A Social Framework for Big Data

Socialising Big Data Project Team[1]

In a very short time what was initially referred to as the data deluge, information overload or tsunami of data has come to known as 'Big Data.' While variously defined, Big Data generally refers to digital content stored in social, commercial, scientific, and governmental databases and often generated as a by-product of digital transactions, communications, interactions, and so on. According to a popular definition, what makes this data distinctive is not only its volume but also the velocity of its generation (the speed of collecting data in 'real time') and the variety of data sources and formats (increasing array of data types from audio, video, and image data, and the mixing and linking of information collected from diverse sources). From our perspective these definitions ignore what is most significant about Big Data: its *social* composition and its *social* effects. **This framework asserts that composition and effects are related and proposes that giving Big Data a social intelligence requires acting with an ethic of care.**

The value of socially intelligent Big Data

Big Data tends to be approached in one of two ways (i) in a utilitarian framing, in which data is seen as a resource, and where political and ethical debate is focused on the rights and interests of individuals, corporations, and governments, and (ii) in an abstract framing where discussion is conducted in terms of a general population (nation, society, humanity). So, for example, economic approaches to Big Data tend to think of data as a resource to be mined and often fail to address the possibilities of mobilizing Big Data as a common or social good. The utilitarian goal of exploiting Big Data focuses on producing predetermined end results, and ignores the benefits that can be realized in the movement of data. Adding individualised rights-based correctives to such harms, such as private ownership and control, fails to release the potential to realize the social benefits of Big Data that can only be brought into being through enabling free associations. Abstract approaches to whole populations limit the potential for Big Data as the many different ways in which the data is composed, and the specific connections and multiple unforeseen effects it can make are 'black-boxed', hidden by processes of abstraction. They thus limit the possibilities of participation, which can lead to a sense of agency or empowerment.

If the full value of Big Data is to be realized then we need to find a way to move beyond this binary – the utilitarian individual and an abstract society - and think in terms of its potential to compose new kinds of togetherness, big and small and in-between, and the possibilities they offer for new ways of living. We need to give Big Data a *social intelligence*.

To realize this social intelligence the aspects of Big Data that are most significant are its *social* composition and its *social* effects, and the ways in which composition and effects are connected together in feedback loops. The social make-up and capacities of Big Data refer on the one hand to the ways in which Big Data is composed in relations between data and people and on the other hand to how Big Data has effects through the ways in which its use makes and shapes connections. Big Data does not exist in the raw, it is not simply 'out there' to be collected, a natural resource to be mined. It is generated from the interactions of diverse actors and technologies (digital platforms, mobile devices, sensors, sequencers), it has to be formatted (cleaned, linked, packaged, stored, curated), and analysed (visualized, correlated, interpreted). This composition of data in turn connects myriad distributed people (everyday users and specialist practitioners) and technologies (computers, devices, software, algorithms). This data then generates social effects. Social intelligence draws attention to the feedback loops between data composition and effects and the capacity they have to make Big Data a collective good.

---

[1] This framework is an outcome of an Economic and Social Research Council (ESRC) funded project - Socialising Big Data - that took place from June 2013 to Sept 2014. Further information about the project and research team and the background to this framework can be found in an accompanying document: *Background: A Social Framework for Big Data* and working paper: *Socialising Big Data: from concept to practice* (both avail. at https://sloddo.wordpress.com). Socialising Big Data Project Team: PI - Evelyn Ruppert, Goldsmiths, University of London; Co-Is - Penny Harvey, Manchester, CRESC; Celia Lury, Warwick; Adrian Mackenzie, Lancaster; Ruth McNally, Anglia Ruskin. Researchers: Stephanie Alice Baker, Goldsmiths, University of London; Yannis Kallianos and Camilla Lewis, University of Manchester, CRESC.

For example, during the 2009 A/H1N1 influenza pandemic, search engines queries relating to influenza on Google were used to predict the spread of infection. It seemed as if this new source of data – Big Data - might provide the basis for more timely predictions than traditional sources, but in 2012 the accuracy of such predictions was called into doubt. What was at issue was the status of search queries as data: on the one hand, the constantly changing search algorithm prompts searches in constantly changing ways, and on the other, how individuals respond to the information provided by their search also changes as their interests in, and the kind of flu they are experiencing, also change. While the data used in Google Flu Trends is valuable because it can be updated continuously, the status of that data in relation to a predictive model requires constant recalibration. There is no natural constant to be found; to be socially useful the data must be *socially calibrated*.

Nurturing the social intelligence of Big Data requires an ethic of care

Social relations are complex, dynamic and uncertain, and Big Data is no different. In medicine, genetic profiles can aid in the identification of risk and in turn improve interventions; in social and environmental policy, collective benefits such as efficiency can be achieved through services targeted to particular identified groups or communities. However, they can also lead to potential discriminatory, manipulative, and stigmatising practices; in consumer finance risky groups can be identified and denied credit; in social policy, specific communities associated with particular behaviours can be targeted for increased surveillance. Furthermore, Big Data can undo or devalue other ways in which people get connected or identified. These issues suggest that Big Data has social consequences such as its constitutive (identifying) and distributed (targeting) effects that are typically not addressed by either utilitarian individual or abstract societal valuations. We propose that developing the social intelligence of Big Data requires *an ethic of care* that attends to relations of connectedness and interdependence between persons and data and their qualities. Acting with an ethic of care entails taking responsibility for potential social effects, whether these effects are the negative ones of subordination, oppression and exclusion or the positive ones of empowerment that flow from recognizing the diversity and (reflexive) dynamism of the social relations of Big Data. An ethic of care for Big Data is not about applying universal principles, but acting responsibly in light of particular experiences and actual situations that are not external to but part of the many relations that compose data. This demands responsiveness to each other because without the connections between people and technologies Big Data would not exist.

An ethic of care provides an alternative to the current understanding in which the legal specification of ownership as private property secures corporate or individual rights. An ethic of care calls for an attentive concern for the social relations of data composition and effects, and looks to develop models of *social ownership* that stress sharing, collaborative, or co-operative possibilities. It encourages us to think about stewardship or the curation of data, and to imagine Big Data as a social resource rather than a natural resource. As a social resource, Big Data can be regenerated and enriched in its circulation, rather than diminished by specific and limited relations of property, exploitation and extraction.

Feedback loops between social composition and social effects of Big Data