## Social Intelligence

[n.b. this section written earlier as an exploration of how we might be using the term 'SI.' Left here for reference. Main part of genomics section is beneath this section.

n.b.2: our section refers quite a lot to genomics collaboratory – as agreed in the article strategy from – yikes – oct 2015!]

BD is a form of Social Intelligence. This is what distinguishes it from what it displaces or replaces; from other ways of knowing and acting. The SI of BD is constantly reforming or transforming the social fabric and changing what counts as social/natural.

Characteristics of the SI of BD:

* it is a specific mode of knowing the socio-natures. It constructs affinities and likenesses between entities and uses these affinities to associate them – people like you like things like this; things like this like people like you. The SI of BD is somewhat alien to us. The groupings it identifies are new and sometimes nameless. We may not know what they are. Classifications are emergent rather than pre-coded. That may have to be measured.
* it is creating new forms of the social (and the natural). People relate to the new categories of actants (humans and non-humans) that the SI of BD creates by relating people and things. Is like Rabinow’s bio-sociality but generalised to data sociality. (Data nature? data naturality?)
* it is a manifestation of social relations. The SI of BD is the upshot of the actions and inter-relations and collaborations of many different people and things. A redistribution of method.
* it attracts the social. It invites interaction and invention. The SI of BD is often crowd-sourced.

The above properties are what make the SI of BD so generative and sometimes apparently autonomous or even 'intelligent.'. They are also what make it (its composition and effects) so unpredictable. Practitioners in our domains just want to use BD to learn more about how to cure diseases, manage waste, know the population but these new forms of the social emerge as often unintended and unpredicted / predictable side-effects. All of which poses problems for an Ethic of Care.

The SI of BD knows and does different things in each of our domains: [[1]](#footnote-2)

* In genomics, the Social Intelligence of genomics Big Data opens new horizons for *knowing (about) life*. Indeed the life sciences (and life itself) are becoming more and constituted through the Social Intelligence of (NGS) BD. The goal in genomics/the life sciences is to harness SI of genomics to re-configure knowledge of living things in ways that re-configure biopolitical practices and institutions (e.g. healthcare, social care, agriculture, etc.).
* The Social Intelligence of interest in National Statistics is how BD *knows the population* and how well it knows it. The SI of BD from a variety of sources is being used to know populations. The question for ONS is whether and how to use the SI of BD to create official statistics worthy of the name.
* In waste management, the practitioners are trying to master the SI of BD to help them to *know and govern citizens/waste*. Practitioners in this domain are engaging with BD because of its potential to help them meet waste disposal targets.

Differences in the role of SI in the different domains shape their social fabric in specific ways. Each domain has different matters of concern. These specify what could, or should, constitute the Ethic of Care for that domain.

(But we work on the assumption that ethical problems are ones for which we don’t have solutions.)

Genomics and Social Intelligence (as an advanced form of accuracy)

Has the shift in modes of knowing that participants in the national statistics collaboratory referred to with respect to their field has already taken place in genomics?

Compared to the other two domains we engaged with, the field of genomics is a highly institutionalised form of big data science, perhaps even a prototypical one. All of the participants at our collaboratory were already integral to the SI of NGS BD – they are part of this new era. Thus, in contrast to the National Statistics Collaboratory domain, we did not witness debate about the pros and cons of the paradigm shift from older ways of knowing and doing (hypothesis-driven science) to the new mode of Social Intelligence (data driven science). In contrast to NS, our collaboratory participants did not discuss how the SI of NGS BD differs from traditional ways of knowing life/nature. The genomics practitioners at our collaboratory were already actively engaged in and committed to managing, allowing, capturing, exploiting, extending the SI of genomics as a new and generative way to know, and re-know and discover life and living systems.

The taken-for-granted identification with big data amongst genomic scientists does not, however, imply that social intelligence is integral to genomics from the start. Indeed the field of genomics illustrates how big data and social intelligence are not the same thing. Whilst the Human Genome Project explicitly sought to generate big data (Hood and Kevles, 1992), social intelligence only emerged in the field of genomics with the advent of genome-wide association studies (GWAS) and other forms of knowing through differences and likenesses and the rise of data-driven science (Kragh-Furbo et al., 2015). Despite its proximity to disease, healthcare and clinical research, genomics continues to grapple with the problem of ‘socialising’ big data. It constantly seeks to move the focus from big data itself – the many sequenced genomes and the vast streams of DNA-related measures produced by DNA microarrays and other such devices – onto the practices through which social intelligence is achieved in the form of a re-figured engagement with life.

## The collaboratory itself displayed the work of data compositional social intelligence

Genomic social intelligence first of all concerns the coordinated composition of data. The extraordinary measures – ever-faster and smaller sequencing devices, ever-larger population cohorts and scientific consortia focused on expanded or deeper coverage of biological variations – that genomics implements in relation to its data (DNA sequences) seek to do more than curate data. They seek to marshal data in quantities and varieties sufficient to maximise the chances of new differences, new norms and relations coming to light by virtue of the accumulation of data.

The genomic commitment to logistically cutting-edge movement of data (Mackenzie et al., 2015), statistically and mathematically deeply ordered organisation of relations between data (Mackenzie, 2015), and highly-leveraged (in the promissory and mechanical senses of that term) engagements between scientists and particular domains (the hospital, the home, the wheatfield) immediately loom large in any conversation around genomics. In the collaboratory events we organised around genomics, first hand accounts of some of the problems, uncertainties and presumptions of mastering and capitalising on genomic data quickly surfaced.[[2]](#endnote-2)

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 in genomics. 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.

Our participants too, exemplified the diversity of institutions and activities that participate in the social composition of genomics, including: a university genomics facility (Liverpool); sequencing and bioinformatics facility (TGAC); large-scale transnational research infrastructures (EBI); public repository (EBI SRA); commercial cloud computing services (Amazon); commercial data services (Eagle); large scale genome project (100,000 GP rare diseases); public health genomics.

Three main instabilities characterise genomic social intelligence.

### 1. Project complexity and unpredictability

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.

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.

The variable social composition of the genomic social intelligence and its complexity leads to uncertainty. Findings depend on whose bodies are sequenced, which instruments and algorithms are used, and which reference sequence is referred to. Epistemically highly leveraged, the SI of genomics constantly encounters variations and differences whose significance requires careful evaluation:

*Applications and platforms … can produce substantially different results. Part of the problem is that the analytical tools and gene sequencing devices designed to read and process genomic information have different properties, errors, costs and read lengths. These result in methodological differences in terms of content, quality and scale* (Neil Hall, Advanced Genomics, Liverpool U).

This drives the demand for more metrics to take these variables into account. However, it is a relentless task as the SI of genomics constantly drives towards ever increasing project complexity, hence project complexity remains chronically difficult to quantify and measure.

### 2. The amorphousness/intricacy of the non-human

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 a genome can best be assembled constituted as a data object.

*Working with complex plant genomes and large genomes can raise a series of challenges. When working with genomic data, the problem is that the metrics you want to know before conducting your sequencing experiment are often missing. Researchers do not always know the ploidy, heterozygosity levels or genome size, for example, before they start sequencing. In effect, this means that a degree of pilot sequencing is always required to get an idea of the scope and complexity required for the project design* (Sarah Ayling, TGAC).

### 3. Post-normal conditions

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.

## Genomics and Ethic of Care – Who cares for genomes?

As a relatively experimental form of SI, genomics instituted principles and infrastructures that could be viewed as part of an EoC from an early stage. Thus, the field of genomics has co-developed with principles for data sharing (the Bermuda Principles), an international data sharing agreement (International Nucleotide Sequence Database Consortium (INSDC)) and community repositories in Japan, USA and Europe for data deposition. These practices and infrastructures are underpinned by principles of open data and data sharing and deposition. Under these principles big data itself is the object of care; they are about caring for the data.

These principles align, however, with a long-standing conceptualisation of scientific object as stable, finite albeit immensely intricate entity. It stems from , an older promise of total knowledge.

Just as in the case of national statistics, whilst not being abandoned, the underlying data principles of genomics are undergoing a transformation as they are adapted to care for genomics knowledge as constituted through socio-technical relations and intelligence as much as being an intrinsic property of DNA big data. Hence, as well as developing common formats for the data, genomics has elaborated ways to capture its socio-technical composition through community ontologies and metadata standards. Indeed, insofar as it is possible to separate data from metadata, much of what is shared and measured and monitored in the archive is metadata rather than data.

Trouble around the epistemic principles of totality ('keep all the data') take various forms. Cochrane and Birney paper about how to decide which data to compress by drawing on practices in visualisation data?. They are developing a principled way to decide which data can be discarded or compressed.

The rising importance of social intelligence of genomics is not to say that genomics has changed its bedrock position regarding the foundational truth of DNA.

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 sequences as ground truth or the bedrock of variation animate the complex projects, the multiplication of metrics and the pursuit of post-normal variations that presentations at the genomic collaboratory discussed. 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.

As a relatively nuanced form of SI, genomics has already instituted practices and infrastructures to uphold an EoC. Genomics cares deeply about the data but perhaps in ways that uneasily navigate between an ethics of care focused on norms and standards, and a social intelligence focused on the potentials of data to support the emergence of new norms, new boundaries and orderings of significant differences.

Much care for genomic data, about which our collaboratory participants had much to say, seeks to preserve the social fabric that makes the SI of genomics possible. In support of this goal, the field of genomics has instituted practices ('bioinformatics') and infrastructures (archives, repositories, biobanks) that support the sharing of data and exploring the social composition of that data. Thus, the field of genomics has co-developed with principles for data sharing (the Bermuda Principles), community repositories for data deposition (in Japan, USA and Europe), an international data sharing agreement (International Nucleotide Sequence Database Consortium (INSDC)), and community ontologies and standards, including metadata standards for capturing the social composition of big data. There is also the work of curating the data, literally caring for it, so it is fit for future use by others – preparing the ground for the genomic social intelligence to address the future.

Metrics plays a noteworthy role in the exercise of the genomic ethics of care. 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. Some examples follow.

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 – Social intelligence requires a post-normal ethic of care

Seems to be alignment between SI and EoC in genomics. But not always straight forward.

Former practices of governmentality classify entities in relation to the norm.

However, with social intelligence, the statistical mean is no longer a useful parameter. Social intelligence displaces a key measures of statistics – the mean, variance – as ways of knowing the population and meaningful entities within it. These are displaced by machine learning and patterns and clusters. This way of knowing splits aggregates, e.g. breast cancer paper.

Correlations or causations. Lucy.

.

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.

As the SI of genomics expands – resulting in ever increasing flows of NGS data - how can the EoC uphold the principle of data sharing? SRA at the EBI working to anticipate how to always keep the repository doors open.

Tension – Cochrane and Birney paper about how to decide which data to compress by drawing on practices in visualisation data? Data compression reduces its future utility to the SI of genomics.

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 uncomfortable prevalence 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.

## 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.

Hood L and Kevles DJ (eds) (1992) Biology and medicine in the twenty-first century. In: *The Code of Code*, Cambridge MA: Harvard University Press, pp. 136–63. Available from: http://www.vjf.cnrs.fr/histrecmed/publications-electroniques/Leroy%20Hood.pdf (accessed 18 December 2012).

Kragh-Furbo M, Mackenzie A, Mort M, et al. (2015) Do biosensors biomedicalize? Sites of negotiation in DNA-based biosensing data practices. In: Nafus D (ed.), *Biosensors*, Cambridge, MA: MIT Press.

Mackenzie A (2015) Machine Learning and Genomic Dimensionality: from Features to Landscapes. In: Stevens H and Richardson S (eds), *Postgenomics: Perspectives on Biology After the Genome*, Durham, N.C.: Duke University Press, pp. 73–102.

Mackenzie A, McNally, Ruth M., Mills, Richard, et al. (2015) Post-archival genomics and the Bulk Logistics of DNA Sequences. *BioSocieties*

1. *NB these are generalisations; there is variety within each of these domains, e.g. between different ONS, between different cities and countries dealing with waste, between different life sciences.*  [↑](#footnote-ref-2)
2. The collaboratory itself displayed the compositional work of genomic Social Intelligence on data. The SI of BD is the outcome and depends upon the actions and inter-actions and collaboration of many different people and things. It is a redistribution of method. This was apparent in our collaboratory. [↑](#endnote-ref-2)