**A Social Framework for Big Data**

**Background** - as in previous doc, OR as brief note at the end?

**Big Data: Can it be social good?**

**Google Flu**

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. While the data used in Google Flu Trends is valuable because it can be updated continuously, the status of that data in relation to the predictive model requires constant recalibration. There is no natural constant to be found – to be useful the data must be socially calibrated.

**Nappies**

**Security – taxi, pizza, terrorist**

**A Social Framework**

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 however, the aspects of big data that are most significant for its value are its *social* composition and its *social* effects, and the way 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 and on the other hand to how big data has effects through the ways in which its use makes 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 people living lives in which they get flu, have babies and buy pizzas as well as computer scientists, data handlers, mathematicians, platform designers and others in business, government and the third sector) and technologies (computers, devices, software, algorithms). It generates effects. The social framework for data draws attention to the feedback loops between data composition and effects and the capacity they have to make big data a public good for both the economy and society.

Diagram?

**The value of big data is social intelligence**

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, nation states and (ii) in an abstract framing where discussion is conducted in terms of a general population (the nation, society, humanity). Economic approaches to big data that think only of data as a resource to be mined fail to address the possibilities of mobilizing big data as a common or social good. Equally 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 it can make are ‘black-boxed’ and hidden by the processes of abstraction. If the value of big data is to be realized then we need to find a way to move beyond this binary – the individual and society - and think in terms of its potential to compose new collectives, 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.

**Big Data and an Ethics of Care**

Data is plural: its value is its relationality. But the vocabulary currently being used to address big data only recognizes a very few of the ways in which we relate to each other. This is not smart: we need to expand the vocabulary of data relations to develop social intelligence.

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 police surveillance. Furthermore social attachments through 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 utilitarian valuations. To be able to develop the social ‘goods’ of big data and avoid the social ‘bads’, we propose a social intelligence that recognizes connectedness and interdependence in *an ethics of care*.

An ethics of care provides an alternative to the current understanding in which legal specification of ownership as private property (and secures individual rights at the cost of the common good) by attending to the social relations of data composition and effects, and developing models of ownership that stress collective, collaborative, or co-operative possibilities. It encourages to think about stewardship or curation of data, and to imagine big data as a social resource rather than a natural resource - a resource that can be regenerated and enriched in its circulation, rather than diminished by specific and limited relations of property, exploitation and extraction. An ethics of care makes big data a resource for a social intelligence.