# Genomics as big data: tensions between social intelligence and an ethic of care

The genomics collaboratory highlighted some major tensions around social intelligence and ethics of care. Genomics could be seen as an advanced form of social intelligence in the sense that the field is organised around and perhaps constituted by the desire to coordinate how people relate to data. At the same time, much of what happens under the name of genomics concerns life and the living. If non-normative obligations to care are concomitant to life (Bellacasa 2012, 198), care as what people do in sustaining relations has to be entailed in genomics. The problem is that the social intelligence that genomics embodies in the form of coordinated composition of data might not align always with the non-normative relationality of care.

What counts as social intelligence and what counts as care may change in and around big data knowledge practices such as genomics. If data composition affects the fabric of experience, both social intelligence and care change. More importantly perhaps, we cannot assume that social intelligence and ethic of care as practices and virtues work together. They may in certain respects be in tension with each other. It may be that the interests of social intelligence and an ethic of care cannot be fully aligned. This difficulty would need to be negotiated politically and ethically. As we will suggest, the genomics collaboratory forms a process in which these negotiations can begin to take place.

# The different senses of social intelligence and the ethics of care

Much depends on what notion of social intelligence and what notion of ethic of care we turn to. Both social intelligence and ethic of care have been conceptualised in quite different ways. There is no ready-made social intelligence or ethic of care that we can bring to bear on genomics. Both terms need some re-conceptualisation to accommodate non-human things and infrastructures.

Social intelligence can refer to an individual capacity to maintain and develop relations with others. In certain accounts, this capacity is linked to the evolution of the human brain in social group settings. The SQ (Social Quotient) measures social intelligence alongside IQ, and standard diagnostic instruments for autiism and Aspergers rely on the concept of social intelligence (Kihlstrom and Cantor 2011). Social intelligence can also mean something like collective intelligence. Social intelligence in this sense refers to the collective effect of many people negotiating the same difficulty. This sense appears in more recent work on social media. The social physicist Sandy Pentland uses social intelligence, for instance, in this sense (Pentland 2014). It is more like 'wisdom of the crowd'. Both senses of the term are difficult to apply to big data and genomics. Can we modify their usages both to account for the infrastructural and computational processes associated with big data? Do they help us grasp a deep re-structuring of experience by data and at the same time the social composition of data?

A similar bifurcation appears in ethic of care. The classic social-psychological account by Carol Gilligan (Gilligan 1990) relates to gender differences in interpersonal relations (particularly in settings such as health and family). It also often embodied a normative moral value. More recent versions, however, diverge in many ways. Some such as Maria Puig de la Bellacasa (**???**), focus much more on human and non-human relationships and identify a deeply ontological register of care. Foucaultian accounts of ethics of care of self tend to address, by contrast, the constitution of particular historical subjects in relation to knowledge, truth and ethical practices. Like social intelligence, ethic of care refers to a range of different objects, and this can make it hard, at least in the context of genomics, to decide what cares for what. In the field of genomics and data intensive knowledges more generally, the implication of care in knowledge is perhaps re-configured in important ways. What would care entail in genomic data?

## Organised around accuracy? Forms of social intelligence in genomics

The field of genomics is an advanced form of social intelligence in both main senses of the term. This social intelligence - whether seen as individual relational skills or collective negotiation of problems -- focuses specifically on data composition in the pursuit of accuracy. Data composition or how data elements cohere in aggregate deeply concerns genomics, since a genome is always an assembled data form (Chow-White and García-Sancho 2011). Every presentation at the genomic collaboratory, and indeed the site at which the collaboratory itself took place, displayed the work of this data compositional social intelligence. The architecture and style of Wellcome Trust Hinxton Campus already points to a social composition of genomic data since it juxtaposes glass-sided bioinformatics laboratories and training facilities, a stately home with dining and bar, and a fairly sleek conference centre, in which the collaboratory took place. The combination of high-tech transparency and log-fire comfort suggests both the epistemic and economic prestige of genomic data.

In the genomic collaboratory itself, presenters were asked to address the metrics of genomic data. Presentations from genomic scientists, clinical research scientists, cloud computing service provides, bioinformatics experts and social scientists displayed, amidst much technical detail, a plenitude of powerpoint slides with diagrams and diagrams concerning how sequence data matters and how different groups of people make sense of it, measure it and value it. Some presentations described large scale genomic sequencing projects, with their often vast, hyper-complex organisational arrangements and collaborations between researchers and institutions. Projects such as the 1000 Genomes Project embody complex awareness of the form of genome sequence data, the economic and technical attributes of information infrastructures and sequencing instruments, and most importantly, the complexity of such projects as a social form of coordination often involving thousands of people. Similarly, presentations on the assembly of plant genomes emphasised the difficulties of knowing in advance what size the sequence data for a given biological sample would be since the DNA sample itself -- typically from a plant such as wheat -- is large, highly repetitive and polyploid (containing more than two sets of chromosomes). The social intelligence of assembling genomic data from such specimens entails many tests, assays, and measures of sequence data in order to ascertain how the genome can best be assembled or composed as a data object. Finally, in the attempts to understand rare diseases in terms of variations in DNA sequences, social intelligence negotiates differences between normal and rare. Genomic research into rare diseases paradoxically pursues the isolated or abnormal variation whose existence could explain a specific condition, but whose rarity thwarts any easy generalization or learning. If social intelligence implies a set of normative practices and relations, the sequencing and elucidation of the reactivity of rare variants in a mass of data implies a post-normal social intelligence.

If we do understand genomics as a form of social intelligence focused on data composition, what would be at stake for it? The grounding assumption for all the work done in assembling genomes in ever greater numbers, sometimes on population-wide scale, concerns the ground truth of DNA sequence data: if sequences can be produced with ever greater accuracy, then the variability, changeability and reactivity of living things, people above all, should become more knowable. Epistemic and economic commitments to DNA as ground truth or bedrock animate the complex projects, the multiplication of metrics and the pursuit of post-normal variations that presentations at the genomic collaboratory described. Sophisticated data and publication infrastructures, an ever more intensive commodification of DNA sequencing as a service and highly developed sociotechnical coordination mechanisms generate and accumulate genomic data. But this accumulation with all its effects of scale (and genomic data is certainly massive) pivots on that grounding assumption, and the investments in accuracy or 'veracity' as one of the presenters termed it.

## Who cares for genomes?

Many of the genomic collaboratory participants were either effectively curators, doctors or scientists working on health or food related research. They voice care in different ways. Some of these of obvious. A clinical research working on rare diseases in childhood has a familiar commitment to care. However, the myriad actions undertaken by participants in caring for data and its relationality suggest that care enters into the very composition of genomic data. For instance, researchers at a genome sequencing facility (TGAC) developed KAT, the quality assessment tool to improve analysis of the quality of assemblies, and undertakes pilot sequencing to assess the nature of genomes before working with them; working at the major UK research centre for bioformatics (the EBI), Laura Clarke & Re-Sequencing Informatics – doing various checks (quantity, quality, identity and consistency) to ensure BD is of sufficient quality for the community to use. Speaking about identity and consistency in data, Laura states ‘these require more specialised methods that are created for, and then applied to, the project in question’. This (and piloting by TGAC) would seem to exemplify an Ethic of Care – ‘An ethic of care for BD is not applying universal principles, but acting responsibly in light of particular experiences and actual situations ….’ This was also illustrated by data curators speaking about the growth of data in one of the main repositories of recent genomic data, the European Nucleotide Archive, ENA: 'the rate of growth must be monitored to assess future sustainability of current technologies, to assist in choice and adoption of new technologies, and to factor in the upper limit to identify and store data' (Rasko).

## Tensions between social intelligence and care in the composition of genomic data

The social composition of data increasingly shows signs of tensions between knowledge (with its values of accuracy, predictivity, validity, and sometimes truth) and care, with its valences of relationality, well-being and sustaining differences. In the genomic collaboratory, we actively sought to draw attention to these tensions through forms of 'dissenting-within' (**???**) by pointing to ways in which genomics as a form of social intelligence invested in the coordinated large scale assembly and analysis of data became the target of economic transformation and reorganisation. Caught between the commodification of sequencing and the increasing yet uncomfortable presence of commoditised data platforms (Amazon Web Services), genomic social intelligence constantly seeks compromises between cost and complexity, between greater knowledge of variation and the potential to do anything about that variation.

The kinds of care for thinking and knowing we saw in the genomics collaboratory were not necessarily morally good. They often took the form of care in the sense of repairing and maintaining. If there is an *ethic* of care in this setting, it might concern those moments when the social intelligence of accuracy encountered questions of for whom or what did this accuracy matter, and at what cost was accuracy going to become the measure of what genomics does. Thinking-with genomic scientists, clinicians, data architects and curators about these tensions is not easy, but it does initiate some forms of relatedness and collective becoming that might yield better compromises in the social composition of data.

## References

Bellacasa, María Puig de la. 2012. “‘Nothing Comes Without Its World’: Thinking with Care.” *The Sociological Review* 60 (2): 197–216. doi:[10.1111/j.1467-954X.2012.02070.x](http://dx.doi.org/10.1111/j.1467-954X.2012.02070.x).

Chow-White, Peter A., and Miguel García-Sancho. 2011. “Bidirectional Shaping and Spaces of Convergence: Interactions Between Biology and Computing from the First DNA Sequencers to Global Genome Databases.” *Science, Technology & Human Values*, February. doi:[10.1177/0162243910397969](http://dx.doi.org/10.1177/0162243910397969).

Gilligan, Carol. 1990. *In a Different Voice: Psychological Theory and Women’s Development*. Reissue edition. Cambridge, Mass: Harvard University Press.

Kihlstrom, John F., and Nancy Cantor. 2011. “Social Intelligence.” In *The Cambridge Handbook of Intelligence*. Cambridge Handbooks in Psychology. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511977244.029>.

Pentland, Alex. 2014. *Social Physics: How Good Ideas Spread—The Lessons from a New Science*. New York: Penguin Press HC, The.