JRAI Journal of the Royal Anthropological Institute

7
Data – ova – gene – data

Antonia Walford *University College London*

In this essay, I observe that data is valuable not only for what it is, but also for what it will become: that is, that data is a form of potential. I explore two aspects of this by drawing two comparisons with other forms of potential: ova and genes. First, building on ethnographic fieldwork with environmental scientists and technicians in the Brazilian Amazon, I compare data processing with ova donation in the United Kingdom in order to explore how data processing might be considered a form of reproductive labour. I then turn to emergent big data infrastructures in the environmental sciences, and compare the environmental sciences with genomics, in order to gesture towards some critical questions that need to be asked of such open data initiatives. I end with a reflection on comparison as a privileged means of drawing out the forms understood to be latent within data.

Potential

The possibilities and promises of big data have been getting a lot of attention over the past decade or so, be it in academic research, businesses keen on making a profit, or a Euro-American press animated by the spectre of an 'Orwellian state' (Iliadis & Russo 2016; Watts 2018). On the other hand, data also remains conspicuously unspecified: something that is extracted from people without them even knowing, an informational artefact that is both 'raw' and technically opaque, apparently worth billions dollars but valued as such on elusive markets by data brokers who remain in the regulatory shadows. Data is everywhere, but at the same time strangely absent.

I want to suggest in this essay that, far from being an obstacle to analysis, this incapacity to grasp data is actually a function of one of its specific properties, namely that data is understood to be valuable because it can be transformed into something else.[1] No wonder it is hard to get a hold of; and no wonder it is not always obvious what or where data is. The metaphors about big data that are in general circulation are telling here (Awati & Shum 2015; Raicu 2015): data is the new oil

Journal of the Royal Anthropological Institute (N.S.) **27**, *127-141*

(*The Economist* 2017), the new gold (Peck 2017) – a resource from which wealth can be extracted, a raw material to be 'mined' (see, e.g., Lupton 2013; Puschmann & Burgess 2014; Seaver 2015). But aside from what the specific terms of these metaphors are, it is interesting that there has been so much metaphorical work around big data at all. This is not coincidental. Metaphors work exactly through the capacity for one thing to be – partially at least – something else.

If data is often coveted exactly because it is understood as something that will transform into something else (even if 'there's no such thing as raw data', Gitelman 2013), another way to say this is that data is constituted by its potential; what Marilyn Strathern has pointed to in a different context as a 'capacity for development as yet unrealised' (1996: 17). In fact, taking my lead from Strathern, I want in this essay to suggest that data is in some aspects very like other forms of potential property which are 'difficult to recognise' (1996: 17) in this way: ova, genes, ideas – amongst other forms.

There is a third way in which data metaphors matter, then: it matters what ideas we are using to think data through. This essay is an experiment in thinking data not through metaphors of resource extraction, but through other forms of transformative potential, specifically ova and genes. I will be building on another of Strathern's astute observations regarding the analogical relationship between kinship and knowledge – how, in the English imagination at least, they have always served as resources for thinking each other (Strathern 1995; see also Maurer 2015). However, moving away from metaphors of resource extraction and towards alternative ways of recognizing data is not to move away from the questions of value, appropriation, and rightful ownership that animate current public data debates. On the contrary, it is exactly the entanglement and disentanglement of what we might call 'production' and 'reproduction' that I want to focus on in order to unravel the ways in which data comes to be valued as a form of social relation that has not happened yet.

The essay is structured around two comparisons which allow me to refract the idea of data's potential in two different directions. Drawing on my own ethnographic material from fieldwork conducted with earth system scientists in Brazil, in the first of these I compare the work of small-scale scientific data processing with ova donation in the United Kingdom in order to examine potential as inhering in the tension between what something is and what it will become. In the second, I compare the emergence of transnational big data environmental science with genomics, focusing on how the idea of open data functions in each as a form of potentiality that rests on the tension between the possibility of becoming anything, and the necessity of having to become something. Heeding the remit of this special issue to demonstrate not only how anthropology can contribute to our understanding of 'the data moment', but also how taking data as an ethnographic object might ask us to analytically reshape anthropology, in the conclusion of the essay I reflexively focus on anthropological comparison as a revelatory technique. Comparison has traditionally been understood to work in anthropology by bringing together disparate sets of materials in order that they might show each other in a different light (Strathern 2005; Walford 2021; cf. Candea 2018). Here, I explore the idea that, rather than bringing together elements understood to be distant or previously unconnected, comparison can also work by disarticulating or disentangling internal aspects of an entity or phenomenon that are understood to be imbricated in each other – and as such, I argue that comparison is a privileged technique for drawing out the latent forms that 'data' carries within it.

Data means, data ends

The first comparison I will be making rests on ethnographic work completed with the Large-Scale Biosphere-Atmosphere experiment in Amazonia (LBA), a long-term international scientific programme led by Brazilian researchers at the National Institute for Amazonian Research (INPA) to investigate the role of the Amazon forest in the global carbon cycle. To this end, the LBA has numerous experiments set up throughout the Amazon region, which are collecting large amounts of data on all sorts of variables, from carbon flux to hydrology to photosynthetic activity. The LBA has built several meteorological towers, as well as a 300 m tower in order to measure CO_2 and different sorts of trace gases and their flux, including carbon dioxide. It does this in real time, with the data sent to the HQ in Manaus, and then stored or disseminated to the wider LBA community. There are many different scientists from different disciplines working in it at any one time, as well as a basic team of researchers, technicians, and students. I spent around thirteen months in total with the LBA, conducting ethnographic research, in 2007, 2010, and 2011.

At the time of my research, within the LBA an economy of exchange had grown up centred on the collection and movement of data. The LBA provides infrastructural support for foreign (*estrangeiro*) – non-Brazilian – researchers to set up data collection campaigns in the Amazon, using the LBA's basecamp, paths, quad bikes and the labour of the technicians, often in return for either sharing that data, giving lectures, lending instruments, supervising students, or putting names on published papers; and this data is also then taken back by foreign researchers to funders in their home countries, and used to request further funding, set up new projects, and so on. Although each of these exchanges had its own specific duration and features, most of them were aimed at obtaining or collecting, in the end, more data. As one researcher told me, there is always too much data, but also never enough.

What became clear to me over the course of my research was that there were (at least) two different conceptions of data that people were working with. Very early on in my fieldwork, I had a conversation with a well-known researcher who impressed upon me in no uncertain terms that he considered the flow of data to be in one direction only – from Brazil to the rest of the world – with very little recompense in return. When I asked why, he said:

> Yes, some people say, 'Oh, the data in itself doesn't have any value. It needs a certain amount of intellectual work'. But data has a value. There's an enormous amount of work done on the data already – collection, treatment, and so on – each part is not a single thing … The data can be very different in a complex environment. To install the instrument is already a lot of work, as much intellectual as any other.

On the other hand, the head of the LBA database and IT department who was responsible for the LBA data management had a different idea of data. He told me:

> Data, yes, it generates knowledge … It's extracted, by the sensor, or observations … but looking at this data, it doesn't make much sense … you have 40 MB of data from an experiment but it won't give you any scientific answer. What is there in this data?

The difference he was pointing to, he told me, was between 'data and knowledge. Data doesn't tell you anything. I can get the data and analyse it, but the number isn't going to tell me anything. The data is just a number, it doesn't have a nature (*natureza*). It can be in various scales. The number is not knowledge'. In fact, he explained to me, the point of a database is to keep the data in constant use by someone, anyone – the more used it

is, the better kept the data is and the less 'entropy' sets in; the more your data circulates, I was told, the more 'visible' and 'valuable' it is.

Between these two positions, then, we see that data emerges quite differently. On the one hand, the labour invested in the product of data means that it already has value, and it can certainly belong to someone such that its collection demands recompense – here, data is an end in itself. On the other hand, data's value is exactly in its capacity to be constantly made into something else, and to be kept in constant circulation so that it can be used and reused by as many people as possible – data as means, here a means to 'knowledge'. Where, on the one hand, there is an awareness of the value of data being in what it already is, on the other, it is in what it will, endlessly, become.

Another way in which this tension appeared, more dramatically, was through the idiom of property, specifically through suspicions that data had been 'stolen' (*roubados*). Several members of the LBA knew someone whose data had been unfairly taken or stolen by a foreign research partner: that is, either not shared during a collaboration, or published without any forewarning or acknowledgement of the partnership. These stories of data theft sometimes drew on other forms of theft, like biopiracy, and seemed to be a concern for several of the researchers I spoke to. As one told me: 'Imagine, I've got some raw data, and I'm not working on it. And then someone comes along from Scotland[2] and uses it to write something – you can't do that, that's stealing … Officially, I'm the owner of the data'. At the same time, other researchers expressed confusion as to accusations of stealing; yet others suggested that there were only certain conditions under which stealing data was at all possible. When I asked one researcher based in the United States, for example, about the accusations of stealing data, he was a little nonplussed. 'You can't really steal data like that', he said. This was because the data is free – it belongs to everyone, he said, and he told me how surprised he had been that he had such an unexpectedly hard time getting data from certain institutions in Brazil. Another researcher told me that although he might be considered unconventional for it, he would not have any qualms using someone else's data from an online repository – if it is available online, as far as he is concerned, it's free, and he'll use it whether they had published or not. I was often told it was not really possible to steal data because the data's collection was funded by government bodies or taxpayers' money; or that the data itself didn't have any value, data was just data; or that the person who had collected the data wasn't using it, it was just sitting there doing nothing. The conflict here is not over whose rightful property the data is, but whether data should be considered property at all (see Walford 2012).

The appropriation of data from the Brazilian Amazon, as the comparisons with biopiracy indicate – and especially under the auspices of common knowledge and unrealized property rights – places data clearly within an extractive relationship that characterizes science as a colonial enterprise (Safier 2008; Walford 2018). Rather than simply a reiteration of historical inequalities, however, this relationship is a complex lived experience for those who work with the data, and different inflections of it could appear in any particular researcher's practice or speech, whether Brazilian or *estrangeiro*. Data, in this context, is something valuable in its own right, an object of personal property in part because of the labour of its production; but at the same time, its value is also framed as being exactly because it is common property, and as such can, and should, be continually transformed into something else. Data is both end and means, simultaneously. Although as I have said, these two ways of understanding data could appear together, I want to hold them apart for now as a useful heuristic. This is

because whereas the former, what we might call 'data as ends', is comprehensible within conventional frameworks of labour and ownership, the latter – 'data as means', or data as the potential to become something else – is more difficult to recognize and therefore requires further analytical attention.

In order to give it this attention, I will now turn back to the LBA data 'economy'. Although I initially described scientific data as an exchange object, or an item of barter, that circulates within a singular micro-economy at the LBA, there are in fact different ways in which different LBA data circulates. In the case of campaigns or one-off collaborations, the researchers in question generally collect and clean their own data; they are formally governed by data policies depending on what funding body finances them, which means they have a limited amount of time before they must upload their data to a public database. However, in terms of the informal data economy I described earlier, they can choose whom to share it with and how. Under almost any policy, they are given the chance to publish with it first: that is, to transform it into something else, normally an article or what the database manager above might call 'knowledge'. It is very rare for anyone to share their raw (unprocessed) data (*dados crus*). In the case of the LBA tower data, on the other hand, which is long-term monitoring data, the raw data is collected from the instruments in the forest by technicians, or sent in real time to the LBA HQ in Manaus. It is then processed by the LBA data curators or cleaners and turned into certified data (*dados certificados*), before being made available to the wider LBA community, either on request or automatically. Unlike the individual data collection drives or campaigns, this long-term monitoring data does not officially 'belong' to anyone – the LBA has to be acknowledged when the data is used, but the language of ownership is expressly avoided when talking about it.

However, although the LBA data was often framed, both in terms of regulation and in everyday discourse, as belonging to no one, in the sense that no person or group has exclusive property rights over it, the situation was not as clear-cut as that. This was in part because there are of course people whose labour goes into this LBA tower data, preparing it for others to use. Joana and Raquel,[3] responsible for 'cleaning' (*limpar*) the data, spend hours with the data from two of the LBA's towers, teasing out the errors and flaws in it. They painstakingly go through each variable in the data, trying to piece together whether the errors in the dataset are due to a problem in the battery supply, an unfortunate frog that met its fate in a pluviometer, or a lightning strike. They deal with requests for the data from researchers all over the world. They are paid to do this labour, and so from one perspective, any relation of ownership they might feel towards the data because of the fact they have worked on it, in the Lockean sense, is severed through monetary exchange (Locke 1980 [1690]; Strathern 2012). As Raquel told me, 'Look, I … no, I'm very clear. I don't have any link to this data. Whenever [the head of the group] or anyone says that I do, I say – no, that this doesn't belong to me. This is not my property'. Joana and Raquel exchange their work on the data for money so that *others* may use the data – as several people told me, 'It's just their job'.

But severing their relationship to the data is not that easy. In fact, even though the data cleaning is 'their job', I discovered over time that Joana and Raquel did still feel some sort of claim over the data, a sentiment that was made apparent during one trip to the forest. Several people were discussing a recent field trip by students and researchers from the LBA and other Brazilian institutions, and from a university in the United States, who had come to the tower to learn about data collection campaigns. In the discussion, people made comments like the following: 'Well, they'll write an

article quickly from that data. No one considers that someone worked on that data, or will include her', and 'The data is used so much but no one includes who treated the data … if everyone did, Raquel's CV would be amazing!' Although it had been made clear that the visiting group would not publish anything from the data without permission, that was not enough to satisfy the group. Their concern was, on the one hand, with the value of the data itself, not just the knowledge that would emerge from it; but, on the other, their comments also point to an understanding of data's value well beyond economic remuneration, and rather in terms of professional and social recognition. Raquel's labour on the data was eclipsed, and her name did not get attached to the data. She was, in a sense, made anonymous in order for others to be able to use the data to make themselves known.

The problem of eclipsed labour is of course well rehearsed. Here, the fact that Raquel is paid means that she is excluded from the benefits of her labour: she is alienated from the thing she produces; she does not get the credit she deserves. As the hydrological researcher explained to me, the concern is with data as a valuable object that can travel well beyond its context of production, and the lack of fair recompense for this; it is exploitation. However, in the discussions between the technicians, data also seems to oscillate between being both alienable and inalienable. The scientific knowledge economy is extremely complex when it comes to authorship and ownership, as has been well documented (Biagioli 2008; Biagioli & Galison 2003); and this oscillation hints at the way in which data does not sit easily within the parameters of an alienable/inalienable framing. This is because Raquel's labour is in fact towards producing a specific sort of phenomenon: *potential*. To argue that Raquel is being denied her just deserts requires data to be understood as something of inherent value already as data. But the other side of data's value is held within the idea that it is not yet what it will become. What is in fact eclipsed is that, far from being innate, the crafting of this potential, too, requires work; what is missing is an analysis of the labour that goes into producing something that is not yet anything – what might be called the reproductive labour of data.

Data – ova

There are in fact two senses in which Joana and Raquel might be said to be performing a sort of reproductive labour. In the first sense, it can be argued that their paid work is directed exactly at the social reproduction of the LBA community. Such an analysis rests on decades of feminist scholarship around the relationship between waged labour and reproductive labour: that is, between the formal, public labour market and informal, private, or domestic forms of work (see Duffy 2007 for an overview). Within this body of scholarship, the distinction between the formal and informal economy, or productive and reproductive labour, is being continuously dissolved and refigured (see Bear, Ho, Tsing & Yanagisako 2015; Dalla Costa & James 1972; Hardt 1999), such that it becomes clear that formal or productive labour can have aspects of intimacy, affect, and nurture, and, more generally, can reproduce forms of sociality. In this sense, the 'paid labour' itself of data cleaning – that is, the production of data as an end – has socially reproductive effects. The data exchanges and negotiations going on all the time at the LBA are the material and specific means by which the LBA community – and sociality – is generated (Walford 2012; cf. Hilgartner & Brandt-Rauf 1994). I described this earlier in the essay as an 'economy of exchange' featuring data, names, instruments, access to research sites, and from which collaborations, partnerships,

friendships, enmities, and institutions emerge. Joana and Raquel's work is thus not only about producing data as a knowledge object, but also about producing data as a form of social relation – as a way to build and sustain the LBA community. Furthermore, Joana and Raquel developed an intimate, embodied relation with it – what they called a 'feeling' (*feeling*) for the data. As I was told, they had to learn how to 'woo' (*namorar*) the data, get to know all its idiosyncrasies and quirks. Although they rarely visited the towers themselves, they would get to know the capricious instruments, and they would pore over instruction manuals for the notoriously troublesome dataloggers, the small computers on the towers which store the data ready to be downloaded. Through this intimate work, then, it can be argued that Joana and Raquel are quite literally providing the means by which scientific communities are made and constantly remade, as data circulates between people (or does not), acting to suture them together (or drive them apart). Shifting into an understanding of the socially productive effects of data labour is the first step towards expanding our understanding of what the value of the so-called data economy is.

But there is another form of reproduction hidden within this one. Beyond this general sense in which Joana and Raquel's work with data is socially generative lies a much more specific way in which the work they do is about reproducing personhood. Joana and Raquel are producing a very particular sort of entity or effect: potential. This often gets overlooked in debates about data's value, because there is a presumption that data has to already be 'something' to be valuable, socially or otherwise. Data as means – potential – then gets folded into data as ends.

In order to disentangle data as means from data as ends, I want to propose a comparison between data and an apparently very different form of reproductive potential: ova. Specifically, a comparison between Joana and Raquel, the data cleaners, and British ova donors, as investigated in the work of Monica Konrad. Here, too, there is a tension between what something is and what it will become, but it is framed as that between ova and the people they might grow into. This comparison allows me to shift into a different idiom to that of labour, ownership, exploitation, and economies; it allows us to think of the value of potential through ideas that have already been hovering at the edges of my analysis and which are often contrasted with the language of economy: kinship and the making of persons.

Konrad's study examines how British egg donors, as 'partible persons', detach parts of themselves so that their agency can circulate, making these parts of themselves into 'substance for others' (1998: 645). This detachment is very clear in the way they talk about it: in the words of one of the women Konrad spoke to, 'I don't think the eggs are mine, they're not something physical that they're my eggs. I don't even think of them as eggs'; others told her, 'It's not the eggs that are the actual thing, they're not like a physical thing that have come from my body', or 'They're just like a fingernail or something … they're just a normal part, like any other part' (1998: 650). This detachment, Konrad argues, is crucial to the efficacy of their social agency and to the nurturing of a certain kind of kinship. The egg donors she speaks to are often unable to put into words the relation that they feel towards the women they donate their ova to; but it is exactly this 'diffuseness' that 'is the form of the relatedness making up the connection' (1998: 650). In Konrad's analysis, the donors' circulating ova thus become a form of 'redirected social relation' (1998: 655), but a relation that is governed by anonymity, because their identity will always be kept a secret from the women in question, and their children. Konrad thus develops a notion of what she calls 'transilient kinship': a 'spiritual kinship

where connexions and disconnexions between persons are to be recognized as forms of extensional relatedness dispersed through multiple (exteriorized) persons. These persons cannot always be located, or even nameable, and, most importantly of all, do not have to be grounded in specific, discreetly bounded persons' (1998: 659). Their anonymity, the open-ended nature of their donations, is at the basis of this form of relatedness.

What the egg donors do not lay claim to is the future of the egg they produce; they might start the process off, but another person takes it on. It is this very diffuseness and anonymity, this eclipsing of their agency as the origin of the process, that generates the sense of kinship they feel. But also important is the way they relate to the ova. Understanding ova as ends, as the final product of their (bodily) work which someone else gains from, would allow us to argue that these donors are being exploited – donors are not paid for their donations, for example;[4] but it would hide from us that what generates the affective attachment that animates the donors is the fact that their ova are constituted exactly by being 'not yet' (persons).

Returning to Joana and Raquel, they, too, give away potential without any claim to its realization. Here, too, 'what signifies is being the origin of a process that another carries forward' (Strathern 2003: 185). The data they produce is 'not theirs', they say, or only theirs incidentally – it belongs to those who will go on to transform it, for example by writing an academic paper based on it. Furthermore, as we have seen, their work on the data allows it to travel well beyond them and, in so circulating, elicit an LBA 'community'. If my comparison turns implicitly on the fact that both data and ova are in a sense difficult to recognize, however, it also demonstrates that there are different ways, or effects, of being unrecognizable. Konrad can argue that the unrecognized position of the British egg donors in her study – their anonymity, the inchoate forms they donate – generates a sense of diffuse, affective attachment. But although in one sense Joana and Raquel know they are instrumental in sustaining the LBA 'family' (a term I heard used several times), the effect of producing the means of reproduction is very different for them. Gametes hold within them the potential to turn into people, but data can also 'make' certain kinds of people, albeit in a very different way: the data cleaners are also reproducing the potential for (other) people to grow. The different ways in which data is transformed into knowledge contribute substantially not only to making 'humans into particular kinds of subjects called scientists' (Haraway 1997: 142), but also to making them into particular kinds of scientist subjects. This is a notion of a scientific subject based on the fact that what scientists produce gives them the ability to relate in certain ways, and it is this relational agency that allows them to grow as researchers, to amass value to their name, to make the claims that they do, to circulate. What they grow through is the ability to transform data, to take the potential and realize it.

However, this access to data's potential was denied to Joana and Raquel themselves. At the time of my research, they both had master's degrees, and they both expressed an interest in doing a Ph.D. Both of them had been struggling during the period of my fieldwork to find the time to write an article based on the data from the towers. Their frustration was quietly held, and rarely surfaced explicitly. If I asked them about it, they often avoided the question, but it was clear that they felt in some way stuck where they were, 'doing the work of technicians', as one person put it: that is, without the chance to turn the data into knowledge. The 'technical' work they do (*trabalho do técnico*) is understood to be simply a means to something else more important; they have master's degrees but no time or support to do a Ph.D. or to publish themselves, and as

such, following a sentiment I often heard, 'at the LBA, people with master's degrees do the work of technicians' (*No LBA, quem tem mestrado faz trabalho de técnico*). It was often aired as an expression of frustration, and it signifies the difficulty in classifying or categorizing Joana and Raquel as scientific subjects; they did not see themselves as technicians, but neither were they researchers. They are therefore stuck in what might be thought of as an 'unrecognizable' position – in the sense of unfamiliar, but also in the sense of unrecognized by others. Whereas in Konrad's analysis, it is through the nameless flows of donated ova to unknown others that 'women themselves can make their procreative powers into an ovular economy of intersubjective (cross-corporeal) agency' (1998: 653), what Joana and Raquel experience is rather different: a sense of truncated growth and a lack of agency. The data bestows upon them the potential of relating, but they cannot fully transform it (cf. Walford 2019). The promise of potential's endless transformativity is also always the possibility of not transforming at all.

Approaching the social world of the ova donors through a framing of labour and ownership, or alienability and inalienability – how can a person detach from themselves in this way? we might ask – obscures the generative effects of their reproductive labour, and the kinship that their anonymous donations reproduce. Likewise, trying to contain Joana and Raquel's relationship to data within the frame of waged labour, understanding data as an alienable object, is equally insufficient; but what it obscures is the extent to which a form of agency and personhood which they give to others is denied to them. Understanding data as a form of reproductive potential, as a means for persons to grow, is necessary to understand better not only how data has the value it does, but also how highly valued those who work with data, often in ways that are ignored, should be.

Opening data out

If in my first comparison I was interested in holding apart data as means and ends just long enough to allow a different set of concerns to emerge, in this final section it is another aspect of data as potentiality that I want to explore: not the tension between what data is and what it will become, but the tension between the possibility of becoming anything at all and the necessity of becoming something in the end – between potential and its realization.

I will be moving away from the LBA to look at what feels like a very different scale of data practice. Across the sciences, there has been an increasing imperative to operate policies of open data, where the data collected is free to anyone who wishes to download it. This pervasive and powerful imaginary of 'open data' is not new in the sciences by any means, and can be attributed to a number of different initiatives that resulted in widely followed guideline and principles.[5] Several scientific funding bodies make continued funding reliant on putting your data in public databases: this has been the case for some time for projects funded by the National Science Foundation in the United States, for example. In the environmental sciences specifically, there is emerging what might be thought of as a new data 'paradigm' (Gabrys 2016; Hey, Tansley & Tolle 2009) that relies heavily on the promises of open data. A striking example of this is the Group on Earth Observation (GEO), a partnership of 103 nations, the European Commission, and ninety-five participating organizations 'that envisions a future wherein decisions and actions for the benefit of humankind are informed by coordinated, comprehensive and sustained Earth observations and information'.[6] To this end, the GEO community is creating a Global Earth Observation System of Systems (GEOSS).[7] On one of the original demos on the GEO website, called 'GEO in Action',[8]

we are confronted by the slogan 'Countries have borders. Earth Observations don't'. As I have written elsewhere (Walford 2018), the imaginary of the GEOSS frames data as an endless flow that observes no barriers and is in constant circulation. As the press statement for the GEO 2015 summit in Mexico City reads: 'GEO commits to unleash the power of open data to address global challenges'.[9]

However, despite its pervasiveness, as several scholars have pointed out, there are problematic assumptions woven into the implementation of open data systems. There are issues of differential access and different technological capabilities: that is, how open data actually is tends to be dependent on infrastructural capacities that the Global North takes for granted (Bezuidenhout, Leonelli, Kelly & Rapper 2017). As we have seen in the previous section, open data policies also tend to eclipse the labour that goes into the production of that data, as if it emerges already as part of the commons. Several of the concerns about open data in the sciences intersect partially with debates in other domains about accessibility to data and transparency. Although these latter often have more to do with changing forms of governmentality and surveillance (Ruppert 2015) and the recursivity and performativity of audit cultures (Carrier & Miller 1998; Strathern 2000), there is also shared concern that open data might be a proxy for other, more ominous practices: for example, how calls for open data, particularly in the environmental sciences, might be used as a form of political obstacle to environmental policy reform (Levy & Johns 2016).

With these critiques in mind, I wish to pause here at what it means for data to be 'open' in the first place. Again, there is something about data which lends itself to being opened up in this way. Several of the factors we saw in the ethnographic material I presented previously (at least from one side of the debate) play a role here: that data is 'not yet' knowledge; that it has been publicly funded; that it is a resource that should belong to everyone. But there is also a sense that, with the right infrastructures and platforms, the data could become anything; opening data is aimed at multiplying the possibilities presumed to be contained within it. Opening data is, as the GEO put it, 'unleashing' its power – understood, until then, to be hidden, inert, or contained. The idea of openness thus carries with it a number of implicit aspirations, ambitions, and understandings. In the next section, I want to use a comparison to make some of these explicit; and to think about what might happen when the openness of open data comes up against its realization – when anything become something, and how that is decided. This necessitates we shift from thinking of data as means, to questions of meaning.

Data – ova – gene

To end, I want now to turn to another science that has seen an explosion in data – genomics – in order to gesture towards where this argument might take us. Jenny Reardon (2017) has charted the development of genomics over recent decades, from lab-based small-scale genetics, where people collegially pipetted DNA samples into micro-arrays, to the massive inhuman infrastructures of automated sequencing producing petabytes of data that characterize genomics today. What she is particularly interested in is what happened after the Human Genome Project (HGP) was completed. In 2003, geneticists, faced with the map of the human genome, were also faced with the question: now we have it, what does it mean? It is this turn to social meaning that Reardon flags as characteristic of her object of inquiry: the *postgenomic condition*. Reardon points us towards the realization that dawned on geneticists that mapping the genome was not going to provide the sorts of answers they expected. Instead, she says,

a much richer terrain opened up, made up of spaces where the 'meanings and values of the genomic data are being forged' (2017: 4), or, as she later puts it, 'in-formed'. The absence of meaning therefore simultaneously seemed to engender a potential excess of it.

Particularly relevant for my concerns here is Reardon's description of what she calls 'genomic liberalism' (2017: 7), the way the genome has come to stand for, or 'mean', a set of liberal ideals, such as participation, openness, transparency, and inclusion. The genome was framed as knowledge for all 'mankind', revealing 'our common humanity' (2017: 8) and often touted as heralding the 'end of race', in a deliberate attempt to move away from its roots in eugenics and the race controversies that dogged twentieth-century genetics (see also M'charek 2005). Importantly, data and data infrastructures were a crucial element in constructing this new face of genetic science, which drew on 'turn-of-the-millennium hopes for justice through data and democracy' (Reardon 2017: 8). But exactly how those hopes were to be realized became the sticking point. This is because, Reardon tells us, data is not meaning (2017: 174). What Reardon finds herself exploring, then, is not so much the information of life, but the life of information: 'Rather than reveal meaningful knowledge about life itself, genomics instead has given life to a deluge of data. How to make anything of value out of this data is now, quite literally, the million dollar question' (2017: 13).

What Reardon returns to throughout her description is what happens when the open-ended potential of the genome comes up against its own realization: genomes contain/become/already are specific persons. The quest for a generic human genome that could be shared by all was thwarted again and again by this realization: by the 'particular bodies, histories, and constraints of the people asked to provide their genomes' (2017: 21). That is to say, the postgenomic condition is characterized by the conflict between infinite potential and its necessary manifestation. This moment of tension is generative in Reardon's account. She examines how the hopes contained within genomic knowledge hover between different sets of meanings, and the conflicts this engendered – and in so doing, she shows how the genome is an ongoing site for the playing out of this dynamic: the opening up and closing down of meaning.

There is then a double imaginary, a resonance, of potentiality here. Although, as Donna Haraway has argued, genes are in one sense 'another kind of thing, a thing-in-itself where no trope can be admitted' (1997: 134) because they reduce life to a lowest common denominator, it is nevertheless true that genes are also understood as a form of potential: as the blueprint for life, they function through endless possibility for recombination, and, of course, they contain persons within them. Genes, then, both close down and open out meaning – it is not surprising that understanding people genetically led to both eugenics and the means to push back against it by allowing for the possibility for alternative futures, however utopian they may sound (the 'end of race'). And data in turn is also both decried for being reductionist, and pregnant with hopes for alternative political forms – openness, as we have seen, being one of them.[10]

Drawing a comparison with genomics as Reardon describes it, we might say that the environmental sciences are likewise entering into a 'post-environmental condition'. The 'environment', like the genome, is a potent symbol which has shifted its signification over time, carrying with it in its singular form ('the environment', as with 'the genome') notions of a shared commons, a 'global consciousness' that overcomes local concerns (Jasanoff 2001). Like genomics, in the environmental sciences we now see technological revolutions ushering in enormous amounts of data; partnerships with

private companies and Big Tech; and the flourishing of liberal ideals of openness, democracy, and freedom through information – 'Countries have borders. Earth Observations do not', as the GEO tells us confidently. But a comparison with Reardon's account also forces us to ask about what all this data means, and for whom. This is not an obvious question from the perspective of open data initiatives, which are aimed precisely at deferring such questions endlessly. Contrary to Reardon's claim, it is not so much that data has no meaning that is the problem; rather, it is that its meaning lies exactly in its lack of immediate interpretation, its potential, and its capacity to carry latent forms within it. An unreflexive commitment to open data, therefore, does not allow us to ask whom exactly this data is benefiting and how. The environment in this new data-driven landscape is a site for both the opening out and closing down of new meanings, commodification by big tech companies, reinscriptions of colonial hierarchies, as well as claims for grass-roots environmental justice, via citizen science initiatives, for example (Gabrys, Pritchard & Barratt 2016). Like the genome, there is suddenly a space for new or different meanings to be forged. How to make sense of this 'post-environmental' moment is exactly what now requires careful and critical analysis.

Data – ova – gene – data

Anthropological comparison is often understood as a practice of bringing together different sets of material so that one might cast a different light on the other (see Candea 2018 for a comprehensive account). So in this essay, for example, I have developed a comparison between data and biological reproductive material in order to move away from what feels like an obvious analogy between data and natural resources; I make a comparison with reproduction rather than production, one might say. However, another way to understand comparison is almost the opposite to this: that it allows us to disentangle analytical elements that have become seemingly inextricably tangled up in each other. For example, here my comparisons are also intended to draw out the way that 'production' and 'reproduction' already participate in each other. Take my comparison between data and gametes, which relies implicitly on the way in which knowledge and kinship are always already implied in each other, at least historically in Euro-American understandings: the 'repeated echo between intellectual propagation and procreative acts' (Strathern 1995: 8). But such entanglements can also be seen in the resonances between scientific framings of reproductive processes and capital extraction (Martin 1992), or in the way that gametes and other forms of tissue like stem cells are increasingly understood explicitly as resources of biovalue and biocapital (Franklin & Lock 2003; Rajan 2006; Rose 2006); or how genetic imaginaries are understood through reference to ideas of information or code. That is to say, it can seem as if gametes *already are* capital, or genes *already are* knowledge – they somehow contain these forms latently within them, as Stefan Helmreich has argued in the context of the biotechnological imagination, in which 'biological process itself already constitutes a form of surplus value production' (2007: 293). In the material I have analysed, data and datafication are both explained by and contribute to this effect: data is understood to already contain hidden forms within it, and transforming phenomena into data is also to impute to them this potential for future realization and transformation. My comparisons in this essay are intended to make these imbrications explicit by disarticulating different imaginaries of data that are otherwise hidden inside itself. In this way, the comparisons turn data inside out – not by bringing data into relation with totally different imaginaries, but

by holding the elements contained within data apart for just long enough to make it differently recognizable.

Comparison has therefore worked in my description to disentangle different aspects of data. Data, on the other hand, seems to, potentially at least, contain almost anything within it. This might be why it is so easily understood as an analogue to currency, or money (e.g. Eggers, Hamill & Ali 2013; cf. Holbraad 2005). This has led to critiques of data that push back against quantification and datafication, on the grounds that there are things in life that cannot be contained or captured in this way (much like, of course, the critiques of money or monetization). My approach, however, proposes a new form of critical engagement with data. Data's power comes not just from voracious encompassment, in the way capitalism is understood, but also from its *reproductive* framing, how it is understood to contain other forms already within it. It is this we must also learn how to question.

NOTES

I would like to thank the two anonymous reviewers for their supportive engagement with this essay, as well as those others who read it and generously commented: Alice Elliot and my co-editors Rachel Douglas-Jones and Nick Seaver. The ethnographic research on which this essay is based was funded by the IT University of Copenhagen.

[1] As when we are told cryptically in a *Harvard Business Review* article from 2013 that 'the value of big data isn't the data' (Hammond 2013).

[2] A deliberate reference to where I grew up.

[3] I have used pseudonyms throughout on the request of my interlocutors.

[4] In the UK, it is illegal to pay for egg donation (Cooper & Waldby 2014; Stoeckle 2018; see Strathern 2012 for a similar discussion around organ donation).

[5] See, for example, the widely cited Panton Principles from 2010 (*https://pantonprinciples.org/about/index.html*, accessed 22 January 2021):

> By open data in science we mean that it is freely available on the public internet permitting any user to download, copy, analyse, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. To this end data related to published science should be explicitly placed in the public domain.

[6] https://www.earthobservations.org/geo_community.php (accessed 22 January 2021).

[7] https://www.earthobservations.org/geoss.php (accessed 22 January 2021).

[8] Subsequently taken down.

[9] https://www.doi.gov/blog/geo-commits-unleash-power-open-data-address-global-challenges (accessed 22 January 2021).

[10] Of course, information was also there all along in our ideas about genes: genes were understood explicitly as a code, a form of information that needs deciphering to be comprehensible. This might also explain why it made so much sense to place 'the free flow of information and knowledge at the moral heart of genomics' (Reardon 2017: 29); in a sense, genetics and information – kinship and knowledge – have served as means to understand each other already for some time in Euro-America (Strathern 1995), to the point where they are sometimes hard to distinguish. Data and genes, then, share a longer history than might be imagined from Reardon's time-frame.

REFERENCES

Awati, K. & S.B. Shum 2015. Big data metaphors we live by. *Towards Data Science*, 14 May (available online: *https://towardsdatascience.com/big-data-metaphors-we-live-by-98d3fa44ebf8*, accessed 22 January 2021).

Bear, L., K. Ho, A.L. Tsing & S. Yanagisako 2015. Gens: a feminist manifesto for the study of capitalism. *Theorizing the Contemporary, Fieldsights*, 30 March (available online: *https://culanth.org/fieldsights/gens-a-feminist-manifesto-for-the-study-of-capitalism*, accessed 27 January 2021).

Bezuidenhout, L.M., S. Leonelli, A.H. Kelly & B. Rapper 2017. Beyond the digital divide: towards a situated approach to open data. *Science and Public Policy* **44**, 464-75.

Biagioli, M. 2008. Documents of documents: scientists' names and scientific claims. In *Documents: artifacts of modern knowledge* (ed.) A. Riles, 127-57. Ann Arbor: University of Michigan Press.

——— & S. Galison (eds) 2003. *Scientific authorship: credit and intellectual property in science*. London: Routledge.

Candea, M. 2018. *Comparison in anthropology: the impossible method*. Cambridge: University Press.

Carrier, J.G. & D. Miller (eds) 1998. *Virtualism: a new political economy*. New York: Berg.

Cooper, M. & C. Waldby 2014. *Clinical labor: tissue donors and research subjects in the global bioeconomy*. Durham, N.C.: Duke University Press.

Dalla Costa, M. & S. James 1972. *The power of women and the subversion of the community*. Bristol: Falling Well Press.

Duffy, M. 2007. Doing the dirty work: gender, race and reproductive labour in historical perspective. *Gender & Society* **21**, 313-36.

The Economist 2017. The world's most valuable resource is no longer oil, but data. *The Economist*, 6 May (available online: *https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data*, accessed 22 January 2021).

Eggers, W.D., R. Hamill & A. Ali 2013. Data as the new currency: government's role in facilitating the exchange. *Deloitte Review* 13 (available online: *https://deloitte.wsj.com/riskandcompliance/files/2013/11/DataCurrency_report.pdf*, accessed 27 January 2021).

Franklin, S. & M. Lock (eds) 2003. *Remaking life and death: towards an anthropology of the biosciences*. Santa Fe, N.M.: School of American Research Press.

Gabrys, J. 2016. *Program Earth: environmental sensing technology and the making of a computational planet*. Minneapolis: University of Minnesota Press.

———, H. Pritchard & B. Barratt 2016. Just good enough data: figuring data citizenship through air pollution sensing and data stories. *Big Data & Society* **3**: **2** (available online: *https://journals.sagepub.com/doi/full/10.1177/2053951716679677*, accessed 27 January 2021).

Gitelman, L. (ed.) 2013. *'Raw data' is an oxymoron*. Cambridge, Mass.: MIT Press.

Hammond, K.J. 2013. The value of big data isn't the data. *Harvard Business Review*, 1 May (available online: *https://hbr.org/2013/05/the-value-of-big-data-isnt-the*, accessed 22 January 2021).

Haraway, D. 1997. *Modest_Witness@Second_Millennium.FemaleMan©_Meets _Oncomouse*™. London: Routledge.

Hardt, M. 1999. Affective labour. *boundary 2* **26**: **2**, 89-100.

Helmreich, S. 2007. Blue-green capital, biotechnological circulation and an oceanic imaginary: a critique of biopolitical economy. *BioSocieties* **2**, 287-302.

Hey, T., S. Tansley & K. Tolle (eds) 2009. *The fourth paradigm: data-intensive scientific discovery*. Redmond, Wash.: Microsoft Research.

Hilgartner, S. & S.I. Brandt-Rauf 1994. Data access, ownership, and control: toward empirical studies of access practices. *Science Communication* **15**, 355-72.

Holbraad, M. 2005. Expending multiplicity: money in Cuban Ifá cults. *Journal of the Royal Anthropological Institute* (N.S.) **11**, 231-54.

Iliadis, A. & F. Russo 2016. Critical data studies: an introduction. *Big Data & Society* **3**: **2** (available online: *https://journals.sagepub.com/doi/pdf/10.1177/2053951716674238*, accessed 2 January 2021).

Jasanoff, S. 2001. Image and imagination: the formation of global environmental consciousness. In *Changing the atmosphere: expert knowledge and environmental governance* (eds) C.A. Miller & P.N. Edwards, 309-37. Cambridge, Mass.: MIT Press.

Konrad, M. 1998. Ova donations and symbols of substance: some variations on the theme of sex, gender and the partible body. *Journal of the Royal Anthropological Institute* (N.S.) **4**, 643-67.

Levy, K.E.C. & D.M. Johns 2016. When open data is a Trojan Horse: the weaponization of transparency in science and governance. *Big Data &Society* **3**: **1** (available online: *https://journals.sagepub.com/doi/full/10.1177/2053951715621568*, accessed 27 January 2021).

Locke, J. 1980 [1690]. *Second treatise of government* (ed. C.B. Macpherson). Indianapolis: Hackett.

Lupton, D. 2013. Swimming or drowning in the data ocean? *This Sociological Life*, 29 October (available online: *https://simplysociology.wordpress.com/2013/10/29/swimming-or-drowning-in-the-data-ocean-thoughts-on-the-metaphors-of-big-data/*, accessed 22 January 2021).

Martin, E. 1992. The end of the body? *American Ethnologist* **19**, 121-40.

Maurer, B. 2015. Principles of descent and alliance for big data. In *Data, now bigger and better!* (eds) T. Boellstorff & B. Maurer, 67-86. Chicago: Prickly Paradigm Press.

M'charek, M. 2005. *The Human Genome Diversity Project: an ethnography of scientific practice*. Cambridge: University Press.
Peck, R. 2017. Mark Cuban: 'Data is the new gold'. *Credit Suisse*, 22 June (available online: *https://www.credit-suisse.com/about-us-news/en/articles/news-and-expertise/mark-cuban-data-is-the-new-gold-201706.html*, accessed 22 January 2021).
Puschmann, C. & J. Burgess 2014. Metaphors of big data. *International Journal of Communication* **8** (available online: *https://ijoc.org/index.php/ijoc/article/view/2169*, accessed 22 January 2021).
Raicu, I. 2015. Metaphors of big data. *Vox*, 6 November (available online: *https://www.vox.com/2015/11/6/11620416/metaphors-of-big-data*, accessed 22 January 2021).
Rajan, K.S. 2006. *Biocapital: the constitution of postgenomic life*. Durham, N.C.: Duke University Press.
Reardon, J. 2017. *The postgenomic condition: ethics, justice, and knowledge after the genome*. Chicago: University Press.
Rose, N. 2006. *The politics of life itself: biomedicine, power and subjectivity in the twenty-first century*. Princeton: University Press.
Ruppert, E. 2015. Doing the transparent state: open government data as performance indicators. In *A world of indicators: the making of governmental knowledge through quantification* (eds) R. Rottenburg, S.E. Merry, S.-J. Park & J. Mugler, 127-50. Cambridge: University Press.
Safier, N. 2008. *Measuring the New World: enlightenment science and South America*. Chicago: University Press.
Seaver, N. 2015. Bastard algebra. In *Data, now bigger and better!* (eds) T. Boellstorff & B. Maurer, 27-47. Chicago: Prickly Paradigm Press.
Stoeckle, A. 2018. Rethinking reproductive labour through surrogates' invisible bodily care work. *Critical Sociology* **44**, 1103-16.
Strathern, M. 1995. *The relation: issues in complexity and scale*. Cambridge: Prickly Pear Press.
——— 1996. Potential property: intellectual rights and property in persons. *Social Anthropology* **4**, 17-32.
——— 2000 *Audit cultures: anthropological studies in accountability, ethics and the academy*. London: Routledge.
——— 2003. Emergent relations. In *Scientific authorship: credit and intellectual property in science* (eds) M. Biagioli & P. Galison, 165-94. London: Routledge.
——— 2005. Useful knowledge. *Proceedings of the British Academy* **139**, 73-109.
——— 2012. Gifts money cannot buy. *Social Anthropology* **20**, 397-410.
Walford, A. 2012. Data moves: taking Amazonian climate science seriously. *Cambridge Anthropology* **30**, 101-17.
——— 2018. If everything is information: archives and collecting on the frontiers of data-driven science. In *Ethnography in a data-saturated world* (eds) H. Knox & D. Nafus, 105-27. Manchester: University Press.
——— 2019. The properties of property: scientific data. *Theorizing the Contemporary, Fieldsights*, 29 March (available: *https://culanth.org/fieldsights/the-properties-of-property-scientific-data*, accessed 22 January 2021).
——— 2021. Analogy. In *Experimenting with ethnography: a companion to analysis* (eds) A. Ballestero & B.R. Winthereik, 209-18. Durham, N.C.: Duke University Press.
Watts, R. 2018. Hack, spy, swing an election: Orwell game sums up life in a tech dystopia. *The Guardian*, 2 July (available online: *https://www.theguardian.com/games/2018/jul/02/orwell-game-hack-spy-tech-dystopia-government-surveillance*, accessed 22 January 2021).

Donnée – ovules – gène – donnée

Résumé

Cet article fait observer que la donnée a de l'intérêt non seulement pour ce qu'elle est, mais également pour ce qu'elle est amenée à devenir : en ce sens, la donnée est une forme de potentiel. L'autrice en explore deux aspects, en établissant deux comparaisons avec d'autres formes de potentiel : les ovules et les gènes. En premier lieu, à partir d'une étude ethnographique réalisée dans l'Amazonie brésilienne auprès de chercheurs spécialistes et de techniciens en environnement, l'autrice compare le traitement des données avec le don d'ovules au Royaume-Uni, afin d'explorer en quoi le traitement des données peut être considéré comme une forme de travail reproductif. En second lieu, l'article s'intéresse aux infrastructures émergentes des mégadonnées dans les sciences de l'environnement, et compare ces dernières à la génomique, afin de soulever des questions critiques nécessaires au sujet de ces initiatives de données ouvertes. L'autrice conclut par une réflexion sur le recours à la comparaison comme moyen privilégié de déceler les formes perçues comme latentes que comportent les données.