**A Social Framework for Big Data**

**Background** - as in previous doc

**The Social Life of Big Data** - first 8 lines of Context very useful.

From our perspective however, the aspects of Big Data that are consistently underplayed are its social composition and its social potential. 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 makes connections. Big Data does not exist in the raw, it is not simply ‘out there’ to be collected. Big Data 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 (mined, visualized, correlated). The process of generating Big Data connects myriad distributed people (computer scientists, data handlers, mathematicians, platform designers and diverse users) and technologies (computers, devices, software, algorithms). The data generated in turn produces a social intelligence, an understanding of social patterns, that in turn feed back into the processes through which the data is generated.

Thus, when we argue that data is social we are referring to the relations that constitute Big Data, and the ways in which Big Data organizes relations between other entities (people, organizations, technologies, devices, other data forms, etc.). From a policy perspective, the promise and the threat of Big Data has everything to do with the on-going relations between these two dimensions of the social (i) the social composition of the data and (ii) the social connections that the data produces.

**The value of a social perspective**

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 social framing where discussion is conducted in terms of a general population (the nation, 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 good. Equally abstract approaches to whole populations limit the potential for big data as a dynamic ‘social intelligence’ as the specific ways in which the data is composed, and the specific connections it can make are ‘black-boxed’ and hidden by the processes of abstraction.

We argue that a focus on the social life of big data can open up a space for alternative approaches to issues of ownership, value, ethics, privacy, and access. Policy debates could take a different approach by thinking about the specific relations that are involved in the composition and use of big data. Our focus on the social life of big data invites considerations of social projects that have a commitment to exploring mutual responsibility, collective ownership, co-operative practice. If the social value of big data is to be realized then we need to find a way to move beyond extractive idioms and think in terms of composing new collectives.

**Big Data and an Ethics of Care**

An alternative to the legal specification of ownership (that secures individual rights at the cost of the common good) would be to develop a more general ethic of care that would attend to the specific relations of composition and effect, and posit models of ownership that stress collective, collaborative, or co-operative possibilities. We might want to think about stewardship or curation of data. Discussions of access could include issues of visualization, of translation and interpretation. The data itself could be imagined as a social resource rather than a natural resource - a resource that could be regenerated and enriched in its circulation, rather than diminished by specific and limited relations of extraction.

To follow through on this more social agenda we would need to mobilise a whole range of idioms for thinking of how we act on data, and for specifying the particular perspectives implied by these idioms. When we discuss privacy, for example, can we distinguish the borrowing, stealing, giving, and loaning of data? When we think about the circulation and re-purposing of data can we keep debate focused on relational effects and possibilities. Social relations are complex 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 and improvements 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 possibly 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 of social goods.

An ethic of care acknowledges that there is never a once and for all solution. A social ethics recognizes connectedness and interdependence.