. Life expectancies have risen over the last century and yet 65+ is still defined as “old” because it aligns with the current allocation of pensions in many countries, rather than an understanding of the biological or social meaning of age, or indeed the implications of these for politics and policies.

This politics of the social has become ingrained in information systems through the categories we use including those associated with social phenomena such as gender, sexuality, ethnicity, disability and so on. These categories are malleable but in an information system they tend to be highly reductive including the field name and the binary “yes or no” common in this particular area. This is one way that past identity politics are carried forward in information systems and analytical assumptions. The need to re-hypothesise social data is essential in this big data era.

But big data approaches don't share these inherited limitations. The long-standing assumptions behind such traditional methods no longer apply. Now we can easily look for ‘natural’ breaks in the data and analyse accordingly or simply analyse *all* of the data that was originally collected or go back and re-analyse old data. It is now clearly possible to collect, store and analyse so much more data than could be done in the past. Our inherited concepts and methods need to change to catch up to the technology instead of reducing the data to the technology we once had.

A major problem with ‘big data’ at the moment is this inherited knowledge many people have ranging from basic statistical assumptions on through to what ‘science’ is, that were built on and around these technical constraints. These limitations aren't shared by big data analytics but the risk is that they persist as normalized knowledge in our information systems and societies.

Not only does data have a socio-political life, but a considerable amount of ‘big data’ is actually social data produced through the growing array of social networking software and systems. De Roure has already suggested that the real nexus of interest is where “**big data meets big social**” in what he calls the fourth quadrant. Many of these social data environments suggest that people can re-present themselves and debate past category limitations. But to be successful, this process needs to transfer from information systems into those bureaucratic systems that frame data analysis. If there are risks to being included in information systems of this size and scope and there are risks to being excluded from them, what sorts of options exist?

These issues are political because there is a real importance to inclusion and exclusion in the information systems of contemporary society. Inclusion means an individual or group gets counted, and as such may have access to resources that those who are not counted generally do not. How you are included or excluded also matters. To be included through an extensive range

THE THINKBELT: THE UNIVERSITY
THAT NEVER WAS

Focus: The difference between feminism
and women's liberation

DISCOVER POLICY PRESS



Bringing social research to
the heart of public debate.



DISCOVER SOCIOLOGY, DISCOVER SOCIAL POLICY



DISCOVER THINKING ALLOWED



of reductive processes has the capacity to make an individual a digital abstraction in information systems – one barely present in the external world. Being excluded means that the person or group can take on varying degrees of risk from the system because to be excluded is not the same as being ignored. Marginalised groups have a long history of being excluded from official counting systems only to have a great deal of, usually negative, attention paid to them by those same systems. This is Phoenix's concept of **normalized absence and pathologised presence** in both theory and practice.

Questions of data collection are a two-edged sword. Dealing with complexity has always meant that decisions were taken to reduce the scope and detail of the data collected and then analysed. However, current arguments about whether or not completing census forms in some countries should be mandatory speak profoundly to the socio-political and economic implications of data collection. As imperfect as the definitions, mechanisms and analysis are, the reality that the most vulnerable are the least likely to have the time, or in some cases the capacity, to voluntarily fill in such forms leads us to a future where inequities can be downplayed for lack of systematic "evidence".

Kitchin, amongst other theorists and critics, argues that **information systems and the socio-technical assemblages that implement them are far from neutral**. In reality they are politics in motion. This is a key point in the growth of big data processes. Their outcomes will rely on the way people are constituted in such systems via their 'datafication'. Big data may offer an opportunity to reform the last century or so of social data collection and analysis by providing a broader and deeper source of data. To be truly "big" in both scope and impact, data practitioners must engage with the politics of categorization and therefore with individuals, communities, and social scientists about the origins and "fit" of inherited and emerging data categories. This will be an important political project in its own right.

References:

Agar, M. (2013) *The Lively Science: Remodeling Human Social Research*. Mill City Press, Minneapolis, MN.
De Roure, D. (2013) *Big Data meets Big Social: Social Machines and the Semantic Web*, International Semantic Web Conference ISWC 2013, Sydney.
Gitelman, L. (Ed.) (2013) *"Raw Data" Is an Oxymoron*. MIT Press, Cambridge MA.

Hamish Robertson is a geographer at the University of New South Wales with experience in healthcare including a decade in ageing research. He has worked in the private, public and not-for-profit sectors and he has presented and published on a variety of topics ranging from ageing, diversity, health informatics, Aboriginal health, patient safety and spatial science to cultural heritage research. Hamish is currently completing his PhD on the geography of Alzheimer's disease and recently finished editing a book on museums and older people. **Joanne Travaglia** is a medical sociologist at the University of New South Wales with experience in the health field as a practitioner, manager, researcher and educator. Her research addresses various aspects of health services management and leadership, with a particular focus on the impact of patient and clinician vulnerability and diversity on the safety and quality of care.



1 COMMENT RESPONSES

LEAVE A COMMENT

Comment:

SUBSCRIBE TO OUR RSS FEED

FOLLOW US ON TWITTER

LIKE US ON FACEBOOK