DOI: 10.1111/dewb.12183

ORIGINAL ARTICLE

WILEY developing world bioethics

To share or not to share: Incentivizing data sharing in life science communities

Louise Bezuidenhout

Correspondence
Louise Bezuidenhout, Institute for Science, Innovation and Society, University of Oxford, Oxford, United Kingdom.
Email: Louise.bezuidenhout@insis.ox.ac.uk

ABSTRACT

Most scientists recognize the importance of sharing data online in an open fashion. Nonetheless, many studies have documented the concerns that accompany data sharing activities, including loss of credit or IP, misuse and the time needed to curate interoperable data. To this end, discussions around data sharing often identify incentives that could potentially ameliorate these disincentivising concerns.

Nonetheless, current Open Data discussions often rely on evidence-based studies to identify the disincentives to overcome. This results in highly specific and directed interventions. In contrast, this paper offers a different interpretation of these concerns. To do so, it makes use of the Thomas Theorem which suggests that: *"If men define situations as real, they are real in their consequences"*.

Using empirical evidence from sub-Saharan African (bio)chemistry laboratories, this paper illustrates how individual perceptions of research environments – whether associated with evidence or not – are highly influential in shaping data sharing practices. It concludes with the suggestion that discussion on incentivising data sharing amongst scientific communities need to take a broader set of concerns into account and offer a more creative approach to ameliorating environmental disincentives.

KEYWORDS

data sharing, incentives, low/middle-income countries, Open Data, Thomas Theorem

> *"If men define situations as real, they are real in their consequences"*[1]

In 1928 William Isaac Thomas and Dorothy Swaine Thomas formulated the above *Thomas Theorem*. The key concept of this theorem was the observation that an individual's interpretation of a situation causes their action. The interpretation itself is not objective, but rather informed by a range of personal traditions and concerns. As a result, actions become influenced by these subjective perceptions of situations. This dominance of the subjective will persist even when there is an objectively correct interpretation available to the individual.

In the paper below I discuss how the Thomas Theorem can be used to unpack the complex milieu of (dis)incentivising the sharing of scientific data. In particular, I focus on the sharing of *non-human* data by individual scientists as part of their daily research practice. Such sharing activities could include disseminating data as supplementary material in publication; on personal, institutional or project webpages; in university, disciplinary or general repositories; on professional networking sites and to peers.[2] Using empirical work conducted in four African laboratories, I focus on how current discussions overlook the scientists' individual perceptions of their situations, and thus fail to identify areas of intervention that could contribute to marked changes in data sharing practices.

[1]Thomas WI, Thomas DS. *The child in America: behavior problems and programs*. New York: Knopf; 1928. p. 571.

[2]Ferguson L. How and why researchers share data (and why they don't) [Exchanges] C2014 [cited 8 Aug 2017]. Available at: http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges.

 © 2018 John Wiley & Sons Ltd

1 | DATA SHARING AND OPEN DATA

The importance of sharing scientific data online is a topic that needs little introduction. A wide range of publications,[3] policy documents,[4] and codes of conduct[5] all detail how opening up the availability and re-usability of scientific data will lead to exponentially increasing benefits for both science and society. The Open Data movement has recently emerged as an umbrella movement to support advocacy, awareness raising, standard setting, and many other key areas that facilitate data dissemination, re-use and interoperability. Moreover, the Open Data movement advocates limit ownership and control of data to the requirement to attribute and share-alike, and to remove any financial barriers to data. Key concepts that play important roles in understanding this movement are availability and access, re-use and redistribution, and universal participation.

Despite widespread support for the Open Data movement, however, the challenges of achieving global openness and data interoperability cannot be underestimated. To this end, it is commonly recognized that there can be no "one size fits all" when it comes to data sharing activities.[6] In most cases, disciplinary or user communities develop their own data standards and practices that suit their particular needs.[7] Such diversity is even more evident when examining the debates distinguishing the sharing of human/non-human data.[8] The sharing of human data often has additional considerations associated with it, in particular bioethical concerns such as privacy, anonymity and non-maleficence.[9] Moreover, the complicated relationship between sovereign ownership of personal data raises continual challenges for consent mechanisms and makes participant engagement a prominent element of human data sharing discussions.[10]

These challenges have led to data sharing policies and practices being developed in a piecemeal fashion. Institutions, regions and countries around the world have implemented a range of regulations and guidelines to govern the flow of data within and across institutional and political borders. Establishing international data sharing practices have been challenged by territorial jurisdictions and often struggle with efficiency, legitimacy and sustainability. As a result, the Open Data movement continues to rely on community buy-in, commitment and vision, and the establishment of "bottom-up" communities of sharing. Indeed, the momentum of the Open Data movement has long been – and continues to be – dependent on the often-voluntary effort of scientists.

Nonetheless, despite widespread support for the Open Data movement, penetrance of data sharing amongst researchers remains incomplete. There has been quite a lot of recent literature that problematizes this lack of buy-in. Studies on why scientists do or do not share data that highlight problems with data sharing that transcend disciplinary (or practice) boundaries. Many of these related to the current challenges of regulation, including lack of enforceable or coherent guidance from publishers, funders and research institutions.[11] Similarly, discussions about how best to mediate intellectual property requirements, credit for research and data ownership both nationally and globally continue to demonstrate the complexity of these issues.[12] The literature also regularly highlights the need for the development of sustainable and coherent infrastructures to support data sharing and dissemination.[13] The construction and maintenance of these structures is by no means straightforward and involves many different stakeholders, including user-communities, information and communication technologies (ICT) experts, and funding bodies. Finally, and unsurprisingly, there are practice-oriented challenges relating to the (often unsupported) time and money needed to share data.[14]

Studies of perceptions of Open Data also continue to demonstrate how support and hesitation for data sharing can exist concurrently within the scientific community (and individual). A large 2015 survey conducted by the Wiley publishing group involved 2250 scientists from 7 different countries.[15] This survey identified a wide range of reasons why scientists felt they would or would not share data. Motivations supporting sharing data included community norms, public benefit, and facilitating transparency and re-use. Importantly, 55% of respondents also recognized that sharing data increased the impact and visibility of their research. This belief correlates with a wide range of other research that highlights the benefits of openness. This includes higher citation rates for Open Access publications,[16] improved visibility online and the possibility of further collaborations.

[3]Molloy JC. The Open Knowledge Foundation: Open Data Means Better Science. PLoS Biol. 2011; 9(12): e1001195.

Murray-Rust P. Open data in science. Serials Review. 2008; 34(1): 52–64.

[4]International Council for Science, InterAcademy Partnership, International Social Science Council, & World Academy of Science. Open Data in a Big Data World. Paris: ICSU; 2015.

Royal Society. Science as an open enterprise. London: Royal Society; 2012.

[5]Wellcome Trust. Policy on data software and materials [Wellcome Trust] C2017 [cited 8 Aug 2017]. Available at: https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing

[6]ICSU, *op. cit*. note 4.

[7]Borgman C. The conundrum of sharing research data. Journal of the American Society for Information Science and Technology. 2012; 63(6): 1059–1078.

Chen CLP, Zhang CY. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. Information Sciences. 2014; 275: 314–347.

[8]Leonelli, S. When humans are the exception: cross-species databases at the interface of biological and clinical research. Social Studies of Science. 2012; 42: 214–236.

[9]Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics — re-shaping scientific practice Nature Reviews Genetics. 2012; 10: 331-335.

[10]Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. Expert Review of Molecular Diagnostics. 2014; 14: 1053–1065.

[11]Savage CJ, Vickers AJ. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. PLoS ONE. 2009: 4(9); e7078.

[12]Carroll MW. Sharing Research Data and Intellectual Property Law: A Primer. PLoS Biol/ 2015: 3; e1002235.

[13]Bastow R, Leonelli S. Sustainable digital infrastructure. EMBO Reports.2010: 11; 730–735.

Leonelli S, Smirnoff N, Moore J, Cook C, Bastow, R. Making Open Data Work in Plant Science. Journal for Experimental Botany. 2012: 64; 4109–4117.

[14]Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions. PLoS One. 2011; June 29.

[15]Ferguson, *op. cit*. note 2.

[16]Elsevier. Citation metrics and open access: what do we know? C2017 [cited 25 Aug 2017]. Available at: https://www.elsevier.com/connect/citation-metrics-and-open-access-what-do-we-know.

In contrast, the respondents to the Wiley survey also reported a wide range of reasons why they would potentially be hesitant to share their data. These included lack of resources (time, money, expertise) and issues relating to individual credit (scooping, misuse, misapplication). A number of scientists also highlighted absent, confusing or conflicting policy requirements as reasons why they didn't bother to share. In addition, the Wiley survey results clearly showed that the prevalence of regular data sharing varied by country, as did the top incentives and disincentives. Such results clearly show how support for Open Data does not necessarily translate into data sharing practices.

Surveys such as the Wiley one provide key insights into discussions on incentivizing data sharing amongst the scientific community. Perhaps most importantly, it highlights the disjunction between the "ideal and the real", whereby the pragmatic concerns of daily practice override normative motivations towards "right action". Thus, the very personal concerns that scientists have with respect to losing credit, ownership or control of their data often trump their commitment to the recognized benefits of sharing and shape their data sharing actions. Recognizing that these concerns are rooted in their research environments – or their *perceptions* of their research environments – makes discussions of incentivizing data sharing highly complex and contextual.

2 | UNPACKING PROBLEMS WITH DAILY DATA SHARING: USING EMPIRICAL EVIDENCE FROM AFRICA

If one unpacks discourse around data sharing, it becomes evident that two key issues are often overlooked. First, that support for data sharing does not necessarily translate into personal research practices. Second, that while the incentives of sharing data were focused on a membership to a global community of scientists, the disincentives stopping individual scientists from sharing data tended to be linked to their daily working experiences. This strongly suggests that further incentivizing data sharing without removing the contextual disincentives that curtail behaviour will continue to be suboptimal.

Furthermore, if one applies the Thomas Theorem to this dichotomy the conversation becomes even more complicated as it becomes evident that the disincentives are linked to *perceptions of the research environment* – and not necessarily to the actual research environment. Thus, how respondents view their research environments and interpret their personal situations could strongly influence their data sharing actions. Indeed, how individual scientists identify and evaluate the strengths and weaknesses of their daily research context will influence how they perceive the dangers of sharing their data. It is important to note, however, that whether these perceptions of strengths and weaknesses were individually held or communally recognized, it potentially makes *little difference to the daily data sharing activities* that the scientists engage in.

As part of a larger project examining data sharing practices in sub-Saharan Africa, I spent 3 – 6 weeks in four different (bio)chemistry laboratories in Kenya and South Africa.[17] These four laboratories were selected as examples of flourishing "homegrown" research in Africa, meaning that they were unaffiliated to large research networks or consortia. All sites engaged in a combination of research and teaching activities and regularly graduated postgraduate students via dissertation. While all sites received research funding, the source (national or international) and the quantities varied considerably. Importantly, however, the majority of daily running costs in the laboratory were covered by these funded grants, and all host institutions provided little in the way of core funding to cover daily expenses.

During the course of these visits I conducted 56 semi-structured interviews with postgraduate and research scientists. During these interviews participants were invited to talk about their perceptions and understanding of data sharing and the Open Data movement. They were also asked to elaborate on what data sharing activities they engaged in, and why they chose those over other avenues of data dissemination. The interviews were accompanied by extensive laboratory observations in which the author examined the physical, social and regulatory aspects of the researchers' environments.

The interviews and observational data were thematically analysed by the author and collaborators. The final thematic analysis is available online.[18] The results of this broader project are extensively discussed in other publications.[19] In the rest of this paper we will discuss how the interview data also highlights the importance of the Thomas Theorem can be used to understand the disjunction between data sharing endorsement and daily practice. In the section below I will identify some key issues that contribute to these discussions.

3 | IT'S A NICE IDEA … IN THEORY

Nearly all the scientists I interviewed recognized openness as valuable part of modern science. As succinctly put by a South African professor: *I think it leads to better science* (SA1/3). No participants found fault with the idea that science progressed through the release of data and the dissemination of results. Unsurprisingly, few participants had any concerns about Open Access publishing.[20] In contrast, however, the discussions surrounding data sharing and Open Data were markedly less coherent. As described below, many of the participants concurrently held – and espoused – conflicting ideas on data sharing and good practice.

In addition to the community-based benefits, many participants recognized the potential personal benefits of being open with one's data.

[17]This study was funded by the Leverhulme Trust and approved by the University of Exeter research ethics committee as well as the appropriate bodies at each institution.

[18]Bezuidenhout L, Rappert B, Leonelli S, Kelly AH. Beyond the Digital Divide: Sharing Research Data across Developing and Developed Countries [Figshare] C2016 [cited 25 Aug 2017]. Available at: https://doi.org/10.6084/m9.figshare.3203809.v1.

[19]Bezuidenhout L, Kelly AH, Leonelli S, Rappert B. "$100 Is Not Much To You": Open Science and neglected accessibilities for scientific research in Africa. Critical Public Health. 2017; 27(1): 39–49.

Bezuidenhout L, Leonelli S, Kelly AH, Rappert B. Beyond the Digital Divide: Towards a Situated Approach to Open Data. Science and Public Policy. 2017; 44(4): 464–475,

[20]Aside from comments about author processing charges, which will not be addressed in this paper.

A South African postgraduate student clearly recognized the benefits that openness provided in terms of getting a wider audience for their work. (S)he said: *I'd like to consider myself as a researcher so in a way that would mean that I would want my work to get as much attention as can be, so I'd favour [openness] so that as many people as able are able to read it – they don't have to cite it – to get to know what's going on – in malaria, in TB, HIV. So I'm more likely to want the [open] mode than the closed access mode* (SA1/2). Similarly, a Kenyan postgraduate mentioned that they supported openness in research: *because that is the way you'll get some of the key challenges and know how to handle* (KY1/5). Such observations correlated closely with the findings of the Wiley survey. The participants were identifying a combination of community benefit and pragmatic gains as reasons to engage in data sharing activities. Interestingly, however, more participants highlighted the community gains of "improving science" than those who identified personal benefits.

In contrast to the unanimous support for the *idea* of data sharing, very few participants had any practical examples to support their endorsement. As one South African participant commented: *people are just locking it [data] away in their computer* (SA2/7). When asked what Open Data practices they personally engaged in, participant responses gathered around statements such as: *I won't release data unless I first of all publish* (KY1/11). Other participants went further to say: *[o]h, I think it is very dangerous* (SA1/6).

Ultimately, these conflicting concerns led to an impasse, well-described by a statement from a South African scientist, who said: *I feel an obligation to share, but I don't think I'd share data as I don't like the idea of people trawling the raw results to write papers and take credit* (SA1/11). Thus, it becomes important to ask, why were the Kenyan and South African scientists interviewed so committed to maintaining rather conflicting perspectives when it came to sharing data?

4 | HAVE YOU SEEN THE CHALLENGES WE FACE?

In each interview the participants were asked to detail the data sharing activities they regularly engaged in. From these descriptions it became evident that very few of the participants engaged in any data sharing activities, and that their data dissemination strategies were solely linked to publishing in academic journals. While all sites had institutional repositories, only one automatically uploaded publications and theses. At the other sites the use of institutional repositories was via voluntary contribution. The awareness of the existence of these institutional repositories was very low amongst participants, and there were no reports of any voluntary contributions. Similarly, three of the four sites had very basic (or no) departmental or personal websites, and the research in these laboratories was poorly visible online. Interestingly, very few of the participants mentioned sharing data via supplementary material accompanying journal publications, which suggested that the journals they published in did not request – or make provision for – data sets to accompany publication. From these observations it became evident that there was little in the way of data sharing occurring at the different fieldsites.

In a large number of my interviews, participants explicitly pointed out how their research environments differed from those in high-income countries. Participants often went into considerable detail to enumeratethe daily difficulties they encountered due to the physical and social structures of their research environments. Key to these discussions was the underlying perception that because of the difficulties of their environments they *could not (be expected to) share data.* Whether these difficulties were actually stopping them from sharing data was immaterial, and the participants identifying them as insurmountable challenges was sufficient to stop them engaging in data sharing activities. When correlating the challenges they mentioned to my laboratory observations, it became evident that while some were easily observable/experience-able, some needed to be accepted as true *because they were reported by the participants*. This section details a small sample of these issues.

4.1 | Observable issues

The participants reported a number of issues that correlated closely to my experience of their physical research environment. These issues inhibited their (and my) ability to share or re-use online data. Most obvious was the slow/unreliable wifi – particularly at 3 of the 4 sites – and frequent downtime. A participant from South Africa highlighted this, saying: *[l]ike I said you can get around it with a lot of patience - waiting when the internet is not strong enough to allow you to download things. But they're always promising us that things will improve, but they are promising one year after another but maybe it will improve. (SA2/12).* This, the participant mentioned, complicated access to all online resources – data sets, figures, supplementary materials and journal articles. It also meant that contributing data and articles was very difficult.

Internet provision was also a problem for many participants when they left their place of work. One individual in Kenya detailed this problem, saying: *here I'm using wifi, so the moment you step out of the college you're shut off and again in the estates [less-formal residential areas] where we stay as of now the internet is a bit expensive. It's not affordable. So I do as much as I can here so that when I go back home I'm going to rest*. (KY1/3). Related issues that were commonly reported were to do with power outages and the time it took to get the servers back up and running efficiently.

Concerns were also raised about the information and communication technologies (ICTs) that the participants had available. In 3 of the 4 sites, participants had to buy their own personal computers for use in their office. Similarly, only one of the sites offered any free software downloads to staff or students. This meant that they were often making use of older ICTs which slowed down their online activities and made some websites and databases unavailable.

4.2 | Positional concerns

Nonetheless, despite the challenges of ICTs, laboratory equipment and research funding, all four sites had bustling laboratories and a range of masters and PhD students enrolled for research-based theses. Interestingly, however, the number of postgraduate students,

traditions of supervision and the expectations of degree requirements were also often highlighted as a barrier to data sharing. One South African participant elaborated on this, saying: *[b]ut also with the students you know we have a very high turnover of students and then those students leave and they haven't stored their data properly or they leave with their hard drive* (SA2/7). Despite the students generating data and successfully completing theses, the structures of the institutional environment were seen as undermining the possibility of collating, curating and disseminating the postgraduate research data. Supervisors felt hard-pushed to keep track of all the data that their students produced, and a lack of standardized data annotation, curation and storage seemed to exacerbate these problems.

Many of the participants also suggested that they perceived data sharing to be a luxury that "other people did". Citing lack of funds – particularly in relation to publishing in Open Access journals – they clearly identified their environments as not supporting any desired activities. For example, a South African participant mentioned that: *probably people who are established who will do [publish in OA journals] for there because now myself if I'm here I must get the chemicals, I must get funding* (SA2/7). Similarly, a Kenyan participant stated that: *if we want to publish we must pay out of our own pocket. It is important to publish, but it is also expensive* (KY2/1). Perceptions that openness was something that was financially "out of their reach" meant that many participants actively dissociated themselves from the idea of engaging with Open Access platforms. This also led to cross-over dissociation from Open Data, with the perception that it was something that *Western people do* (KY1/4) because they had the necessary funds.

These perceptions stood in direct contradiction to the availability of author fee-waivers for scientists from low/middle-income countries (LMICs)[21] and similar schemes. Nonetheless, the perception of this financial barrier served as sufficient justification for not investigating these possibilities or requesting support for publication costs, data management or depositing data. Data sharing, it would seem, is something that happened to "other people", and for whom systems and assistance were designed. This is evident in another quote from South Africa: *[a]gain I'm going to say my view will be it will be established researchers who would do that [share data] now its people who, let me say, they don't have to prove themselves to anyone so they know that okay it's there but they can always have something on the side* (SA2/7). Moreover, the lack of tangible incentives for sharing through promotion criteria or teaching buy-out further exacerbated such perceptions.

Interestingly, however, a number of higher level academics recognized that it was not "all about the money". A professor in South Africa tellingly commented on this, saying: *here [in comparison to well-established universities] it's a different culture. I think we have to learn to crawl before we walk or compete. So, we must do some things right and perhaps we don't need money for that - even though money is useful. I mean from my experience I always found good researchers can attract funding - that is the rule. It's different in Europe because you could be good and still battle for funding as the competition is stiff but here if you have a story to put together I am sure you get funding. So here funding is not the biggest problem we have* (SA2/1).

Similarly, another South African researcher with experience collaborating with other African countries further elaborated on this conundrum, saying: *[t]he other confounding thing in this picture is that it's not at all the case that all these institutions exist in a culture of deficit – or that they don't have money, to be blunt. So [a university in a neighbouring country] had a brand spanking new version of [a digital scanner] – they'd just invested millions in a new research management system that wasn't being used because nobody knew how to use it. In the time that we were there – ok, it's not true to say that there isn't deficit – what there seems to be is intense disparity. Incredible spending of resources in one area and then complete poverty right next to it* (SA1/9).

The views of different staff members contrasted very strongly when it came to assessing issues such as time and teaching loads, funding and responsibility for sharing data. Understandably, early career researchers felt under more pressure to publish and gain promotion than those further up the career ladder, which altered the manner in which they chose to (not) share data.

4.3 | Speculative concerns

Another key concern that participants expressed was that they would be scooped after they released their data. While this is, of course, not a concern limited to LMICs,[22] the manner in which these concerns were presented represented a particular view of not only the LMIC research environments, but also their global counterparts. Many participants drew attention to the fact that their relative lack of resources (in comparison to what they understood as high income country (HIC) labs) meant that their research took longer to complete. This, they felt, put them at risk. As one South African scientists commented: *if you haven't finished your project and you contribute there's other people with a lot more resources in terms of physical actual lab resources that can do what we do in a year in a couple of months. So if you were to share what you were doing without having finished they will finish the work for you and basically your work is rendered obsolete* (SA1/7). Similarly, a Kenyan scientist elaborated, saying: *because the results, you know, they can be taken away. You're dealing with colleagues and there are some – that which has taken you like 2 months, they can come and do it overnight with a whole research lab* (KY1/1).

What is evident from these quotes is how many participants viewed their HIC counterparts – or the "faceless global" online community as somehow predatory. In a way, this suggested a deep-seated concern that someone was looking over their shoulder, waiting to pounce on their data. This was coupled with a perception of the relative lack of protection offered to them as LMIC scientists in LMIC institution. A South African scientist verbalized these fears, saying: *for example with the size of [the university,] we don't have the same legal power like a university in Australia or America. If someone steals their ideas they will go for them. But we are small and who is going to believe me when I say "this was my idea". So there is that fear* (SA2/11).

Interestingly, these concerns existed side-by-side fears that even if the participants were to share data that it would not be valued and re-used. The awareness that their data were created using older equipment and methodologies, and less expensive reagents was raised by

[21]Lawson S. Fee waivers for Open Access journals. Publications. 2015: 3; 155–167.

[22]Ferguson, *op. cit*. note 2.

a number of participants as reasons why, even if they were to share, "no one would care". A participant in Kenya exemplified this concern, saying: *[h]ow much do we do to develop our own data? What processes do we need to convince people that the data are good* (KY2/13)? Together, these two sets of concerns presented a confusing perception of the participants' research situations, where they were concurrently at risk of being scooped and of being ignored.

4.4 | Say what you want …

What became apparent was that many participants were using the perceptions of their research environment to "other" themselves from their HIC colleagues. As one participant said in conversation: *[Data sharing is] a Western issue* (KY1/4). This perception was held by a number of participants, who cited the prevalence of data sharing discussions in HICs, the foreign data sharing policies, the lack of African presence in the Open Data movement and a variety of other issues as evidence that data sharing just "wasn't an African thing".

The implications of such statements are twofold – first, that many of the participants were linking data sharing activities to geographic, economic or cultural localities outside of their frame of reference. Second, because the participant perceived their physical environment to differ from these localities that data sharing "wasn't their problem".

5 | USING THE THOMAS THEOREM TO UNDERSTAND INCENTIVES AND DISINCENTIVES

When discussing data sharing with one South African professor, (s)he made an interest observation. (S)he said: *where I find it difficult is people don't understand our situation – it's not bad will, it's just not being able to figure it out* (SA2/12). What she was alluding to was the difficulty of uploading data to certain online sites due to inappropriate design that did not support low-bandwidth settings. Nonetheless, her comment also offers real insight into some of the problems with incentivizing data sharing in LMICs.

From the data above it is evident that the participants used a range of different issues to personally define and make sense of their research environments. These included the highly visible (such as the poor bandwidth) to the more speculative. These perceptions together yielded a personal – and potentially speculative - interpretation of their research environment that directly influenced their data sharing activities. These interpretations involved a range of personal traditions and concerns which were more or less grounded in evidence.

Nonetheless, as these perceptions shaped how the participants were engaging in data sharing activities, they remain valid to Open Data discussions. Indeed, ameliorating concerns about being scooped, or having data under-valued are as important as addressing issues of connectivity and bandwidth. Indeed, it is only through addressing the multifaceted concerns of a local scientific population that data sharing activities will become embedded within these communities. And this can only be done by engaging with the community and *taking their interpretations of their research environment seriously.*

5.1 | Re-examining data sharing in Africa: Areas for further action?

In recent years, there has been a marked proliferation of Open Data activities aimed specifically at LMICs. These range from data embargo policies to assist LMIC scientists to publish sufficiently from their data[23] to fee waivers for a variety of dissemination avenues.[24] There has also been a rapid rise in the number of institutional and national repositories developed by LMIC science stakeholders and interest in regional coordination.[25]

Nonetheless, to return to the comment from the South African professor, we must question whether these initiatives are really designed to ameliorate the concerns discussed above. Do they really get to the heart of the perceptions and concerns that my participants detailed in relation to their physical and social research environments? Do they in any way offer ways in which to change perceptions and thus change practice? Without such directed activity, it is highly possible that many LMIC scientists will remain "interested bystanders" in the Open Data movement, rather than embedded participants as their definition of their current situation still makes inaction a safe(r) option than data sharing.

5.2 | Lack of evidence

A key issue confounding attempts to embed data sharing practices in African science is a lack of evidence. This absence is twofold. First, there is a lack of detailed descriptions of African research conditions, and a marked absence of empirical engagement with African scientists on the topic of data sharing. Indeed, without such evidence it is difficult to see how Open Data discussions on (dis)incentives for African scientists will be able to progress. Indeed, by assuming that the global community not only understands the concerns of these communities, but is able to prescribe "sticks and carrots" may be seen by some as bordering on hegemony.

Second, there is a lack of evidence offered by the Open Data movement detailing the incentives of data sharing for African scientists. Amongst the scientists that I interviewed there was little evidence of gains from data sharing incentives - either personally or via peers. For example, when talking about contributing to a professional networking site, one Kenyan scientist made the following comment: *I can't see what [a professional networking site] has contributed to me. I don't know why. They say it is another way of measuring how successful a researcher is. And they say that normally I am better than 90% of RG users, so I'm wondering how that is good for me. Because I don't see any*

[23]For example, see the MalariaGen consortium's approach to sharing data. MalariaGen C2017 [cited 28 Aug 2017]. Available at: https://www.malariagen.net/data/our-approach-sharing-data. While such initiatives predominantly cover the sharing of human data, these practices are expected to extend more broadly throughout all activities in the network. See de Vries J, Bull SJ, et al. Ethical issues in human genomics research in developing countries. BMC Medical Ethics, 2011: 12; 1-10.

[24]For example, see the BioMedCentral waiver of author processing charges for authors from LMICs. BiomedCentral C2017 [cited 28 Aug 2017]. Available at: https://www.biomedcentral.com/getpublished/article-processing-charges/open-access-waiver-fund

[25]For example, see the developing African Open Science Platform. Details can be found at: Academy of Science of South Africa [Assaf] C2016 [cited 28 Aug 2017]. Available at: https://www.assaf.org.za/index.php/news/322-african-open-science-platform-to-boost-impact-of-open-data-for-science-and-society

good news coming out of it – someone saying we want you here to do this or that, or give a talk. I've never seen anything (KY1/8).

Lack of evidence of personal gains was something that was regularly repeated. In part, this was due to a lack of discussion about data sharing within their institutions, but also due to a marked lack of national discussion, evidence of "good examples" and effective mobilization of professional networks and organizations. This situation appeared to be pervasive amongst all levels of academics, as one South African scientist observed. *Yeah, it's not something they teach you in undergrad. It's often not something even your supervisor has worked with a lot because I guess it's a kind of a very modern way of doing research because never before has there been this much data available. So, that's the other thing. The student is almost, usually, the first one in the group to have the experience, so it's hard* (SA2/5). However, if the levels of sharing data remain low, the chances of building up examples that clearly demonstrate the value and utility of data sharing remains remote.

6 | CONCLUDING COMMENTS

The observation that the *definition of a situation* influences action, and therefore both the present and the future should be a key element in discussions on data sharing. Nonetheless, this is by no means straightforward. While most discussions recognise the concerns of scientists towards disseminating and re-using data online, they are often used to elaborate on the imperative that *data sharing is good*. Thus, discussions about these concerns (such as those highlighted by the Wiley survey) are often focused on how these perceptions lessened by the incentives, can be addressed by strengthening the prevailing norms that scientific communities endorse.

Alternatively, they suggest that existing policy or directed changes could ameliorate these concerns.[26] As a result, little credence is given to the core tenant of the Thomas Theorem: that the concerns held by scientists' regarding data sharing activities are based on their personal interpretations of their situations, and thus may persist despite the provision of evidence or guidance. The key, it would seem, is to change actions through *changing perceptions*.

Improving data sharing practices can thus not solely be a case of increasing evidence-based incentives and ameliorating evidence-based concerns in response to an objective view of research environments. It is about changing how scientists view their environments as well. Indeed, as the participants of this study all recognized both the importance and the benefits of sharing data, it is hard to see how externally constructed codes of conduct or policy documents would change their current behavioural patterns. Similarly, while mandated sharing could enforce change, there is always the question of how much scientists are solely paying "lip service" to the expectations instead of embedding them in their daily practice.[27]

Incentivization should thus not just about finding a carrot or stick to shape behaviour in response to external interpretations of environmental pressures. Instead, incentives must be thought of as highly complex and contextual with strong cultural content. Properly addressing incentivization in policy must therefore include adding long-term value and advance individual and community aspirations, pragmatic short-term rewards, are in line with accepted community values and have community support. Similarly, addressing disincentives must consider daily research challenges of individual scientists, achievability of data requirements *in situ*, full scope of data engagement activities. In this, it is very important that scientists are canvassed to get a proper understanding of the research conditions on the ground, so as to create policies and practices that truly address the context in which the scientists are working.

ACKNOWLEDGEMENTS

The author would like to thank all participants at the four fieldsites. She would also like to thank her collaborators on this project, Prof Brian Rappert, Prof Sabina Leonelli and Dr Ann Kelly, for all their input and support.

CONFLICT OF INTEREST

No conflicts declared.

ORCID

Louise Bezuidenhout http://orcid.org/0000-0003-4328-3963

Louise Bezuidenhout is a research fellow at the Institute for Science, Innovation and Society, University of Oxford. Her research interests are broadly centred on data sharing issues within the life sciences. In particular, she is interested in how the data produced during scientific experimentation enters into circulation, and how it is valued by potential downstream users. Her work involves a strong empirical component, including a number of ethnographic studies in laboratories in the UK, USA, Kenya, Uganda and South Africa.

Trained in both the life and social sciences, Louise holds a PhD in Cardiothoracic Surgery from the University of Cape Town (South Africa, 2007) and a PhD in Sociology from the University of Exeter (UK, 2014). She also holds an MA in bioethics from KU Leuven (Belgium, 2008). She has previously worked as a research fellow at the University of Exeter and the University of Notre Dame (USA). In addition to her work at InSIS, Louise remains an honorary lecturer at the Steve Biko Centre for Bioethics at the University of the Witwatersrand (South Africa) and is active in various initiatives advocating for better education on Open Data for scientists.

How to cite this article: Bezuidenhout L. To share or not to share: Incentivizing data sharing in life science communities. *Developing World Bioeth*. 2019;19:18–24. https://doi.org/10.1111/dewb.12183

[26]Tenopir. *op. cit*. note 12; Kaye. *op. cit*. note 9; Mauthner N, Parry O. Open access digital data sharing: principles, policies and practices. Social Epistemology. 2013; 27(1): 47–67.

[27]For another example see Bezuidenhout L. Ethics in the minutiae: examining the role of the physical laboratory environment in ethical discourse. Science and Engineering Ethics. 2015; 21(1): 51–73.