DATA SCIENCE JOURNAL

Bezuidenhout, L 2017 Technology Transfer and True Transformation: Implications for Open Data. *Data Science Journal,* 16: 26, pp. 1–13, DOI: https://doi.org/10.5334/dsj-2017-026

RESEARCH PAPER

Technology Transfer and True Transformation: Implications for Open Data

Louise Bezuidenhout
University of Oxford, GB
louise.bezuidenhout@insis.ox.ac.uk

When considering the "openness" of data it is unsurprising that most conversations focus on the online environment – how data is collated, moved and recombined for multiple purposes. Nonetheless, it is important to recognize that the movements online are only part of the data lifecycle. Indeed, considering where and how data are created – namely, the research setting – are of key importance to Open Data initiatives. In particular, such insights offer key understandings of how and why scientists engage with in practices of openness, and how data transitions from personal control to public ownership.

This paper examines research settings in low/middle-income countries (LMIC) to better understand how resource limitations influence Open Data buy-in. Using empirical fieldwork in Kenyan and South African laboratories it draws attention to some key issues currently overlooked in Open Data discussions. First, that many of the hesitations raised by the scientists about sharing data were as much tied to the speed of their research as to any other factor. Thus, it would seem that the longer it takes for individual scientists to create data, the more hesitant they are about sharing it. Second, that the pace of research is a multifaceted bind involving many different challenges relating to laboratory equipment and infrastructure. Indeed, it is unlikely that one single solution (such as equipment donation) will ameliorate these "binds of pace". Third, that these "binds of pace" were used by the scientists to construct "narratives of exclusion" through which they remove themselves from responsibility for data sharing.

Using an adapted model of technology first proposed by Elihu Gerson, the paper then offers key ways in which these critical "binds of pace" can be addressed in Open Data discourse. In particular, it calls for an expanded understanding of laboratory equipment and research speed to include all aspects of the research environment. It also advocates for better engagement with LMIC scientists regarding these challenges and the adoption of frugal/responsible design principles in future Open Data initiatives.

Keywords: technology; low/middle-income countries; data sharing; research; pace

The issue of increasing the openness of data online is a global priority. Indeed, Open Data is increasingly featuring on agendas of both high- and low/middle-income country development plans (Schwegmann 2012). Nevertheless, data sharing in low/middle-income countries (LMICs) is challenged by a number of widely-recognized issues. These include a lack of resources for sharing activities (Bull 2016) as well as for research activities more generally. Strategically increasing research capacity in LMICs – and thus the ability of LMIC researchers to participate in the Open Data movement – is intrinsically tied (at least in part) to the need for increasing the availability of laboratory and ICT equipment.

Unpacking the Links Between Laboratory Equipment and Open Data

It is recognized that the lack of up-to-date laboratory equipment hampers not only the ability to conduct certain types of research, but has an overall impact on the pace and efficiency of research. How to best address this lack of physical research resources is becoming a topic for directed intervention, and a number of different organizations have been set up to address issues relating to equipment provision. These include

databases of equipment,[1] equipment donation schemes[2] or equipment collaborations, as well as increased equipment budgets in many funded grants.[3]

Despite the value of these initiatives, a coordinated and sustained approach to research equipment in LMICs remains elusive for two key reasons. First, a lack of empirical evidence detailing the contextual heterogeneity of LMIC research environments challenges targeted interventions. Second, the absence of LMIC scientists in more general discussions on scientific research practices makes it difficult to pinpoint key issues that may be prevalent within these research settings. Thus, capacity building initiatives are often challenged by the absence of a clear picture of what equipment are needed and best deployed in LMIC regions. It is therefore highly possible that other interventions are critically needed if this resource shortfall is to be effectively addressed.

The challenges of increasing research capacity through equipment-related interventions have far-reaching implications for LMIC research. In this special edition, and in related papers (Bezuidenhout et al 2016; Bezuidenhout and Rappert 2016; Bezuidenhout et al forthcoming), we argue for a stronger connection between the discussions of Open Data and the research environment in which data are generated. The physical – as well as the social and regulatory aspects of research environments – influences how scientists are able to create, curate and disseminate data, and thus the ability of scientists to contribute and re-use data online. Moreover – and often overlooked – the characteristics and challenges of personal research environments can influence the importance that scientists attach to the Open Data movement (Bezuidenhout et al 2016; Bezuidenhout and Rappert 2016; Bezuidenhout et al forthcoming).

Nonetheless, in many discussions on Open Data there is an absence of robust discussion on the influence of the physical research environment on data engagement activities. This paper examines this issue in more detail examining four interlinking questions: first, to what extent do issues relating to technology affect the pace of research in these laboratories? Second, could these issues of pace be ameliorated by the directed provision of more equipment – particularly high-level, specialized machinery? Moreover, how can reflecting on issues to do with technology contribute towards more inclusive discussion surrounding Open Data? Finally, how can a better understanding of research technologies enable more contextually-sensitive discussions about data engagement?

In order to unpack these questions in detail, the paper discusses qualitative fieldwork conducted in four African laboratories between 2014 and 2015. This fieldwork was designed to investigate data engagement activities amongst scientists working in resource-limited environments. From these interviews the paper highlights how issues of data engagement and issues of equipment provision were inextricably intertwined and often interdependent. If these issues are to be effectively addressed in Open Data discussions, the paper suggests that an expanded definition of "research technologies" is necessary. Using a model proposed by Elihu Gerson, the paper then offers key ways in which the critical issues of technological contextuality can be effectively implemented into Open Data discourse.

It's Not Just the Equipment

When considering laboratory equipment and research it is tempting to make the assumption that more – and newer – equipment leads to an more productive research that is conducted at a faster speed with increased outputs (such as data). Indeed, such assumptions drive many of the equipment-focused initiatives mentioned above. Similarly, it is tempting to extend such assumptions to Open Data conversations. If more equipment will facilitate the faster production of increased amounts of data, the argument would go, then scientists will be more able (and willing) to share their data online.

While these arguments make a compelling case, examination of the current status quo indicates a need for caution. Indeed, if the causal links between equipment provision, increased research pace and improved open outputs were that straightforward, data sharing should be markedly increased by the provision of (any) laboratory equipment. Such questions motivated a period of embedded fieldwork in Kenya and South Africa between 2014 and 2015. I wanted to examine how scientists in low-resourced research settings engaged in Open Data activities and discussions – and whether their physical laboratory environment had any influence over this engagement.[4] Over the course of the year I spent 3 – 6 weeks in 4 different chemistry laboratories and conducted 56 semi-structured interviews with researchers and postgraduate students to find out what

[1] Such as the EPSRC's database www.equipment.data.ac.uk (discussed later).
[2] Such as Seeding Labs (discussed later).
[3] For example, see http://www.esrc.ac.uk/funding/guidance-for-applicants/changes-to-equipment-funding/ (accessed 03/05/2017).
[4] A full description of the methodology is given in the appendix.

was working in their research environments, and what challenged their ability to generate, curate, store, share and re-use data online.

When coming to analyze the interviews, the issue of *pace* in research was unavoidable. Indeed, it was everywhere. Concerns about the slowness of research, and the pressure to speed it up pervaded how the scientists talked about their research, valued their data, identified threats to their sovereignty and acquisition of credit, positioned themselves within the scientific community, and evaluated the international community's efforts to assist them. These issues have been discussed in other papers (Bezuidenhout et al 2016; Bezuidenhout and Rappert 2016; Bezuidenhout et al forthcoming) and will not be covered here. Instead, this paper takes a step back to look at *why* there was this overwhelming awareness of pace in these laboratories. What aspects of the laboratory equipment played key roles in controlling the pace of research, and consequentially the engagement of scientists in Open Data activities.

The equipment is …

The laboratories that I visited were not members of high profile consortia or integrated into well-funded foreign research networks. Rather, they were good examples of home-grown science. They produced high quality research, but were dependent on their funding from multiple national and international sources. Moreover, their facilities – and the budget to maintain or upgrade them – were provided by their host institutions. This created a bind for the researchers, as the facilities provided were often minimal and/or badly maintained, and their institutions did not have large amounts of "core funding" for upgrades. As one Kenyan participant said:

> *"We get no funding from the government. We get paid from the government, we get bills of power and water by the government but otherwise, other than that, the materials that we need for research we have to source from funding agencies."* (KY1:8)

Similarly, as most of the funding for their research came from project-specific grants, the researchers had few opportunities to secure money for standard laboratory equipment or general laboratory maintenance. A participant in South Africa eloquently said, when talking about her research that:

> *"[it] is a challenge because the university doesn't offer a start-up fund for equipment. … I would need to pay bit by bit and one by one. When I have funding then buy one piece of equipment and maybe after 5 years I would have my lab."* (SA2:11)

Moreover, even when the money was there, many of the participants said that they experienced problems accessing it, or using it to address the challenges that they identified in their daily research environment. This is evident in a quote by another South African participant who said:

> *"It's really bad – the bureaucracy of it. It's how the money is transferred, technical services, procurement, all those … but those are like "grand problems" that you can't solve."* (SA2:6)

Thus, a lot of the discussions I had about research and data engagement became discussions about equipment and research environments. The researchers I interviewed highlighted a number of key issues that affected the pace of their research in comparison (in their opinion) to well-resourced laboratories. In particular, the statements related to the "un-usability" of the equipment that was available for them to use. These statements are broadly grouped under the headings below.

… not there …

One of the most common complaints I heard in all four laboratories was that the equipment available for research curtailed the types of research that could be done by the researchers. While this is, of course, an issue for scientists around the world, for many of the researchers that I interviewed this was almost a deal-breaking aspect of their research plans. As one Kenyan participant observed:

> *"the lack of equipment limits the extent to which you can do research – and even the type of research that you want to do. And you ask yourself, ok, so I want to do this kind of research but do I have the machinery?"* (KY2:3)

Similarly, a participant from the other Kenyan site said:

> *"[o]ur labs are not even there for synthesis – synthetic work – the environment is not there. So when it comes to that I either have to skip it or I have to go to a lab that has such facilities."* (KY1:3)

These constraints not only shaped the research being conducted in these environments, but also necessitated that a number of researchers change the direction of their research in order to fit in with the equipment available. Particularly in Kenya there were a number of lecturers and professors who had done postgraduate training in the UK or the USA, but were unable to capitalize on their research experiences back home. This was described by one Kenyan professor who said:

> *"the kind of research which is taking place here is a bit different from what I was doing – like in the UK I was doing synthetic organic chemistry. And the kind of equipment and the rest, it was purely on silicone chemistry and the reagents and the rest I couldn't get them here. So what I had to do was to look for things which are relevant for this institution."* (KY1:1)

In addition to shaping the *types* – and thus the broad pace – of research, the lack of equipment also had an impact on the daily pace of research activities in the laboratory. This is evident in the exchange below, where the participant (a postgraduate student) explains day-to-day practices within the laboratory. In particular, he highlights how sharing basic equipment plays a highly influential role on how much he can work on a day-to-day basis, and thus how much data he can produce. As there were six postgraduate students sharing one evaporator, one can only imagine their frustration.

Participant: *the solvents and reagents we have all, but the equipment – some equipments are missing. But we do the best we can.*
LB: *and with so many in the lab there must be high competition to use the equipment.*
Participant: *yeah! For example, this evaporator, we all use it. So we have to use it at a certain time and you when you leave it the other person wants to use it and so on and so on.*
LB: *so there is a schedule.*
Participant: *so for us to work very well, so everyone should have at least an evaporator like this so that you can use it at any time. In that case it can become very easier, instead of sharing – it's not easy* (KY1:6).

The absence of laboratory equipment thus created two different pace-related binds for the researchers that were interviewed. Not only did it shape the types of research that could be conducted, thus affecting the long-term pace of research, but it also shaped the pace of daily research. In this, it was often the absence of multiple copies of generic equipment – evaporators, Gilson pipettes, glassware, water baths and so forth – that played key roles in slowing down the amount of experiments that could be done by one individual on a daily basis.

... broken ...

At one of the Kenyan universities I was given a tour around the laboratories and shown the available equipment for research. I took the following note in my field diary after some discussion with my guides:

> *"This department has been donated an NMR machine by a laboratory in the USA. When it arrived it needed to be calibrated and set up. It would also seem that some parts needed to be replaced in order to get it working. However, there is no technical support for this make and model [it is an older version of the current one on the market] in East Africa, and the only place with spare parts and a qualified technician is in South Africa. This creates situation in which they are expected to be grateful for donations, but age of machine and lack of funds for upkeep makes it obsolete before it is delivered."* (KY2 field diary: day 3)

Indeed, the NMR machine had never been in use, as the laboratory lacked the money to fly the technician from South Africa. Such lack of technical support and funds available for maintenance and upkeep were often key issues for the researchers interviewed. It was apparent that even the equipment that was bought using project funds was vulnerable to this situation after the end of the grant. Thus, while it may be assumed that many of the laboratories in sub-Saharan Africa possess quality research equipment, the lack of technical support – together with the rapid obsolescence of models of research equipment – cause this equipment to stand un-used.

It must be noted that many of the participants made use of some sort of equipment sharing – either by partnering with geographically close institutions, or by sending samples away. One South African participant described this, saying:

> *"[i]t's only now we are starting collaboration in terms with sharing equipment because previously they didn't have any equipment so they were using ours but now ours is broken down and we are going back to them."* (SA2:2)

Nonetheless, every single participant who discussed equipment sharing mentioned the time and frustration of not being able to do experiments *in situ* – and the waste of time and resources necessary to take experiments to a different laboratory.

The inability to make full use of the equipment available was a source of considerable frustration to many of the scientists interviewed. Moreover, they perceived a lack of agency in being able to ameliorate these situations due to the constraints of project-specific funding, lack of core funding, and an absence of other pots of money that could be tapped into for repairs and maintenance. As one PI in South Africa said:

> *"[y]ou know they call us to meetings and they say we have funding for this and that. And I think "great stuff", but I wish they would ask me what the real issues are. I'll probably tell you 100 other things outside of the money [permitted to be spent on the grant]."* (SA2:1)

… not running …

Related to the problems experienced with broken equipment were not having the reagents or infrastructure to use working equipment. This was eloquently described by one of the Kenyan participants, who said:

> *"[o]ur equipment is not running or idle. We have an AS that is not operating, because we have no fume hood and now no acetylene gas. Because of this it has been idle for 6 years."* (KY1:9)

Similarly, in South Africa, one participant described the challenges of working in a geographically-isolated university, saying:

> *"it has been very challenging [having the NMR machine] – it's a baby that you have to nurse all the time. Also for the liquid nitrogen that we need at first we couldn't get a source of liquid nitrogen north of [a major metropolitan area 6 hours away]."* (SA2:6)

The difficulties of ensuring regular supplies of reagents, electricity and internet connection often had a significant impact on the ability of the researchers to run what equipment was available to them. Consequently, the pace of their research slowed down almost as much as if the equipment were broken or missing.

Open Data, Technological Difficulties and the Slow Pace of Research

As detailed from the fieldwork above, the scientists in the laboratories I visited often experienced challenges to their ability to work effectively. Absent, broken or poorly maintained laboratory equipment slowed down their research, and delayed the production and subsequent analysis of research data. Interestingly, these challenges played a big part in how they discussed their involvement – or lack thereof – in Open Data activities. Indeed, while most of the interviewees were supportive of data engagement activities in theory, there was not much data engagement occurring on a daily basis. These issues are elaborated on below.

A need for speed

Many of the scientists that I interviewed believed that the slower pace of their research (in comparison to HICs) left them at a disadvantage when it came to data release – particularly in terms of pre-publication data release. This is evident in the exchange below:

Participant: *But no in the fact that maybe I'm here in the lab doing something and someone is out there in Europe and they do the same research as me and published before me so my work will be null and void.*
LB: *so you're concerned that by making your research available other people might beat you to the post.*
Participant: *Yes. Because it may be null and void but you've been in the lab for almost a year.*
LB: *do you think it is influenced by the resource difference between the North and the South?*

Participant: *We're in Africa, right. That is the West – they definitely have more advanced stuff than us. So if I'm doing this research for one year, someone in Europe of the US they can do it in 3 or 4 months. So that is where now the issue* (KY1/4).

This concern about speed of data analysis has been reported by other researchers (The Malaria Genomic Epidemiology Network 2008; Parker et al 2009), and has already influenced a number of data release expectations by funders and consortia (such as MalariaGEN and H3Africa consortia). These initiatives focus predominantly on ensuring that scientists in LMICs get extended periods of time on completion of the project to process and analyze the data generated (The Malaria Genomic Epidemiology Network 2008; Parker et al 2009).

While extended data moratoria at the end of the project is undoubtedly valuable to enable the maximum number of publications from a research project, the quote above highlights that more is needed. What became apparent from the conversations with fieldwork participants was that they were conscious of the pace of their research *throughout* – and that being slow at producing data was as pertinent as taking longer to analyze the final product. This highlights a key oversight in current data discussions, where there is no sensitivity to how mid-project data releases can be safeguarded for researchers who necessarily take longer to complete their research projects due to resource limitations.

What the fieldwork identifies is the need for corresponding efforts to address the issues relating to the varying pace of data generation. More reflection and productive policies are needed that address the multitude of issues that cause this slower pace in daily research activities. Specifically, this links directly to the types, availability, maintenance and provisioning for the equipment in the laboratory. If the entire research process occurs at a slower pace, it is unlikely (as the quotes show), that many researchers will risk sharing pre-publication data, methodologies, and other resources. This is of particular importance for scientists not involved in international research networks, and who do not have extensive support systems to draw on.

Data quality

Another key theme that emerged from the discussions about data sharing was that, many of the participants were concerned that – even if they did release the data – it would not be re-used by their international peers. One researcher in Kenya highlighted this, saying:

> *"[t]here is a constraint. Even the conditions aren't right, so you cannot work as fast. One of the limitations is of facilities. I mean facilities that can't be considered credible for some publication. If the instruments that are there are really elementary so you have to search for instruments that aren't here and that takes some time."* (KY2:15)

This was eloquently reiterated by another of his peers, who said:

> *"how much can we do to develop our own data? What processes do we need to convince people that the data are good?"* (KY2:13).

Such statements show a distinct anxiety over the data that are being produced that is linked to the types of equipment being used to produce it. If, as is suggested in the quotes above, the equipment is older, and the methodologies are more basic, how will the data be viewed by international peers? Would it, as some of the participants suggested, not be viewed as of equal value if it is shared? In other words, would the data created in low-resourced settings be re-used at all if it is released online? Such observations link the pace of technological change within research communities to perceptions of data sharing – something that has not yet been examined in Open Data discussions.

Perceptions of data becoming obsolete based on the equipment and methods used to generate it are contentious and the validity of such positions may be argued. Nonetheless, it has far reaching consequences for the Open Data movement. In a way, it may be said to offer an example of the Thomas theorem.[5] If the researchers believe that their data will be judged based on the age of their equipment and methods, they

[5] The Thomas theorem was formulated in 1928 by W. I. Thomas and D. S. Thomas and states that "if men define situations as real, they are real in their consequences".
The child in America: Behavior problems and programs. W.I. Thomas and D.S. Thomas. New York: Knopf, 1928: 571–572.

will be less inclined to go through the effort of sharing. This is particularly the case if they believe that their work will be heavily scrutinized, overlooked, or rendered obsolete upon arrival. Consequentially, the pace of research is slowed down by researchers not sharing data, or delaying its release.

Together, these two issues were highly influential in mediating the interviewees predominant lack of involvement in data engagement activities. In particular, these two issues had key effects on the lack of pre-publication release of data and the participation in online knowledge transfer (through posting of presentations online, contribution to discussion forums, release of methodologies, and so forth). The specter of "being scooped" due to the slower pace of research, coupled with the helplessness of changing the pace at which data were generated thus led to a situation in which the scientists recognized the value of increased openness in science but did little to engage.

Technology Transfer: the Solution to Pace and Openness?

The fieldwork described above clearly highlighted two key issues. First, that the speed of research in the laboratories that I visited was influenced by the technologies available. This impacted on research productivity, data production, research efficiency and the optimal use of funding resources. Second, that the issues of pace in research were intimately connected to how scientists valued their research – and subsequently how they conceived their responsibilities to be involved in data engagement activities.

In a way, the researchers that were interviewed constructed a "narrative of exclusion" in which they (in) voluntarily opted out of Open Data activities. This narrative was constructed around perceptions of the pace of high-income country science, and their inability to match this pace in their own research. The existence of these perceptions, and the preferred exclusion that the narrators often choose, is rarely acknowledged in Open Data discussions.

The obvious solution to these problems – the solution to slower research in LMICs, to less data engagement, to more visibility of LMIC research – would appear to invest in the equipment present in these low-resourced laboratories. By providing more equipment to researchers currently working with the pressures described by the interview participants would seem the logical step out of this current conundrum … or does it?

Problems with technology transfer

Initiatives such as Seeding Labs[6] and the Sustainable Sciences Institute[7] have been influential in partnering HIC donors of equipment with LMIC applicants, and have considerable testimonials to bear witness to their positive impact. Nonetheless, a recent article on Seeding Labs noted that: "recipients pay a fraction of the equipment's cost to offset the logistical expenses — although Seeding Labs refuses to say how much — and, as buyers, they assume responsibility for setting up and maintaining the equipment".[8]

In light of the difficulties experienced by the Kenyan and South African laboratory in setting up their NMR machines (see fieldnote above and SA2:6), or the Kenyan laboratory's struggles with their AS machine (KY1:9), it is important not to see these initiatives as a blueprint examples for generic success. Rather, the careful matching of donors and recipients, mentoring during the process of donation and a careful analysis of what is required from the target sites are all necessary to ensure success.

Similarly, efforts to create equipment databases to facilitate inter-institutional sharing in LMICs have also struggled with similar problems. Informal discussions with scientists regarding such initiatives have brought to light key contextual concerns, such as how the user/provider relationship will cope with issues such as payment and sourcing of reagents, maintenance, technical support, and possible damage instances. Such concerns, it must be noted, are not unique to LMICs and have similarly been discussed in relation to the EPSRC equipment sharing portal for UK universities.[9]

These problems highlight two key concerns. First, that current approaches to technology transfer often do not take into consideration the limitations of the context in which it will be used. Providing equipment without the researchers having a sustained ability to get the reagents necessary to run it is highly problematic. Second, current approaches to technological transfer often do not take into consideration the difficulties of moving technologies across different contexts. As described by my fieldnotes from the first Kenyan society, there is

[6] http://seedinglabs.org (accessed 03/05/2017).
[7] http://sustainablesciences.org (accessed 03/05/2017).
[8] http://www.scidev.net/global/capacity-building/feature/recycled-kit-equips-african-labs-.html (accessed 03/05/2017).
[9] www.equipment.data.ac.uk and personal communication with UK institutional research officers delegated to collate institutional equipment lists.

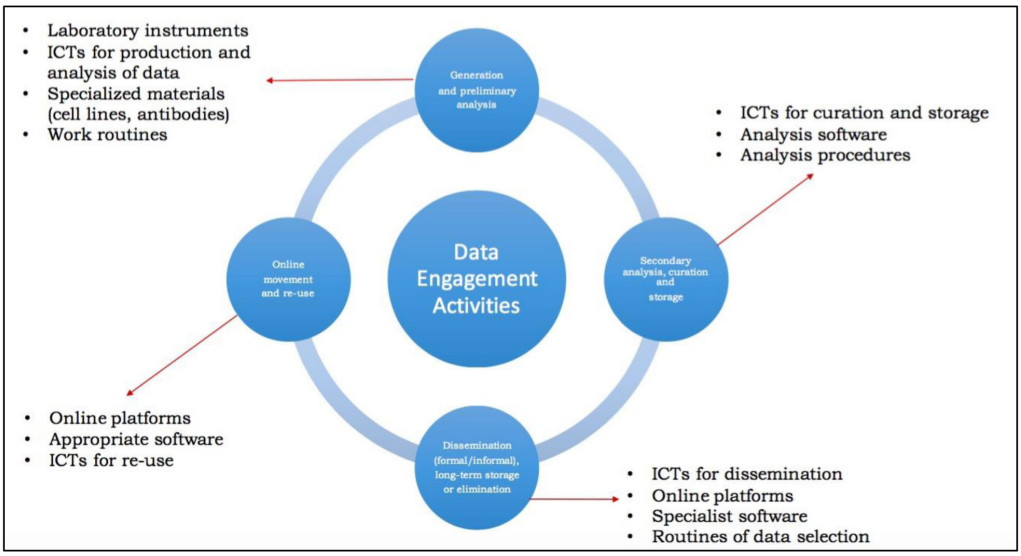

Figure 1: Technologies associated with data engagement.

no value in a piece of equipment that cannot be calibrated or maintained due to a lack of qualified technical support.

While the implications for broader, capacity-focused discussions are apparent, these observations also have important consequences for future Open Data discussions. What the evidence clearly suggests is that relying on LMIC researchers receiving *more* equipment will not necessarily influence the speed of their research, nor their willingness to share data. Thus, structuring projections on LMIC involvement in the Open Data movement based on a linear model of technology transfer/research productivity is highly problematic. What Open Data discussions need is a new model of technology/pace that takes these issues into account.

Gerson's model of technology transfer

First and foremost, it is important to critique exactly what is needed from a definition of "technology". It is often tempting to equate "technologies" to the equipment used within the laboratory – in particular, the specialized (and often high-tech) machines such as NMR, PCR, chromatographic apparatus and so forth. In contrast, as evident from the fieldwork, these pieces of equipment are not the only causes of the pace issues experienced by the researchers interviewed. Indeed, the student discussing the lack of multiple evaporators in his lab (KY1:6), or the researcher who struggled to get liquid nitrogen for the AS machine (SA2:6) experienced similar problems of pace. Moreover, dissociating discussions of equipment from those relating to their running costs and infrastructural requirements is evidently limiting.

With this in mind, it is helpful to make use of a recent definition of "technologies" proposed by Elihu Gerson in his 2015 lecture at the International Society for the History, Philosophy and Social Studies of Biology (ISHPSSB).[10] He proposed that:

> *"[t]echnology can include instruments, specialized materials such as cell lines, model organisms, enzymes, antibodies etc. It also includes specialized codified procedures, such as those used in psychology, field observations etc."*

This expanded notion of technologies allows us to draw in the many different issues that were raised by the fieldwork participants – including the difficulties of getting reagents, of setting up laboratories and protocols, of having instruments available, and also having the expertise necessary to utilize the equipment.

Second, Gerson draws attention to the difficulties of moving technologies across research contexts. He highlights six ways in which attempts to introduce technologies into new environments may be problematic. These include:

[10] *Gerson 2015 ISHPSSB: Resituating new data collection technologies.*

- Materials and equipment are recalcitrant.
- Researchers can't anticipate every contingency in a situation.
- Resituating new technologies requires coordination between source and target sites.
- Repertoires for work at the target must be developed.
- New technologies must be registered at the target site.
- New technologies address phenomena in new ways.

Moreover, it is apparent that a failure to address these concerns can result in incomplete re-situation of technologies that significantly slows down research processes and stops effective data engagement. This is evident in many of the quotes from the fieldwork, such as the idle AS machine at one of the Kenyan sites (KY1:9), or my field notes description of the donated NMR machine. It would thus appear that effective data engagement – generation, storage, curation, analysis, dissemination and re-use – is directly tied not only to issues of technology provision, but also of (in)effective manner in which the technology is introduced into the new context. Thus, without careful attention to the contexts in which technologies are to be deployed, making assumptions about its efficacy is highly problematic.

Such observations are of particular importance to Open Data discussions that should – ideally – consider the process of data production in its entirety. **Figure 1** outlines this in detail, identifying key areas of the research data cycle. While each stage is associated with technologies that could speed up research, introducing these technologies without correctly situating them within the specific context could undermine these efforts.

Figure 1 – and the accompanying empirical evidence – highlight the potential issues of that could accompany well-meaning technology provision and undermine effective data engagement. Using the adapted version of Gerson's model clearly highlights how Open Data discussions – instead of relying on the provision of high-level laboratory equipment as a means of speeding up research and encouraging sharing – needs to carefully consider the research contexts, use expanded interpretations of "technologies", and pay attention to the repertoires necessary to use available technologies. Such awareness also needs to be reflected in the design of future initiatives, so as to safeguard against the hidden binds of pace that accompany inappropriate re-situation of technologies. Without such an expanded focus it is likely that LMIC scientists will continue to underperform in Open Data initiatives.

Avoiding "Insidious Inequalities"

Issues associated with the pace of research and incomplete technology transfer also extend beyond the sites of data creation in LMICs. While many LMIC countries are rapidly expanding their internet capabilities, incomplete integration of these capabilities into immediate working environments – through lack of expertise, older equipment, poor maintenance and technical support and infrastructural challenges (such as power provision) – continue to challenge effective usage. As one Kenyan participant observed:

> *"[i]n Kenya people say that we have internet everywhere but really how much can you download, and you have to have the equipment to be able to. You bought the data bundle but what you have is not enough for you to download any publication or anything like that. Some areas in Kenya we know that people can't even access. Although we know the networking has been done but there is an assumption that everyone can access."* (KY1:2)

For LMIC scientists, it would therefore seem that the connection between "pace" and technology is inescapable. Without directed interventions that address the integration of technologies *in situ* it is likely that scientists will continue to operate at the slower speed that currently characterizes much of the research in these areas. The evidence from the fieldwork presented in this paper suggests that this has multiple implications for the Open Data movement – both in terms of practical engagement and ideological buy-in.

Co-partnership with other initiatives aimed at addressing equipment provision, the integration of responsive design principles into data platforms, and the provision of funds to ameliorate hurdles to data generation and dissemination will all assist in changing this current paradigm of insidious inequality. As said by a PI in South Africa:

> *"[t]he disadvantaged are still disadvantaged and that is the fact of the matter. The government may be willing to address the gaps, but there are still gaps."* (SA2:1)

But with an increased awareness of the contextuality of data engagement it is likely that the Open Data movement can move beyond such accusations.

Building Capacity in Research … and in Open Data

Combining the fieldwork above with Gerson's model of technology transfer makes a compelling case for prioritizing the issue of *pace* in Open Data discussions. Moreover, it clearly highlights the need for further, expanded discussions on technical capabilities necessary for data engagement. It thus becomes important to ask, what can be done to effect such a change?

Recognizing problems, designing solutions

When discussing the issues of equipment limitations, a South African researcher made a telling remark. She said:

> *"[b]ut 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)

In this comment she was specifically talking about the difficulty of registering online for international conferences where the high-resolution of the websites made it difficult to use in her low-bandwidth area. She described a long and annoying process of attempting to get the site to work, after which she ended up giving up and emailing the organizers. Nonetheless, as she said, creating websites that were not usable in low-bandwidth areas was not an intentional slight by the organizers. Rather, they were not aware that it could be problematic for many researchers in low-resourced environments.

Unpacking this comment leads to a number of different issues. First and foremost, it draws attention to the lack of awareness by scientists, funders and affiliated stakeholders in high-income countries about these problems. While many are aware that there are some "equipment issues" in low-resourced research settings, very few have a consistent and coherent impression of what these problems actually are – particularly when using an extended interpretation of technologies.

This lack of awareness is compounded by the rarity of detailed ethnographic analyses of technological challenges in low-resourced settings. While a number of studies on working conditions in low-resourced environments are gradually emerging (Bezuidenhout 2015; Fine 2007; Harle 2010; Bezuidenhout et al 2016), the focus of these studies are diverse and a systematic collation of the evidence with regards to research technologies is urgently needed.

Such awareness will be of critical importance to future Open Data discussions. As evident from the fieldwork, issues of the availability of effective technologies– and the resultant impact on research pace – were key contributors to the lack of involvement in data engagement issues in all four institutions that I visited. Not only did the lack of effective and integrated technologies slow down the pace of data generation, but also the lack of up-to-date ICTs (and the corresponding infrastructures and repertoires) cause difficulties in all aspect of data engagement.

Without support for these issues – and policies that directly address these issues – it is difficult to see how LMIC researchers can be effectively drawn into Open Data discussions. Indeed, current policies do little to assuage the fears that they have relating to the *pace*, leading (at least in part) to the lack of daily data engagement evident from the scientists I interviewed. As a result, the lack of contextually-sensitive policies not only led to lower levels of data contribution and re-use from scientists in these regions, but also influenced the manner in which the ideals and responsibilities of Open Data were discussed in these setting.

To counter this, Open Data discussions need to expand discussions on responsible innovation to include the expanded version of research technologies suggested by Gerson. Key lessons from the "frugal innovation" movement (Radjou and Prabhu 2015) could also be effectively incorporated to promote cost-effective and efficient design of research technologies that will more easily adapt to low-resourced settings. Similarly, best practice by consortia such as the Global Health Network[11] – where the capabilities of the "lowest common denominator become the guiding set of criteria for which software and webpages are designed – offer important lessons that could be extended further into discussions on Open Data. What issues, it is necessary to ask, are really slowing down research and altering data engagement, and how can current policies and initiatives be designed to address these issues more productively?

[11] https://tghn.org (accessed 03/05/2017) and personal communication with The Global Health Network website developers.

Creating "safe spaces" to discuss these issues

Most of the fieldwork participants I interviewed were very forthcoming about the challenges of their research environments. This forthrightness was consistent with the numerous informal discussions I've had with LMIC scientists at conferences, socially and in related projects. Nonetheless, the same issues that they were so willing to discuss – and had such robust opinions on – were rarely raised to the university governance, funders, collaborators and stakeholders that might be able to make some difference in ameliorating them.

This led to many discussions about why the fieldwork participants were both so willing and unwilling to discuss their contextual challenges. One Kenyan scientist put it very succinctly, saying:

> *"I was worried about applying for international funding because the facilities are poor and we have to deliver."* (KY2:2)

The fieldwork participants described a tension inherent in drawing attention to the limitations of their environment – particularly that it may have some negative impact on their ability to secure funding, disseminate results or form collaborations.

Challenges with the effective situation of technologies within research contexts, and the accompanying binds of pace that they create, thus rarely get raised to those in positions to effect change. It is thus apparent that scientists in LMICs need to feel comfortable and confident to raise these issues. How this can be done, of course, is open for discussion and by no means apparent. Nonetheless, creating a "safe space" in which they can do so is urgent. It may be possible that by operationalizing the micro-finance scheme described by Brian Rappert in this issue such a space may be created. A failure to encourage LMIC scientists to effectively discuss these binds of pace and technology has significant impact on the research in these areas. Moreover, it also has downstream effects on the uptake of Open Data ideals and the contribution of these scientists to the online data milieu (Bezuidenhout et al 2016; Bezuidenhout et al forthcoming).

In order to stimulate LMIC scientists' involvement with Open Data initiatives it is thus vital that they are able to raise the technical challenges in their work without fear of reprisal. Such fears, it must be noted, may be largely unfounded and funders, networks and collaborators may be happy to engage with LMIC scientists on ways to ameliorate these issues. However, the current stalemate between lack of awareness of the former's part and lack of forthrightness on the latter's can only be broken if stakeholders are explicit about their intention to help, and their willingness to engage in productive discussion without any suggestion of reprisal.

More holistic approaches to capacity building

In a similar vein to Rappert's article in this issue, the fieldwork evidence makes a very strong case for the need for alternative approaches to capacity building and research funding. Combining this observation with the Gerson's extended view on technology it is evident that what is needed are funding avenues that allow researchers to address the difficulties of technology transfer and the establishment of robust technology landscapes.

While project-specific funding is, of course, of vital importance to LMIC research, it is important to recognize that without corresponding support on research structure, infrastructure, technology re-situation, and establishment of social practices, research in these settings will continue at a slower pace to their HIC counterparts. This has significant implications for capacity building initiatives, and for the advancement of research agendas in these regions.

Similarly, this has significant implications for advancing Open Data ideals in these regions. This was recognized by a number of researchers, who drew strong correlations with the slow pace of research and the idea of Open Data. As one researcher in South African observed:

> *"[y]es we end up spending much more time than it would be in a western country. But even the power failures, for example, they are not part of data sharing but are part of the vicious circle."* (SA2:12)

Similarly, the incomplete integration of technologies into the *in situ* research context – particularly ICTs – was highly problematic for many researchers who would otherwise have been interested in participating both as a data contributor and user.

It would therefore be of particularly importance that Open Data initiatives start problematizing the range of research technologies necessary for effective data engagement. It is possible that considerable traction for promoting openness in research might be gained from partnering on equipment sharing initiatives, or on funding avenues that facilitate sample or staff exchanges between resource-limited settings. By tapping

in to initiatives that already address technological insufficiencies and their effect on the pace of research it is possible that a new area of discussion on openness can be leveraged.

All in all, this paper makes an argument for the importance of considering the technologies that underpin data engagement activities, and the ways in which they insidiously control the speed of research. By controlling the pace of research, in turn they control buy-in from science communities in ways that are usually overlooked. Without directed attempts to address the insidious inequalities in data engagement that accompany the binds of technology/pace, it is likely that LMIC scientists will continue to struggle to make full use of the Open Data movement, and that in turn will undermine the egalitarian ideals of the Open Data movement. In particular, a multifaceted understanding of the binds of "pace" is needed to ensure the visibility of scholarly products from LMICs. Without careful and sensitive attention to these issues, it is likely that LMIC scholars will continue to exclude themselves from opportunities to share data, thus missing out on improved visibility online (Neylon et al 2014). This impacts not only on their credibility (Piwowar and Priem 2013) as scientists, but also precludes the effective re-use of their research efforts (OECD 2007; Piwowar and Vision 2013).

Additional File

The additional file for this article can be found as follows:

- **Appendix.** Methodologies used in study. DOI: https://doi.org/10.5334/dsj-2017-026.s1

Acknowledgements

Thanks to Brian Rappert, Sabina Leonelli and Ann Kelly for their collaboration on the project *Beyond the Digital Divide: Sharing Data Across Developing and Developed Countries* (Leverhulme Trust, (RPG-2013-153). Thanks to the two reviewers whose comments strengthened the initial manuscript.

Competing Interests

The author has no competing interests to declare.

References

Bezuidenhout, L 2015 Ethics in the minutiae: examining the role of the physical laboratory environment in ethical discourse. *Science and Engineering Ethics*, 21: 51–73. https://www.ncbi.nlm.nih.gov/pubmed/24510311 (accessed 02/05/2017).

Bezuidenhout, L, Kelly, A, Rappert, B and **Leonelli, S** forthcoming Beyond the Digital Divide: Towards a Situated Approach to Open Data (accepted to Science and Public Policy 20/03/2016).

Bezuidenhout, L, Leonelli, S and **Kelly, A** 2016 "$100 Is Not Much To You": Open Access and Neglected Accessibilities for Data-Driven Science in Africa. *e-published in critical Public Health*: 17/11/2016. http://www.tandfonline.com/doi/full/10.1080/09581596.2016.1252832 (accessed 02/05/2017).

Bezuidenhout, L and **Rappert, B** 2016 *What hinders data sharing in African science?* Presentation and paper at Fourth Codesria Conference on Electronic Publishing (Dakar, 30 March–1 April). http://www.codesria.org/spip.php?article2564&lang=en (accessed 02/05/2017).

Bull, S 2016 *Ensuring Global Equity in Open Research*. London: Wellcome Trust. https://dx.doi.org/10.6084/m9.figshare.4055181 (accessed 02/05/2017).

Fine, J 2007 *Investing in STI in Sub-Saharan Africa: Lessons from Collaborative Initiatives in Research and Higher Education*. Washington DC: Global Forum (Building Science, Technology and Innovation Capacity for Sustainable Growth and Poverty Reduction). http://siteresources.worldbank.org/INTSTIGLOFOR/Resources/Investing_in_STI_Paper_Feb06.pdf (accessed 02/05/2017).

Gerson, E 2015 *Resituating New Data Collection Technologies*. Presentation at the ISHPSSB (Montreal, June 2015). https://www.academia.edu/20947756/Resituating_new_data_collection_technology (accessed 02/05/2017).

Harle, J 2010 *Growing Knowledge: Access to Research in East and Southern African Universities*. London, The Association of Commonwealth Universities. https://www.acu.ac.uk/focus-areas/arcadia-growing-knowledge (accessed 02/05/2017).

Neylon, C, Willmers, M and **King, T** 2014 *Illustrating Impact. Applying Altmetrics to Southern African Research*. Cape Town: Scholarly Communication in Africa Programme. https://open.uct.ac.za/handle/11427/2316 (accessed 02/05/2017).

OECD 2007 *OECD Principles and Guidelines for Access to Research Data From Public Funding.* Paris: Organisation for Economic Co-Operation and Development. https://www.oecd.org/sti/sci-tech/38500813.pdf (accessed 02/05/2017).

Parker, M, Bull, S J, de Vries, J, Agbenyega, T, Doumbo, O K and **Kwiatkowski, D P** 2009 Ethical data-release in genome-wide association studies in developing countries. *PLoS Medicine*, 6(11): e1000143. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000143 (accessed 02/05/2017).

Piwowar, H and **Priem, J** 2013 The power of altmetrics on a CV. *Bulletin of the American Society for Information Science and Technology*, 39(4): 10–13. http://onlinelibrary.wiley.com/doi/10.1002/bult.2013.1720390405/pdf (accessed 02/05/2017).

Piwowar, H A and **Vision, T J** 2013 Data reuse and the open data citation advantage. *Peer J*, Published October 1, 2013. https://peerj.com/articles/175/ (accessed 02/05/2017).

Radjou, N and **Prabhu, J** 2015 *Frugal Innovation: How To Do Better With Less.* London: Profile Books Ltd.

Schwegmann, C 2012 *Open Data in Developing Countries.* European Public Sector Information Platform Topic Report No. 2013/02. Brussels: EPSI Platform. https://www.europeandataportal.eu/sites/default/files/2013_open_data_in_developing_countries.pdf (accessed 02/05/2017).

The Malaria Genomic Epidemiology Network 2008 A global network for investigating the genomic epidemiology of malaria, *Nature*, 456: 732–7. http://www.nature.com/nature/journal/v456/n7223/abs/nature07632.html (accessed 02/05/2017).

How to cite this article: Bezuidenhout, L 2017 Technology Transfer and True Transformation: Implications for Open Data. *Data Science Journal,* 16: 26, pp. 1–13, DOI: https://doi.org/10.5334/dsj-2017-026

Submitted: 16 January 2017 **Accepted:** 05 May 2017 **Published:** 22 May 2017

]u[*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS