

## **Raw Data: Making Relations Matter**

***Abstract:*** This article takes scientific ‘raw data’ as its ethnographic object in order to investigate the co-implication of nature and culture in scientific knowledge practices. The article traces out some of the activities that are involved in producing numerical climate data from the Brazilian Amazon. Although science and technology studies (STS) makes a strong case for associating relationality with certainty, the article argues that a particular form of data, ‘raw data’, complicates this association. It further argues that scientific data is not simply composed out of relations, but is a relation itself. The article ends with a brief reflection on the possible repercussions of shifting from thinking of science as producing multiple natures and cultures to thinking of it as producing the potential for relations.

**Keywords:** Data, climate science, Brazilian Amazon, relations.

## Introduction

Drawing on an ethnographic study of the every-day scientific practices pertaining to the production of observational environmental data in the Brazilian Amazon, this article demonstrates some of the different sorts of relationality that are constitutive of scientific data.<sup>1</sup> The article will concentrate on a form that observational scientific data takes known as “raw data” (in Portuguese, *dados brutos*), which is data that has come straight from the instrument, and has still not passed through any sort of quality control process. It is data with noise, gaps and errors in it, which contains “impossible” measurements, and does not display the relations between variables that it needs to in order to be considered “certified data” (*dados certificados*), ready for use by other researchers. By concentrating on uncertain, unstable and ambiguous raw data rather than certain and stable certified data, the article seeks to interrogate the relationship between stability and relationality from an empirical perspective. This also provides a vantage point from which to rethink the way that facts and relations might themselves relate to ideas of nature and culture.

I will start by reiterating an observation that has been made several times, (including in the introduction to this special issue): how the relationship of nature to culture seems to kaleidoscopically shift, forming a “matrix of contrasts” (Strathern 1980: 177) that does not hold steady across space or time. As also noted in the introduction, from a relational perspective the fact that there might be more than one way to relate nature to culture, across history and across (nature)cultures, means not only that there is more than one ‘culture’ in the world; it also means that there is more than one ‘nature’ – an insight that has galvanized a great deal of contemporary work in Anthropology and Science and Technology Studies (STS) (Haraway 1991; Latour 1999; Viveiros de Castro 1998, 2004).

But this observation has other, related and equally kaleidoscopic effects, one of which I will concentrate on in this article: namely, whilst (scientific) ‘facts’ can be opposed to (social) ‘relations’ (in the same way that nature can be opposed to culture – such that Western scientific knowledge-production about ‘nature’ can be seen to be reductively ‘cutting out’ culture); ‘facts’ can also be seen to be themselves constituted by ‘relations’ (in the same way that ‘culture’ has always been ‘in’ nature, in different ways) (see for example, Hayden 2012; Law 2002; Strathern 1992; Wagner 1977). The tension arises from the imbricated co-existence of a relation of opposition and a relation of implication. As Geof Bowker (2013: 168) puts it in another form, “[t]he social, then, is other than the natural and should/must be modeled on it; and yet the natural is always already social”.

An example of this tension plays itself out within STS and related disciplines. One of the most powerful of STS’s (often feminist-inspired) interventions into our understanding of the scientific production of nature is its emphasis on all the unseen or eclipsed relations that go into the production of scientific knowledge, and that permit the enactment and existence of certain sorts of ‘facts’, or ‘objects’. But if certain social critiques then highlight the way in which Western science removes all traces of these socio-cultural relations in its quest for objectivity (thus defining the latter two categories as against each other), STS approaches such as Actor-Network Theory further demonstrate how these social relations were in fact *there all along*. On the one hand, then, whereas scientific objectivity has been accused of “reductionism”, on the other ANT offers the opposite: the accumulation and endless multiplying of connections and relations. From this view, facts, objects and certainty – in short, reality as we know it – are a result of these relations, rather than opposed to them.

Based on an ethnographic study of raw data, this article will take this argument in a slightly different direction by exploring how observational scientific data is not only com-

posed of, or constituted by, particular relations – as if it were the sum result of them - but is also itself a relation of sorts. Given the theme of this special issue (and see Vertesi et al. 2014; Woolgar and Lynch 1990), we might say that data is already a form of relating nature to culture (although its relational capacity is not restricted to this particular dichotomy). In terms more familiar to my informants, observational scientific data is itself a relation between what is data (i.e what is fact) and what is error (i.e what is artifice). Indeed, data can be seen as a particularly interesting case for examining “multiple nature-cultures” in this sense because it slips between these categories. For example, certified observational data is not considered to be the same as a ‘natural’ object, sample or phenomenon; and yet neither is it considered to be an abstract scientific law. As such, this article suggests that we see the relational work that goes into the stabilization of data (and other scientific informational entities) not as the establishment of the existence of this or that fact, but rather as the establishment of specific kinds of relations. By focusing on the raw data, I demonstrate that not only are natures and cultures, or data and errors, enacted through the quotidian practices of scientific endeavour. In fact, it is the relation that constitutes the dualism of these forms that is being forged. That is to say, the relation is also an end, and not simply a means.

### **“Raw Data is an Oxymoron”**

In 2010 I spent a year conducting fieldwork with researchers and technicians involved in a Brazil-led scientific project, the Large Scale Biosphere Atmosphere Experiment in Amazonia (LBA). Initiated in 1998, the LBA is responsible for managing, collecting and making available data from a long-term data collection project that has spanned 10 years, with the intention of ascertaining the role of the Amazon forest in the global carbon cycle. It thus brings together hundreds of different researchers in different scientific disciplines. With help from col-

laborative partnerships with NASA,<sup>ii</sup> and various other Brazilian, European and North American institutions, the LBA built meteorological towers that stretch more than 20m above the top of the forest canopy, up to almost 70m in places. These towers have a profile of equipment on them, which measure different sorts of meteorological variables including wind speed and direction, rainfall, air temperature, and air humidity. Some towers also have specific equipment that measure carbon flux - the rate and volume of carbon moving vertically up and down, being exchanged between the biosphere and the atmosphere.

The data these instruments produce is either collected weekly by a technician, or sent in real time via telemetry to the LBA head office, which is in the city of Manaus, Amazonas. During my fieldwork, I spent a great deal of time with the LBA micrometeorology team that is responsible for this tower data. This involves work both 'in the field' (*no campo*) - that is, in the forest with the towers and the instruments; and back at the LBA head office, on the computer, processing or "cleaning" (*limpando*) the data that arrives from the forest. This clean data is then stored in several databases, and made available to the wider LBA community and beyond.

The data that the LBA micrometeorology team collects and processes evolves and changes as it is transformed from "raw data" (*dados brutos*), into "certified data" (*dados certificados*). The members of the micrometeorology team refer to "the data" (*os dados*) all the time without necessarily having to refer to any specific stage, because the context makes this clear: the question "did you collect the data?" implies one sort of data; "did you send the data to so-and-so?" implies another. Raw data was almost invariably explained to me as something still unfinished – "data that has not been cleaned yet" - or as one researcher put it, that still has to be "lapidated" (*lapidar*). It is data that does not display the relations between variables that it needs to in order to be considered "clean" and ready for use by other researchers.

A recent edited volume makes a case for the provocative argument that “raw data is an oxymoron” (Gitelman 2013; cf Bowker 2005). The thrust is that the imagery of rawness eclipses the fact that raw data is the result of an enormous amount of intervention, manipulation and hard work. Indeed, the sheer amount of hard graft necessary to produce raw data is very apparent in the case of the LBA. Funds have to be obtained from foreign institutions or from the Brazilian government; collaborative links need to be established between institutions and governments; visas need to be negotiated; data policies have to be established. After a series of expeditions to different possible sites in the forest, which can involve trekking in intensely hot and difficult conditions, a research site must be chosen – often as far from urban settlements and roads as possible. Once chosen, it must be ensured that a tower can indeed be built there, which means a team of engineers must be hired to 'probe' the area (*fazer sondagem*). If the tower is being built in an area near to indigenous territory, as was the case with one tower, the LBA need to send someone to ensure the indigenous people know and accept what they are doing.

Once the research site is settled upon, the tower must then travel up the Amazon to Manaus, and be transported to the research site. The difficulty of this journey depends a great deal on the time of year it is, as the river can sometimes be too swollen from rains for a safe passage to be possible, or too low for any boats to pass at all. International and domestic transportation laws must also be negotiated here, and the tower can be held for several weeks at customs. Once the tower has arrived at the research site, it must be built. This involves recruiting a team of people to work shifts in the forest over an extended period of time. Holes must be bored deep enough for the cement foundation of the tower to be laid, which in turn involves the back-breaking work of carrying all the building materials necessary for this to the research site. The tower is then assembled, level by level, often between massive rainstorms

that make it too dangerous to construct the upper levels for fear of lightning strikes. A lodging or camp must also be established near to the research site for those who are involved in constructing the tower, or subsequently, in maintaining the instruments or collecting data.<sup>iii</sup>

Clearly, a myriad of relations is forged between people, objects, instruments, and institutions in order to produce raw data from the forest – relations that are unlikely to make it into the published papers, reports or other artefacts of knowledge that come out of the data. Yet the initial establishment of the research site and tower is only the first step. What follows is an on-going effort involved in then continuously collecting the raw data. And just as building the tower is particularly arduous exactly because the desirability of the research site is based on its inaccessibility, collecting the data that the instruments on the towers produce is difficult for the same reason, and requires just as much effort. For example, raw data from a tower near to the city of São Gabriel de Cachoeira, in the north of the Amazon, arrived by aeroplane. The tower there was very isolated, several kilometres along a dirt road which frequently floods, and then about two kilometres or so along a particularly precarious trail that has been cut through the forest. A single technician collected data from this tower, and maintained the instruments in good working condition. He went to the tower whenever he could, downloaded the data onto an interface and uploaded it to a CD, before putting it on the next flight to Manaus. This was easy for him as he also worked at São Gabriel's tiny airport.<sup>iv</sup>

As such stories make clear, an enormous amount of effort is expended in order to establish a research site, build a tower and collect the raw data from the forest. This effort is necessary because the research sites are so isolated; as a result, the raw data therefore is highly prized by the scientific community of which the LBA is a part. On several occasions, I heard researchers remark that "this has never been done before", or that "this would be the first time

[this variable] was measured in the Amazon". On one occasion, a German researcher told me that they considered it probable that the data produced in the Amazon would be published even if it were of lower "quality" than data from elsewhere, simply because there was no other data like it. This is why researchers will go to such lengths to obtain it. The possibility of obtaining raw data on something that has never been measured before means that many different researchers from all over the world are drawn to the Amazon forest, and to the LBA.

Mark is one such researcher. Based at Harvard University in the United States, Mark came to Manaus to install a new piece of equipment on one of the LBA towers. His instrument could measure not only carbon dioxide ( $\text{CO}_2$ ) and water ( $\text{H}_2\text{O}$ ), but also carbon monoxide ( $\text{CO}$ ). Mark is a specialist in laser engineering, instrumentation, and trace gas measurements. When he arrived at the LBA HQ in Manaus, the head of the micrometeorology team had suggested that they install his instrument on one of the towers.

Mark was very excited because, as he explained, this would be the first time this sort of data had been obtained in "real time" – immediately available and even, Mark hoped, directly transmitted to his desk in Harvard. Furthermore, adding the measurements of  $\text{CO}$  to those already being taken by the instruments on the tower of methane ( $\text{CH}_4$ ) and  $\text{CO}_2$  would provide a unique picture of the chemical relations in the air. As Mark explained to me, the relation of these gases to each other is very important in global atmospheric chemistry directed at understanding climate change.  $\text{CO}$  produces ozone,  $\text{O}_3$ , and  $\text{CO}_2$ , so the global  $\text{CO}_2$  estimate in part depends upon the presence of  $\text{CO}$  in the atmosphere. This in turn depends upon knowing how  $\text{CO}$  behaves in tropical biomes - of which the Amazon forest is the largest in the world. However, just measuring isolated concentrations, Mark told me, "isn't chemistry". You

need simultaneous measurements of all of the gases to be able to see the way in which they are interacting. This sort of chemical measurement had never to his knowledge been done in the Amazon before - which is why he was so excited.

Mark was therefore willing to overcome all sorts of obstacles in order to collect this data. Alongside trying to ensure a clean energy supply reached his highly sensitive instrument, Mark also had to work out a way to protect the instrument from the heat, and from the bees, which like to make hives in any inviting cubbyhole they can find. Mark spent some time toying with ways to outsmart the bees, and decided eventually that he would just have to sit and watch over the instrument up on the tower, in the blazing heat, whilst it was running. When packed in a protective box, the instrument weighed 60-70 kilos, and it took two people to hoist it up the tower using a rope-and-pulley system, and another two people to guide it so that it did not bang into the tower on its way up. Once up there, Mark managed to get the instrument working to his satisfaction, but given the capricious weather conditions, he was worried. In fact, he was so worried about the instrument overheating that he left it running only when he was there - a total of about 5 hours, during which time he sat with the instrument in the tower by himself. Yet he seemed unfazed by the ordeal. "That'll be some good data", he told me, looking pleased.

It required an enormous amount of work to install Mark's instrument in the middle of the Amazon forest. Clearly, he would not have been able to collect any data had it not been for the actions of others, as Mark himself impressed upon me several times. Those difficulties are one reason why Mark's data is one of a kind. In one sense, the raw data is thus unique simply because it is from an isolated research site. Looking more closely at what is involved in its production, however, we might say that it is unique also because of the particular relational

demands that the isolated research site makes. The raw data does not just emerge from the isolated patch of forest. Without the ongoing relational work – that spans international data policies between countries to collectively working out how to think like a bee – this particular, Amazonian data would not exist. Its uniqueness is not due to its isolation, as such, but to the very particular relational configuration from which it emerges.

### **Raw Data's Ambiguity**

However, this highly sought-after Amazonian raw data is not only unique. What in fact makes it “raw” for the researchers and technicians I spent time with is that it is uncertain and not yet “trustworthy”.<sup>v</sup> Whereas Mark was conducting a short pilot experiment and therefore could sit in the tower with his instrument, the bulk of the data from the towers is collected over years, and stored automatically, which means that the instruments are left out in the forest for long stretches of time. The implication for the researchers and technicians is that the raw data from the towers could potentially be full of errors and gaps, and therefore highly uncertain. As I shall now show, this uncertainty is also the result of a certain sort of relationality.

Some uncertainty in the raw data is inevitable, arising from the fact that, as a metrological technician told me, "no sensor is perfect". That is, no sensor can measure without an element of observational uncertainty. One cause of this is the undesired effect of measuring in situ – for example, air temperature sensors heat up exactly by being ‘in’ the heat, so the heat they measure is partly their own, rather than that of the air. There is also uncertainty in the measurement system due to the loss of information in the conversion of analog signals into digital, a conversion that takes place automatically in the instruments in order to give the data.

No less important is the fact that all instruments function with a larger or smaller uncertainty factor, given by the manufacturers. An uncertainty factor of, for example, 10% means that if that instrument measures 25° C, this value actually could be anything in between 22.5° C and 27.5° C. This factor is a result of the metrological testing and calibration undertaken by the manufacturers, and is unavoidable when making measurements.

Further sources of uncertainty are inherent to more complicated measuring methods. Consider the technique for measuring carbon flux, called the eddy covariance method. The carbon flux system installed on the LBA towers comprises a sonic anemometer and what is known as an open-path Infra Red Gas Analyser (IRGA). The sonic anemometer measures wind direction in three axes using sound waves that are pulsed between three triangulated detectors, corresponding to three spatial dimensions: vertical, horizontal and lateral. These pulses are interrupted by the turbulent vortices and eddies created by the interaction of the wind with the tops of the trees in the forest, and from this interference the sonic anemometer infers wind turbulence direction and speed. The IRGA also measures the concentration of CO<sub>2</sub> and H<sub>2</sub>O in the air at very high frequencies. By correlating the concentration of CO<sub>2</sub> with eddy direction and speed, the eddy covariance method can give a measurement of carbon flux - that is, how much carbon is moving per area per unit squared. But these calculations get rather complicated, partly because of the rapidity with which carbon molecules move about, and partly because a lot of different factors have to be taken into account. In fact, some factors that are known to influence flux, such as mixing ratio and air density, are many times knowingly omitted from calculations – and this is despite the method being state-of-the-art. This omission imparts a certain level of uncertainty that accompanies the eddy covariance method irrespective of where it is applied.<sup>vi</sup>

Furthermore, specific uncertainties that in other settings are minimal are magnified considerably when measuring carbon flux in the forest. One of these sources is the footprint of the tower. The footprint of the tower – how far the IRGA and the anemometer can "see" – is calculated based on the assumptions that the terrain in question is flat, as it is in pasture-land, for example. As several researchers impressed upon me, the towers are therefore conventionally understood to measure *vertical* carbon flux only, as the horizontal axis can be taken as homogeneous. However, the Amazon forest is far from flat. This means that there is horizontal or lateral carbon and energy flux as well - the CO<sub>2</sub> "rolls" down hills and collects in basins and valleys. As a result, as one LBA researcher told me, the raw data from the towers also includes the effect of this horizontal carbon flux caused by the topography of the land – something the researchers have only recently become aware of. This poses a serious problem: the researchers had assumed the vertical measurement to be total when in fact it only captured part of the carbon that was moving around the forest.

On top of these issues are the problems of installing and maintaining electrical instruments in the forest, which interferes unremittingly. As the tower can move about an alarming amount in a rainstorm, the anemometer can often very slightly slip, which will affect the calculations of the directionality of the wind eddies. At the more remote towers sometimes the technician would arrive to find the instruments disconnected, or turned upside-down, as if a curious passer-by, or some animal, had decided to have a fiddle. Bees colonize the boxes that house the datalogger and slowly turn it into a hive, coating it in wax. When it rains, the open pathway IRGAs cannot take measurements because raindrops on their detectors do not evaporate due to the intense humidity. The manual for these instruments, intended for temperate climates, only warns users about the effects of snow. But the most substantial problem is also the simplest: dirt. It piles up on the instruments, coats the IRGAs, spreads over the radiome-

ters. It requires someone to constantly remove it, and the instruments should all be thoroughly cleaned as often as possible. Dirt offers the most vivid illustration that the very presence of the instruments in the world, in the forest, affects the data that is being produced.

Despite the best efforts of the micrometeorology team, raw data is therefore always a measurement *plus something else*. As the head of the team told me "dirt kills the curve" (*a sujeira mata a curva*). What he meant was that the dirt on the instrument will affect the shape of the graphed line of the data – that is, the dirt is included in the measurement. What is thus interesting about the raw data is not only that it emerges out of all the relational, intentional work that went into its production. More striking is the fact that what emerges from all this work is not a stable entity at all, but inherently uncertain and unstable, or “untrustworthy” as the researchers called it. As I have tried to demonstrate, this uncertainty is also constituted by a whole host of relations. Indeed, the raw data is constituted as much by these relations – between bees and wax and the datalogger, or between the instrument and the rainstorm – as by the network of social, legal and institutional relations, associations and negotiations that I described earlier. Even with cutting edge instruments, even with meticulously planned logistics, and access to a team of willing people who know the forest and can help you, it is very hard in the first instance to produce Amazonian carbon flux data that is singular in meaning. What is produced instead is very raw data that is inherently semiotically uncertain. Its rawness is a consequence of it being connected to all sorts of entities alongside the phenomenon it is a measurement of - including insects, unwanted meteorological events, and even unknown future variables.

### **Making Relations Matter**

What kind of relation between 'facts' and 'relations' can be drawn from the descriptions offered thus far? In this section, I concentrate on one particular and very influential way in which 'facts' and 'relations' have been understood in STS. Perhaps unsurprisingly, I have in mind the dynamic characterized by Actor-Network Theory (ANT), whereby the accumulation of associations in a network successively establishes the reality of a fact (Callon [1986] 1999; Latour 2005; Law 2003). The central point is that facts, or more radically truth, are a result of collective and contingent processes of connection, enrolment and relating between entities in a network. These processes might include the persuasion of colleagues, the enrolment of funders, the cajoling of objects, the objection of objects, and the creation of an audience, and more, all in order to achieve "*well-articulated actors*, associations of humans and non-humans"(Latour 2004a: 86; cf. Bowker and Star 2000; Gerson and Star 1986).

In Latour's now classical *Science in Action* (1987), the methodological emphasis is on "following" the multiple entities that need to coalesce around a task before scientific knowledge can be made. For this reason, the emphasis in such studies is often on the entanglement between science and entities conventionally deemed to be 'non-scientific' (for example, Derksen 2000). Latour's(1988; cf. Callon 1999: 81; Latour 1999a: 311) examination of Pasteur, for example, traces the articulation and "translation" of the interests of not only the most obvious human (Pasteur) and non-humans (the anthrax), but also French farmers, statisticians, public health workers, Pasteur's colleagues and the French public; as well as a host of sundry non-humans (Latour 1988; cf Latour 1999a: 113-173). More 'focused' laboratory ethnographies (for example, Knorr Cetina 1981, Knorr Cetina and Amman 1990) often centre on showing the process through which alternative interpretations of experimental results or "inscriptions" (Latour 1987: 79) are progressively discarded or refused by the scientists in favour of a single interpretation which comes to be seen as true. This has been variously de-

scribed as "black boxing" (Latour 1987: 131), "fixation of evidence" (Amman and Knorr Cetina 1990: 88), or "controlling interpretative freedom" (Collins 1985: 106). Whatever the designation, and wherever the ethnographic setting, the transformation of ambiguous results into singular meaning through socio-material means often considered to lie outside objective scientific practice, is frequently highlighted as the governing dynamic.

An interesting double quality characterises the reduction in meaning that such descriptions catalogue. Paradoxically, the fact-to-be (or "matter of concern") assumes a singular meaning (becomes a "matter of fact") exactly by being increasingly "articulated" into networks of different sorts of entities. The *more* connections and articulations it has, the more indisputable and therefore *singular* it becomes: as more and more humans and non-humans are convinced of and enrolled into its existence, the fact becomes more stable and real. As Latour writes, "reality grows precisely to the same extent as the work done to be sensitive to differences. The more instruments proliferate, the more the arrangement is artificial, the more capable we become of registering worlds" (2004: 85). Whereas a paucity of connections surrounds a proposition that is ambiguous and plural in meaning, as the number of connections and associations and actors increases - as the network grows - this ambiguity decreases.<sup>vii</sup>

Certainly, the production and collection of raw data I have described entails all sorts of associations, articulations, connections, and negotiations. Nevertheless, from the perspective of the LBA researchers and technicians, the problem of raw data is not only about forging stable relations, but also about dealing with the fact that the raw data is *already* connected in unpredictable ways to all sorts of other entities in the world. The 'problem' itself is that these connections are the *wrong* sorts of relations. The extraneous relations the raw data has with bees and lightning and dirt, or with non-linear effects of the environment, or with curious passers-by, obscure the representational potency presumed to be lurking within it. Whereas

data's singular meaning emerges as a result of a whole array of relations that are constantly being tested, the same can be said of its *lack* of a singular meaning.

Another way of saying this is that “the network”, as it were, does not exhaust the relational repertoire of the raw data.<sup>viii</sup> In fact, what makes the raw data "raw" is precisely the tension between the work of those actors in the network that toil towards eliciting a stabilized set of certified data, and those connections and relations that lie outside this network, anchoring the raw data in the liminal state between meaning and lack of meaning.

We might therefore observe that certainty and uncertainty are equally relational effects. And thus, rather than being a question of the accumulation of associations in a network, the stability of data is created by the gradual substitution of certain relations for others. The raw data might in this way be thought of as being constituted by a sort of relational friction. It is composed of all sorts of relations to entities that the researchers do not want, in combination (they hope) with the relations that they *do want* and have struggled to elicit. Processing or cleaning the raw data, which results in certified data that can be shared between researchers, is a matter of sorting out those relations – keeping some, forging others, and getting rid of yet others. At the end of this process, the result is stable, certified data that can travel to other institutions and researchers, and indeed, forge yet more relations.

That ambiguous, raw data is also composed of relations may seem a somewhat banal conclusion – as everything in ANT is understood to be the result of the network (although see Lee and Brown 1994). Yet, it helps to clarify how facts and relations come to be co-implicated in the production of data. If the dynamic that emerges is not of relational accumulation that produces facts, but rather of a progressive substitution of relations, then in the same way, it follows that not only are there relations always already ‘behind’ or within the data, there are also relations always already behind or within the relations (cf. Strathern 1995). A second im-

plication is that the relations that compose the raw data are relations that are somehow excessive to meaning, and therefore excessive to the network. This in turn means that raw data (or uncertainty) is not, in fact, a result of the network in the way a fact (or certainty) is. In ANT idiom, one might describe raw data as emergent from the overlapping interstices of two networks that are qualitatively different. But it is descriptively and analytically difficult to imagine the bees, the wax, the lightning, the passers-by, the rain, the non-linear affects of heat and so on to themselves be a 'network' in the ANT sense: that the LBA researchers are 'enrolling' these entities into producing an uncertain object. We might say therefore that raw data sits at the point at which the image of the network no longer has that much traction. That is to say, it would be imprecise to say that raw data is ambiguous because it is pulled between two different networks, and that as a result it is composed of too many relations. Rather, it is ambiguous because it pre-exists the one relation that is necessary for it to be considered either a fact - or not.

### **Not Data; Not Not Data**

Raw data is understood by the researchers and technicians I worked with to be awaiting transformation; it is unfinished. In paying attention to it, one finds one's attention constantly directed elsewhere - to what it will become, rather than what it is. Thus, the uncertainty around the raw data is not in terms of content – i.e. what it means numerically as such – but rather concerns the question of whether it is even data at all. If it is, then it will *always* have been data. If it is not, then it will *never* have been data. It becomes in a certain sense what it already is.<sup>ix</sup>

All data that is unable to lose its rawness sufficiently is defined by an unfixed relation of data to error, and this lack of fixity is something the LBA researchers and technicians deal

with on a daily basis. One PhD student I spoke to, for example, had found that their data on water vapour flux was "very strange". The balance did not "add up".

"In fact, I thought they [the data] were wrong. But then, when one of the reviewers of the committee read [my thesis], he told us that it was something he had noticed about the river himself..." According to this committee member, it was not necessarily an error. He had seen the same anomaly in another part of Brazil. "It's a question of the lateral transport of energy. It's like what [another LBA researcher] saw, only he saw it for CO<sub>2</sub>, and mine is humidity (...). So, it isn't an error, but to prove this, I need more data (...). We're not certain, because we need a longer series of data."

"Why not just accept it as an error?" I asked.

"The first time you see it, you think it's error because the difference is really big. So, it could be an error associated with, let's say, the calibration of an instrument. But it is actually the same sensor that measures these two. So...I thought there was a possibility (...) that there might be a calibration error...or it might be the lateral transport of heat. Of vapour in fact, of humidity, right? When the reviewer read my thesis, he was like, "Ah, I saw the same thing. And I think the problem you have is nothing to do with calibration, it's not an error, it's lateral transport (...). I know it wasn't an error...you are never certain, though, right? You imagine that it's an error. But now, after the [committee members'] result, I know that it could be...*could be*, I am also not certain that it is lateral transport. That's why I need more data."

As her supervisor told me, it's a question of "mistrust" (*desconfiança*), because they "don't know exactly what it is".

A similar uncertainty is captured in Susan Leigh Star and Elihu M. Gerson's (1987) work on anomalies in science. Star and Gerson are dissatisfied with what they see as a tautol-

ogy in Thomas Kuhn's description of anomalies, that suggests that an anomaly is only an anomaly because it does not fit with the current paradigm; that is, something is an error if it is wrong. They suggest that this can be refined by looking at the history of an anomaly, and how it can change status very rapidly not simply because it becomes logically wrong or right as a paradigm changes, but because of the constantly changing context (or what they call "work organization" *ibid*: 149) from which both the anomaly and the paradigm emerge. Regarding a certain anomaly in neurophysiological research, they write that "it's centrality to the problem structure of brain research became clear; it was defined and redefined to fit the need of an emerging work organization and the victorious school of thought in a debate. The anomaly went from mistake to artifact to discovery - even through accusations of fraud - until it was made tractable enough to be absorbed into an ongoing enterprise." (*ibid*: 160). Star and Gerson are interested in a very different set of questions from those that animate the present discussion, but this observation resonates. However, what their analysis further demonstrates is the extent to which an anomaly, like raw data, is in constant transformation, and thus always seen as something other than it is. I would like to draw attention explicitly to the conceptual potential of this characteristic. Raw data is neither phenomenon nor artifact, neither data nor error. Instead of this choice, raw data instantiates the moment before either of these positions is even ascertainable.

The technicians and researchers do not know the relation of their raw data to the world when they collect it - if it is error or data. Both these positions are contained within it, because what will end up being data will have *always been* data, and what will end up being error will have *always been* error. What ends up as error is in fact simply the relations that the researchers are not interested in. However, until analysed in this way, the raw data is thus still not exactly data, nor is it 'not data'. From this perspective, all the effort that the LBA puts into

collecting the raw data in the first instance generates neither data nor error - but the endless potential for constantly creating the relation between them.<sup>x</sup>

## **Conclusion**

I will conclude by briefly returning to the question of the co-implication of nature and culture with which I started. In the article, I have tried to hint at an unstable analogy between facts and relations, and nature and culture. I have suggested a shift in emphasis, from seeing the result of scientific work as the production of facts that are constituted by a myriad of relations, to seeing the result of scientific work as the production of the relation that makes the distinction between facts and relations possible. Finding out that raw data is what it has always already been - either one or other of the terms of the relation between data and error - eclipses the fact that it is the potential for this relation itself that is being forged over and over again. Extending the analogy to Anthropology and STS, it is likewise possible that if nature is understood to *always already* contain culture within it - to be a “nature-culture”- we risk eclipsing the potential that both contain to be related very differently in different times and places.

## **Acknowledgements**

This research was funded by the IT University of Copenhagen, and a travel grant from the Danish Research School for Information Systems. I would like to thank Casper Bruun Jensen and an anonymous reviewer for their comments. I would also like to thank the members of the LBA for being so patient with me.

**Antonia Walford** is currently a teaching fellow in Digital Anthropology at University College London, and a Post-Doctoral Research Associate at the Centre for Social Data Science (SODAS), University of Copenhagen. Previously, she was a Research Associate at the ESRC Centre for Research into Socio-Cultural Change (CRESC). Her research explores the effects of the exponential growth of digital data on social and cultural imaginaries and practices, focusing particularly on large-scale digitisation in the environmental sciences. She has published in several major journals and is preparing a book manuscript.

## References

- Bowker, Geoffrey. 2005. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.
- Bowker, Geoffrey. 2013. "Data Flakes: An Afterword to "Raw Data" is an Oxymoron." Pp. 167-172, in "Raw Data" is an Oxymoron, ed. Lisa Gitelman. Cambridge, MA: MIT Press.
- Bowker, Geoffrey C. and Susan Leigh Star. 2000. *Sorting Things Out: Classification and its Consequences*. Cambridge, MA: MIT Press.
- Callon, Michel. [1986] 1999. "Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fisherman of St Brieuc Bay." In *The Science Studies Reader*, ed. Mario Biagioli. London: Routledge.
- Candea, Matei. 2010. "'I Fell in Love with Carlos the Meerkat': Engagement and Detachment in Human-Animal Relations." *American Ethnologist* 37, no. 2: 241–258.
- Candea, Matei, Joanna Cook, Catherine Trundle and Thomas Yarrow. 2015. "Introduction: reconsidering detachment." Pp. 1-34, in *Detachment: Essays on the Limits of Relational Thinking*, eds. Matei Candea, Joanna Cook, Catherine Trundle and Thomas Yarrow. Manchester: Manchester University Press.
- Collins, Harry M. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press.
- Coopmans, Catelijne, Janet Vertesi, Michale Lynch and Steve Woolgar, eds. 2014. *Representation in Scientific Practice Revisited*. Cambridge, MA: MIT Press.
- Derksen, L. 2000. "Towards a Sociology of Measurement: The Meaning of Measurement Error in the Case of DNA Profiling." *Social Studies of Science* 30, no. 6: 803-45.
- Gerson, E.H and Susan Leigh Star. 1986. Analyzing Due Process in the Workplace. *ACM Transactions on Office Information Systems* 4, no. 3: 257-270.

- Gitelman, Lisa, ed. 2013. *"Raw Data" is an Oxymoron*. Cambridge, MA: MIT Press.
- Goldman, Marcio. 2009. "An Afro-Brazilian Theory of the Creative Process: An Essay in Anthropological Symmetrization." *Social Analysis* 53, no. 2: 108-129.
- Haraway, Donna J. 1991. *Simians, Cyborgs and Women: The Reinvention of Nature*. London: Free Association Books.
- Haraway, Donna. J. 1997. *Modest\_Witness@Second\_Millennium.FemaleMan©\_Meets\_Oncomouse™*. New York, London: Routledge.
- Hayden, Cori. 2012. "Rethinking Reductionism, or, The Transformative Work of Making the Same." *Anthropological Forum* 22, no. 3: 271-283.
- Jensen, Casper Bruun and Brit Ross Winthereik. 2015. "Test-sites: attachments and detachments in community-based ecotourism." Pp. 197-219, in *Detachment: Essays on the Limits of Relational Thinking*, eds. Matei Candea, Joanna Cook, Catherine Trundle and Thomas Yarrow. Manchester: Manchester University Press.
- Knorr Cetina, Karin and K. Amman. 1990. "Fixation of (Visual) Evidence." Pp 85-118 in *Representation in Scientific Practice*, eds. Michael Lynch and Steve Woolgar. Cambridge, MA: MIT Press.
- Knorr Cetina, Karin. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.
- Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.
- Latour, Bruno. 1993. *We Have Never Been Modern*. Trans. Catherine Porter. Cambridge, MA: Harvard University Press.
- Latour, Bruno. 1988. *The Pasteurization of France*. Trans. A. Sheridan and J. Law. Cambridge, MA: Harvard University Press.

- Latour, Bruno. 1999a. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
- Latour, Bruno. 1999b. "On Recalling ANT." Pp 15-25 in *Actor Network Theory and After*, eds. John Law and John Hassard. Oxford: Blackwell Publishing.
- Latour, Bruno. 2004. *Politics of Nature: How to Bring the Sciences into Democracy*. Trans. Catherine Porter. Cambridge, MA: Harvard University Press.
- Latour, Bruno. 2005. *Reassembling The Social*. Oxford: Oxford University Press.
- Law, John. 2002. "Objects and Spaces." *Theory, Culture and Society* 19, no. 5/6: 91-105.
- Law, John. 2003. "Networks, Relations, Cyborgs: On the Social Study of Technology." <http://www.lancaster.ac.uk/fass/resources/sociology-online-papers/papers/law-networks-relations-cyborgs.pdf> (accessed 17 February 2017).
- Lee, Nick and Steve Brown. 1994. "Otherness and the Actor Network: The Undiscovered Continent." *American Behavioral Scientist* 37, no. 6: 772-790.
- Pedersen, Morten Axel. 2012. "Proposing the Motion: Morten Axel Pedersen". *Critique of Anthropology* 32, no. 1: 59-65.
- Star, Susan Leigh. 1991. "Power, Technology and the Phenomenology of Conventions: On Being Allergic to Onions." Pp. 26-56 in *A Sociology of Monsters: Essays in Power, technology and Domination*, ed. John Law. London: Routledge.
- Strathern, Marilyn. 1980. "No Nature, No Culture: The Hagen Case." Pp. 174-222 in *Nature, Culture and Gender*, eds. Carol P. MacCormack and Marilyn Strathern. Cambridge: Cambridge University Press.
- Strathern, Marilyn. 1992. *After Nature: English Kinship in the Late Twentieth Century*. Cambridge: Cambridge University Press.
- Strathern, Marilyn. 1995. *The Relation: Issues in Complexity and Scale*. Cambridge: Prickly

Pear Press.

Viveiros de Castro, Eduardo. 1998. "Cosmological Deixis and Amerindian Perspectivism."

*Journal of the Royal Anthropological Institute* 4, no. 3: 469-488.

Viveiros de Castro, Eduardo. 2004. "Perspectival Anthropology and the Method of Controlled

Equivocation." *Tipiti: Journal of the Society for the Anthropology of Lowland South*

*America* 2, no. 1: 3-22.

Wagner, Roy, G. 1977. "Scientific and Indigenous Papuan Conceptualisations of the Innate: A

Semiotic Critique of the Ecological Perspective." In *Subsistence and Survival: Rural*

*Ecology in the Pacific*, eds. T. Bayless-Smith and R. Feachem. London: Academic

Press.

Woolgar, Steve and Michael Lynch, eds. 1990. *Representation in Scientific Practice*. Cam-

bridge, MA: MIT Press.

<sup>i</sup> I will be using a relatively wide notion of relationality, that will take advantage of the slip-page between 'relations' and 'culture' that I go on to explore below, and the corollary sense in which relations are therefore opposed to facts, and often associated with connectivity.

<sup>ii</sup> The National Aeronautics and Space Administration of the United States.

<sup>iii</sup> At one isolated camp, there was also a satellite dish so that the construction workers who spent several months in the forest at a time could watch the telenovela. But mysteriously, the dish only received the right channel if a potato was carefully balanced on the connection between it and the television.

<sup>iv</sup> Unfortunately, the tower at São Gabriel fell down in June 2010, halfway through my field-work.

<sup>v</sup> In a subsequent telephone conversation, Mark told me had not used that data in a publication, in fact. Some of it was good data, but there had been too many errors and there had "not been enough to draw trends".

<sup>vi</sup> When I asked about this, I was told by one researcher that the scientific community knew about these issues, of course, but at the moment the eddy covariance method was considered to be the best way available to measure flux, so they kept using it. That is, just because the method is uncertain does not mean it is redundant.

<sup>vii</sup> As an aside here, it is worth noting that the entities that Latour (and others) discuss are often entities that are discovered/fabricated in a laboratory. In such a setting, it makes intuitive sense to picture the lone researcher at his desk, with suspicions that he might have 'discovered' something of note and the subsequent attempts to convince his colleagues that indeed he has (what Donna Haraway refers to as the heroic depiction of science, 1997: 31). This plays into an idea of the increasing connectivity and stability (and therefore the "constructed" nature) of his discovery in the world. This narrative is arguably more convincing when the setting is a laboratory.

<sup>viii</sup> See also feminist critique of ANT here, such as Star (1991).

<sup>ix</sup> I am here appropriating an idea from Marcio Goldman's analysis of Afro-Brazilian *candomblé*, who in turn cites Gilles Deleuze on Nietzsche (Goldman 2009: 121-123).

<sup>x</sup> The dynamic I am pointing to resonates with Morten Axel Pedersen's argument regarding 'post-relational' anthropology (2012; see also Candea 2010), particularly the challenge he puts to reductionist understandings of the apparently non-relational. I am not so sure however that the relational 're-invention' of scientific practice that I am here engaged in contributes to exhausting the conceptual capacity of the relation, as Pedersen would have it. See also Casper Bruun Jensen and Brit Ross Winthereik's argument regarding the necessary interplay between different sorts of relationality (2015; cf Candea et al 2015).