

RESEARCH PAPER



Data sharing in low-resourced research environments

Brian Rappert^a and Louise Bezuidenhout^{a,b}

^aDepartment of Sociology, Philosophy, and Anthropology, University of Exeter, Exeter, UK; ^bSteve Biko Centre for Bioethics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

ABSTRACT

'Open data' has recently emerged as a label for renewed attempts to promote scientific exchange. As part of such efforts, the posting of data online is often portrayed as commonly beneficial: individual scientists accrue greater prominence while at the same time fostering communal knowledge. Yet, how scientists in non-Western research settings assess such calls for openness has been the subject of little empirical study. Based on extended fieldwork with biochemistry laboratories in sub-Saharan Africa, this paper examines a variety of reasons why scientists opt for closure over openness with regard to their own data. We argue that the heterogeneity of research environments calls into question many of the presumptions made as part of open data. Inequalities in research environments can mean that moves towards sharing create binds and dilemmas. These observations suggest that those promoting openness must critically examine current research governance and funding systems that continue to perpetuate disparities. The paper proposes an innovative approach to facilitating openness: coupling the sharing of data with enabling scientists to redress their day-to-day research environment demands. Such a starting basis provides an alternative but vital link between the aspirations for science aired today as part of international discussions and the daily challenges of undertaking research in low-resourced environments.

Introduction

Questions about who should share what data, and with whom, have long accompanied research. Today, 'open science' serves as an umbrella label for a diverse range of initiatives including the open access, open data and open software movements. Open science has generally been associated with calls to ensure research is available to the maximum extent possible (OECD, 2007). Policies pursued under this label seek to realize the norm of openness, so long associated with science, as well as to achieve greater egalitarianism in research. In this way, the open science movement seeks to maximize the benefits accrued from research for the benefit of humanity.

Nonetheless, as elaborated in the next section, despite the enthusiasm that often characterizes open science, realizing its ideals is challenging. A crucial element is the need for buy-in from practitioner communities. In the light of this, in this paper we ask: what concerns do scientists working in low-resourced research environments have about participating in open data activities?

The findings of the fieldwork draw attention to the critical role contextual factors in the research environment play in shaping opinions about openness and open data (Oyeleran-Oyeyinka, 2006; Olmos-Peñuela *et al.*, 2015). The paper details the observational and interview fieldwork with biochemistry laboratories in sub-Saharan Africa to illustrate how the contextual demands of sustaining research at all in low-resourced laboratories mean the logics for sharing in open data discussions have little traction with many of these scientists, given the perceived downside of sharing. The paper then offers a sense of the complicated intersections between how reward and credibility can be perceived within low-resourced settings *vis-à-vis* forms of data engagement. It makes a contribution to the tradition of critical scholarship regarding the potential for science policy to exacerbate global disparities (e.g. Lamberton, 2001). The paper concludes by considering what alternative support strategies could enable research and thereby data sharing. While the argument is geared towards redressing conditions frequently experienced in laboratories in low/middle-income countries (LMICs) and specifically sub-Saharan Africa, the considerations that delimit sharing and the recommendations offered are meant to be applicable to low-resourced laboratories whatever their location.

The promises and challenges of openness

Openness in data is commonly understood to refer to situations in which ‘anyone is free to use, reuse, and redistribute [the data] – subject only, at most, to the requirement to attribute and/or share-alike’ (Open Knowledge International, 2016). Recent initiatives to promote openness are justified with reference to varied normative, pragmatic and instrumental grounds, including: facilitating ‘self-correction’ through peer scrutiny; realizing the norms of science; making most effective use of public funds; ensuring commercial innovation; responding to demands by citizens for evidence in support of public policies; and enabling novel forms of science through utilizing new computational and communication technologies (see CODATA, 1997; Royal Society, 2012; Leonelli, 2013). Funding bodies, publishers, professional societies and others are bringing in policies justified through appeals to openness and accessibility (e.g. European Science Foundation, 2008; European Commission, 2011; RCUK, 2013).

While the general goals of open data are widely endorsed, what they should mean in more detail is often much less a matter of accord – especially in relation to data sharing. Scholars and policy-makers alike have acknowledged that moving from general principles to specific policies requires governments, funders, universities and others to address potentially thorny questions such as: what counts as ‘data’? Which elements of research need to be open? Who pays for that? How should openness be balanced against other priorities? In what ways does promoting the availability of research facilitate forms of commercial capture? (Wessels *et al.*, 2014).

When it comes to releasing datasets, no one size fits all. Thus, while imploring those involved in science to make their data accessible to those beyond their formal collaborators,

the majority of current data sharing statements leave the how, where and what to the determination of scientists and/or their organizations (International Council for Science, 2015). The non-mandatory nature of many current data sharing calls also raises questions as to the practicability and sustainability of current models (Mauthner and Parry, 2013). Increasing pressure to make data available has heightened the requirements for processing and curating these data too, though the forms of labour needed for such work are generally poorly recognized and rewarded within professional and organizational structures (Ankeny and Leonelli, 2015).

Reasons for data sharing

Against these sorts of widely recognized issues, how scientists can be encouraged to be more open has been identified as a matter of considerable urgency (Hayden, 2010; Leonelli *et al.*, 2013). Recent studies concentrating on Western Europe and North America and on why scientists share data highlight the role of perceptions of intellectual credit and peer recognition (Tenopir *et al.*, 2011; Borgman, 2012; Fecher *et al.*, 2015a). For instance, a survey distributed to scientists (mainly in high-income countries) conducted by Wiley, the publishing house, finds that 55% of respondents think that the increased impact and visibility of their work motivates them to share data (Ferguson, 2014). Similarly, a study by the Research Information Network (RIN) interviewed UK scientists regarding the benefits of sharing data (Research Information Network, 2009). Respondents highlighted the enhanced visibility of research, the increased efficiency arising from reusability and exposure, the identification of new research questions and directions, the fostering of scientific integrity and replication, as well as the enhancement of collaboration and community-building as key reasons to participate in data sharing activities.

Reasons not to share are also identified (Fecher *et al.*, 2015a). The sample of concerns collected by Ferguson (2014) include intellectual property or confidentiality issues (42%), fears of being scooped (26%) and not getting proper credit (22%). A 2010 report by RIN and the National Endowment for Science, Technology and the Arts also highlights perceived lack of evidence of benefits, lack of time and skills, cultures of independence, and concerns about quality and ownership (RIN/NESTA, 2010).

Despite such concerns, the release of data online is recognized by many of the surveyed scientists in North America and Europe as having professional benefits in addition to philanthropic good. Critically, the benefits can justify the additional time and effort spent on curation and dissemination activities. For some, data sharing activities are portrayed as enhancing the ability to win grants, securing greater recognition from peers and advancing careers (Research Information Network 2009, p.2).

And yet, while recognition, credit and reputation are acknowledged as topics that need to be addressed, policy and academic literature about open data typically lack a wider theoretical framework for understanding how reward, recognition and reputation are linked together in the (re-)production of factual claims and professional careers. Fecher *et al.* (2015b) offer some initial pointers along these lines in their proposal to situate data sharing in the 'reputational economy' of science.

Much of the literature in Science and Technology Studies (STS) recommends a broad backdrop for understanding sharing *vis-à-vis* recognition. For instance, through their notion of 'credibility cycles', Bruno Latour and Steve Woolgar (1979) contend that the central

preoccupation of the scientists they observed was building up ‘credibility’, defined generically as the underpinning ‘abilities [to] actually [...] do science’ (Latour and Woolgar, 1979, p.198). Latour and Woolgar’s formulation of credibility cycles emphasize the distinction between *reward* and *credibility*. Reward (including forms of award and reputational credit often mentioned in open data today) refers to recognition for achievements, while credibility pertains to the ongoing capacity required for doing research.

Understanding individual participation in data sharing through calculated decisions – be they based on net benefits, reputational gains or credibility accrue – has important implications for the presumed potential of open data. First, it assumes that those who want to participate in data sharing can. Second, it assumes that the terms of calculative logics of scientists are similar regardless of nationality, physical location and cultural background. The evidence given below offers a critical unpacking of these assumptions. Based on the fieldwork undertaken in support in this paper, we return to a consideration of frameworks for understanding scientists’ data practices.

A limit to openness: low-resourced research environments

Perhaps unsurprisingly, the vast majority of studies of data sharing to date have examined Western laboratories (Carr and Littler, 2015, p.315) – and within these well-resourced labs. Discussions about both the motivations to share data and the ways in which data are made open are tied to specific understandings of resource distribution, infrastructure provision and governmental involvement (as in OECD (2015)).

Open data discussions do sometimes recognize that scientists in some countries are not well placed to make their data open because of shortfalls in research resources and infrastructures (e.g. Royal Society, 2012). In response, calls have been made to enhance physical hardware infrastructure, soft behavioural infrastructure and skill-based capacities in these countries in order to ensure that scientists are better able to participate in the universal call for openness (CODATA, 2014; International Council for Science, 2015). While such initiatives are important, it is open to question how much they can ameliorate the differences in research environments between high-income countries (HICs) and LMICs. Most obvious are disparities in Internet connectivity. Less visible, but not necessarily less severe, are problems relating to the wider research environments for undertaking science. How these issues impact on scientists’ understanding of, and interaction with, open data is largely absent from current policy and academic discussions.

Further complicating this picture is that the existing literature on data sharing in developing countries focuses on comparatively well-funded and well-connected research networks or consortia, dealing predominantly with clinical research (de Vries *et al.*, 2011; Parker and Bull, 2015; de Vries *et al.*, 2015). While these studies raise important concerns, these are often very specific to clinical research with vulnerable patient populations. Indeed, previous studies note a strong divergence in the experimental practices, goals and values between biologists and clinicians (Kelly and Geissler, 2012; Leonelli, 2012). Although this situation is improving with regard to researchers who donate their own data rather than data from others (Bull *et al.*, 2015a), the vast majority of studies on data sharing in LMICs still focus on clinical trials or public health research, with minimal attention given to other fields (e.g. Pisani and Abou-Zahr, 2010).

Research questions and design

In recognition of the need for evidence on data sharing practices in low-resourced research environments, we undertook a study that sought to address two questions:

- Do low-resourced research settings influence scientists' perceptions of the value of data?
- Do the conditions in low-resourced laboratories influence scientists' perceptions of the potential gains and risks from data sharing activities?

We selected field sites through a series of strategic decisions. First, it was decided that all of them would be in Africa, as this continent is largely missing from discussions about open science and open data in particular. Second, two countries were selected – one in southern (South Africa) and one in eastern Africa (Kenya) – with robust national research programmes. As country background, Kenya has 22 public universities, many of which conduct research. It also has a long history of international research collaboration, a prime example being the long-standing KEMRI–Wellcome Trust partnership. While the government encourages research, financial support for it remains limited and the focus of national universities is on undergraduate teaching. South Africa has 25 public universities, all of which conduct research. South Africa has a long history of academic research which is actively supported by the government.

Third, we sought examples of vibrant, 'homegrown' research. While some of the researchers at the sites visited collaborated with others in Europe and North America, by design none of the field sites was formally affiliated to large internationally-funded research consortia or networks. Fourth, within these two countries, four departments/groups in academic institutions were selected for inclusion based on their common disciplinary focus (the interaction of chemistry and biochemistry) and research interests (medicinal chemistry). These decisions were to minimize the differences in data sharing practices and perceptions between scientific disciplines noted in previous open science discussions (e.g. Royal Society, 2012; Wessels *et al.*, 2014) and considered in information studies more broadly (Macdonald, 1998).

Within Kenya, site 1 (KY1) and site 2 (KY2) were both chemistry departments of well-established universities. Both had over 15 full-time faculty members. However, student to faculty ratios were high and teaching loads considerable. KY1 had a large number of M.Sc. and Ph.D. candidates, the majority of whom were full-time and a number of whom had financial assistance. In contrast, KY2 had a very high number of M.Sc. students, the majority of whom were self-funded and part-time (and thus conducted their laboratory work during holidays). In both departments, space in laboratories was at a premium and students shared working space and equipment. Neither department had any postdoctoral researchers.

Within South Africa, site 1 (SA1) was a research group within the large chemistry department of a well-established and comparatively well-resourced university with a tradition of research. Site 2 (SA2) was the chemistry/biochemistry department of a university that had previously been designated as being for marginalized population groups under the apartheid system. Both sites were the recipients of numerous national and international grants. SA2 had one postdoctoral researcher at the time, while SA1 had none.

Empirical data were gathered using a combination of qualitative methods including embedded laboratory observations and semi-structured interviews. Each site visit took between three and six weeks, during which time one of the authors (LB) participated in

departmental activities, interviewed faculty and graduate students, and observed social and physical working environments in the departments and laboratories. Data collection was undertaken over a period of five months between November 2014 and March 2015, with 56 semi-structured interviews in total conducted with faculty and graduate students. Follow-on visits to each site were made in late 2015 by both authors to solicit feedback on our analysis.

Commonalities between sites

While the four sites visited varied in terms of age, financial provision and size, they nonetheless shared certain commonalities. This sub-section briefly highlights some of these similarities. In the next section, we turn to distinctions.

Division of labour. All the departments relied heavily on Masters and Ph.D. students for data generation, and the vast majority of the research conducted in all the laboratories was done as part of graduate degrees. The absence of postdocs, lab managers and dedicated research staff – in combination with high teaching loads for faculty – meant that these students assumed responsibility for undertaking daily research procedures, data analysis, and (peer and undergraduate) laboratory training. Driving research activities according to thesis requirements had additional implications for the ability of principal investigators to create long-term research agendas, to find funding for research projects, and to bring together and synthesize data produced from different students.

Precarious funding. Common to all sites was the problem of acquiring core funding for facility maintenance and improvement. This was not only because of low governmental contributions (particularly in Kenya), but also because most grant awards did not make provisions for facility maintenance or upgrading. Thus, the purchase of general equipment, ICT hardware and software was regularly reported to be problematic.

Systemic issues. Participants at all four sites mentioned challenges in their daily research activities that related to broader infrastructural issues. These included regular power cuts and varying provision of backup generators, complicated and time-consuming border controls and reagent delivery, problematic or absent effective sample transport options, difficulties with adequate technical support, and issues with equipment maintenance.

Promotion systems. At all four sites, the promotion of faculty was directly linked to publication outputs in the form of journal papers. Other forms of sharing or public engagement were not recognized explicitly in promotion criteria. In addition, in the South African sites the publication of journal papers was directly linked to the acquisition of funds through the government's Research Incentive (RINC) scheme. This works by funding universities for each peer-reviewed paper, conference proceeding or book published, thus providing strong incentives for favouring the number of officially recognized publications over other research outputs.

ICT provisions. While all the sites had access to the Internet and at least some computing and library facilities, all interviewees agreed that challenges existed when accessing online resources. Power cuts, low bandwidth and variable wi-fi signal were regularly identified as daily challenges to working online. Complications were noted in off-site access to university resources, as none of the sites (except SA1) had functioning proxy servers. In addition, many participants noted problems associated with the acquisition of software and hardware. Importantly, many were working with older hardware and software as they were required to make ICT purchases from personal, rather than research, funds.

Open access. Interviewees often equated open access (OA) publishing with pay-to-publish journals, and these were consequently viewed as inferior to other journals. None of the Kenyan interviewees was aware of international financial assistance schemes for publishing in OA journals. Some researchers reported using their own money to make their publications OA.

Professional self-promotion. It was salient that none of the interviewees were particularly interested or engaged with professional profiling sites. Although membership of professional networking sites (e.g. LinkedIn) was often reported, the majority of interviewees saw little value in membership and did not actively contribute. None of the interviewees had considered using social media (e.g. Twitter) or professional monitoring tools (e.g. altmetrics) to promote their research. Similarly, personal websites and extensive university webpages were also absent. At the same time, the feeling of professional isolation stemming from geography and peripheral community position was a recurring theme.

Research in low-resourced environments: some site vignettes

The interviews highlighted something that, while perhaps unsurprising, has been little explored in current international discussion about openness: namely that physical and organizational aspects of the research environment significantly influence scientists' involvement in practices aligned with open data and particularly how they think about sharing data. What became evident was that the day-to-day challenges of conducting research in these low-resourced environments – and thus the (reduced) speed at which research progressed – affected scientists' perceptions of data sharing, their fears of being scooped or exploited, and their understandings of the rewards of releasing data online. In the following sections, vignettes from each field site illustrate these concerns.

Funding and research: balancing openness with gain in KY1

At the KY1 site, faculty faced teaching demands that seriously curtailed the time that could be spent conducting research. The faculty interviewed regularly made reference to the high numbers of students in their classes:

Between September last year and August this year, I've had to teach 830 students ... They had a double intake and we weren't told about it last year. Just in the middle of teaching we were told there is a new group that is coming. (KY1/1)

Parallel teaching programmes increased the number of undergraduate students dramatically, thus increasing the workloads of individual researchers. Similarly, the lack of teaching assistants, practical supervisors and marking assistance made teaching duties more time-consuming.

Discussions with faculty members confirmed that purely research appointments were very unusual, and that teaching was the primary duty expected of faculty members. Nonetheless, all promotions were directly based on publications and qualifications. Faculty members, needing publications for promotion, thus faced a difficult situation because they often had little time or support for research: 'It's not an environment that values research and development. It's not a ... I mean, it's kind of a lonely thing. You have to do it out of your own push' (KY1/8).

This situation was compounded by issues relating to funding. Funding for research at KY1 was secured outside Kenya; government support covered only basic core infrastructure costs (such as salaries, electricity and building maintenance) and not laboratory running costs. While some faculty members had secured funding as part of international projects (e.g. from the US National Institute of Health), all reported that finding the time to apply for funding and establish collaborations was taxing. Most research was conducted as part of graduate degrees, commonly small projects with clearly defined objectives (such as isolating compounds from a specific plant with known medicinal properties). These projects were united by methodological similarities and curtailed by resource limitations (such as the need to send compounds out of the country for activity testing, which entailed costs as well as delays).

In order to circumvent the problems associated with building up a body of research necessary for promotion, several faculty explicitly mentioned that they used their own personal money to cover research costs:

For most things I have used a lot of my own money on the research because I don't think you get government [help]. Though occasionally there are some committees called Dean's committees where you can apply for funds, but the funds are limited and they are mostly limited to students who are doing research, PhD students. But for staff, what do you do? (KY1/1)

The use of personal money for research, combined with the high stakes associated with publishing research, meant academics strategically limited the accessibility of their data to others: 'Here you often find that people pay for their own research. They wait to patent their findings, but this means they can sit – for 5 or 10 years – on the data without a patent or a publication' (KY1/9). Concerns about losing control of the data – and thus the benefits of the data – were regularly verbalized as fear of being 'scooped'. This led to reluctance to share data prior to publication: 'I know people tend to handle the data in a way that they do not share before publishing' (KY1/2). This corresponded with wider preferences around revealing visibility online, with a marked preference against too much online openness. While concerns about being scooped through data sharing are not new to discussions on open data (see Ferguson, 2014), the repeated and explicit link of these concerns to the research environment was notably pronounced in KY1. Scientists at the site felt that their working conditions made being scooped the likely outcome of any data sharing.

The pressures of conducting research and publishing it, together with the lack of support for these activities, created an environment in which data were often perceived as personal rather than collective property. Data were seen as a means to an end for the researchers who invested in the data production in the first place. Linking publishing to remuneration from promotion had a significant impact on how scientists viewed their responsibilities to disseminate data, particularly as it took them a lot of effort, time (and, again, often personal money) to generate the data. Such observations were reinforced by a number of statements by senior faculty regarding the (lack of) incentives to do research at their career stage.

Time and space: part-time students ensuring data release in KY2

As (similarly to KY1) the faculty at KY2 relied primarily on graduate students to generate research data, this created a difficult situation in which data were appearing only sporadically and research took a long time to complete. As one faculty member put it: '[i]t's difficult to build up a body of data when the research is all short-term and ends with a student' (KY2/13).

At the KY2 site a part-time Masters programme had recently been introduced. This increased the number of graduates registered at the department and thereby the student to faculty ratio. Most of the part-time graduates were high school teachers and came during school holidays to complete their laboratory research. As a result, the vast majority of graduates at KY2 were enrolled on a part-time basis taking three–four years to complete. Both faculty and students that were interviewed found this a frustrating situation. As one part-time Masters student said:

it is difficult to work like this because you must come for a short period of time, take a little data and then go away. When we are away it is difficult to do work, and also to get hold of our supervisors. (KY2/12)

The duration required for graduate students to complete also frustrated the publication of their research and, as well, publication was not a criterion for passing a Masters. Moreover, without a wider programme of curation, synthesis and re-analysis it was likely that a large amount of the data produced by students – even their theses – would not be effectively used.

Compounding these considerations, research was also hampered by the state of labs which contained little equipment, old benches, few reagents and so forth. The availability of equipment at KY2 in particular was identified as restricting the amount and the kind of experimentation that could be performed (for instance, chemical synthesis was not possible). As in the case of a donated *nuclear magnetic resonance* (NMR) spectroscope, even if the department possessed equipment, it did not have funds for spare parts or a serving technician. One faculty member addressed this explicitly, saying: ‘the lack of equipment limits the extent to which you can do research – and even the type of research that you want to do’ (KY2/3).

The lack of equipment in turn played an important role in some scientists’ self-perceptions of their position within their field *vis-à-vis* their geographic position. As one faculty member stated: ‘[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’ (KY2/15). Similarly, another said:

[h]ere I will publish and unfortunately when I do that here even that person in research and development in industry will not read my paper unless someone has said something. So the tragedy is I do all the work, I publish it, the audience that I am looking for already has a prejudiced view, if I may put it so, about my ability to do the research, so they won’t value it or read it, right, unless they know me – unless they know me as a person ... They don’t even know I exist. So it’s quite a depressing situation. (KY2/4)

The interviewees at KY2 thus highlighted the difficulty of gathering, curating and disseminating data accrued from numerous student projects. They also highlighted the additional difficulties associated with limited equipment and the limitations they put on the type of research possible at their institute. Together, these difficulties contributed towards an environment with low levels of data sharing.

Geographies of research: managing historical and geographic legacies in SA2

The institution in which SA2 was located is in a geographically isolated part of the country, three hours away from the next major city. It had been founded as a university for disenfranchised populations during the apartheid regime and continued to struggle to overcome its legacy as a disadvantaged university. As one lecturer said: ‘The traditionally advantaged

institutions in South Africa are still advantaged. The 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).

A number of interviewees made reference to the difficulties of getting reagents, equipment and technical support, which slowed down the pace of research in the department. One faculty member offered a story that eloquently described these challenges:

... in terms of technical support in the lab, it's not there. At times we have a plug that isn't getting electricity and to have someone to come out and get it fixed may take a week so perhaps you need to move the freezer and send it to another building, find a plug and hook it up for a while. So, sometimes the plug just goes dead for a day and there's nobody on site who will come and find out what the issue is. To make a report may take a few days. In the interim, you need to find a solution to save the biologicals from breaking down and deteriorating. So those are challenges. I find I go to Johannesburg to get stuff, but it speaks to the culture because when you understand what it takes to run research as a program and to put the bits and pieces together such that if you're in the lab you stay in the lab because you're not worried whether things will be supplied or not. (SA2/10)

Geographic challenges also meant that acquiring and maintaining the equipment necessary for research was a continual challenge. A number of faculty members made reference to the NMR spectroscope that is extremely important for the chemistry research undertaken. This machine 'was sponsored by the government. But now it's not working and we are squeezed. You must remember that we are far from the city here and sometimes there are challenges with filling of the liquid helium and the liquid nitrogen' (SA2/2). Similarly, in order to get the liquid nitrogen

we installed liquid nitrogen plant and that has been quite challenging also but it has been working, then when it broke down we had to depend first on [another university] then they were not active anymore we had to send the person on a weekly basis to ... fetch liquid nitrogen [from a city 12 hours round trip away]. (SA2/6)

Despite the considerable advances made by the university in the post-apartheid years, the challenges of the historic legacy were still evident within the university structures. One of the key concerns was the lack of core funding from the institution to improve research facilities. Such issues were a challenge for researchers: '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). Many interviewees also complained that lack of basic infrastructure reduced the speed and efficiency of activities. These were problems not easily addressed by individuals or departments: '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' (SA2/3).

Interestingly, there was also awareness that these systemic issues would not necessarily be resolved simply by more research funding:

Our challenges are unique and very different and so you should come with a purse of money and hand it out. You may address some of the issues but you would not address all. In fact, I don't think you could address even 50% of them. So again, you 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 money. (SA2/1)

Problems with systems of procurement, budget allocation and maintenance were all commonly cited as responsible for slowing down research processes. Interviewees regularly linked these daily pressures into their answers about data sharing. They took pains to

enumerate how much effort it took to do high-quality research in such a geographically isolated environment, and how these daily pressures took time away from possible data sharing activities.

It's a matter of speed: competing internationally in SA1

SA1 was the best supported and resourced of the departments examined. Funded by large national and international grants, this department was producing the most internationally competitive research. Nonetheless, when discussing their science and the dissemination of data, worries about being scooped by better resourced foreign laboratories were routinely voiced. Such fears turned on two important issues. First, despite the resources available at SA1, researchers were aware that the additional time necessary to acquire reagents ('You're twiddling your thumbs trying to find out what you're supposed to do for a couple of months while you're waiting for chemicals' (SA1/7)) and their reliance on graduate students for data meant that it just took longer to do research. Thus, while interview participants were, in principle, eager to share, they were often hesitant to disseminate data because 'It would just mean that half of the stuff that we do would be taken over by people with much more resources and they'd do it much quicker than us' (SA1/7). In practice, data sharing was limited largely to established personal or formal connections – colleagues, long-standing contacts and international collaborators.

Second, there was continual reference to the South African government's structures for research funding. Researchers were under considerable pressure to publish in RINC (research publication incentive)-approved journals. While separate from project funding, the money-for-publishing funding structure was an important part of research life, as it was the primary way in which individuals and departments accrued funding for departmental activities and conference travel. Thus, as one participant put it:

I think [open data is] great, but in terms of funding and output-recognized grant proposals and things like that, I don't think it's working right here. So we're judged by how much output we get out there and first-time publications and first authors and things like that. (SA1/7)

Speed and data sharing

While participants agreed on the importance of openness and sharing data for science as a whole, very little data sharing beyond project collaborators was in evidence in any of the sites visited. Interviewees repeatedly spoke of the professional perils that might be encountered through the general release of data. It would seem, in contrast to more optimistic discussions on open data, that simply telling scientists about the possible benefits of data sharing was nowhere near sufficient to convince them to engage in data sharing.

In all four cases, it is evident that reluctance to share data was in some way related to the relatively long time between planning and publishing experiments caused by day-to-day factors hampering research. As a result of pervasive concerns about being scooped in priority contests, many interviewees said that the extent and nature of sharing had to be limited and expressed reservations about data sharing initiatives. In short, those we spoke with did not regard themselves as having the luxury to partake in the gift economy (Zeitlyn, 2003) or in gift exchange (Hagstrom, 1965) relations with others. Such anxieties echo those

made elsewhere. Both Pisani *et al.* (2010) and Bull *et al.* (2015a) discuss apprehension about sharing clinical and public health data generated in LMICs.

Instead of discussing the benefits from sharing, nearly all participants made such comments as:

To me it's too much of a risk, it's too much of a risk and I'm not at that stage to take such risks, I don't think so. ... We want to be trusting each other but sometimes it's not that easy. (SA2/7)

In particular, no interviewees explicitly endorsed the sharing of unpublished data. Failure to engage in the sharing avenues that increased altmetric exposure created a vicious circle. Many participants explicitly said they did not perceive overall gain from the professional networking sites they used (primarily LinkedIn). This created a negative perception that was commonly extended to all altmetric-enhancing avenues. Unsurprisingly, the interviewees did not identify the kinds of advantages to sharing – increased citations and visibility for research, and kudos within the relevant communities (Ferguson, 2014) – that are reported in Western countries as offsetting fear about data exploitation.

Some of those interviewed reported that if they undertook research on topics most prominent within their specialty fields, they risked losing out to those able to produce publications in shorter time frames. As a result, many contended that openness with data was particularly dangerous for them, given their constrained ability to undertake research. Even garnering international visibility for one's work was a potentially *pyrrhic* accomplishment as it might attract competitors. However, if scientists undertook research on topics more peripheral to the concerns of their fields, then they risked perpetuating their (typically) marginal positions within their field. This predicament was compounded by limitations in research capacity overall. Interviewees had few lines of research they could sustain at any given time because of other demands on their time and the taxing requirements of doing research. As a result, decisions about which agendas to pursue involved high stakes. The most common decision was to play safe by setting low, but achievable, aspirations.

At all four sites, there was widespread perception that relatively well-funded scientists just did not understand the challenges those interviewed faced. One participant at SA2 spoke to this when discussing attempts to engage with web pages:

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 – is for conference registration. Sometimes trying to explain to the conference organiser that I am not able to put my data online because for some reason my system stops. It works for two entries and on the third entry it stops. So, I have tried it several times, I've tried it from other computers on campus, and now I give up. Please help me! Trying to convince people that we are really having a problem of this type is difficult. They cannot imagine what the problem is – they've never experienced it. ... That is because those sites are heavy. They have lots of fancy things that are very beautiful, but then also the templates you have to fit in things and for some reason if the system is weak. The template will not respond. (SA2/12)

As one participant put it, doing research in Africa is 'tough, but tell that to someone who's in America and they would say "What are you talking about?"' (SA2/7).

Credibility and reward

Qualms with data sharing extended far beyond concerns associated with reputation itself; rather, they were rooted in scientists' basic ability to undertake science. This attention

to scientists' ongoing ability to do research is consistent with the analysis of credibility mentioned earlier. As Latour and Woolgar (1979) discover, as part of 'cycles of credibility', different aspects of research – data, outputs, funds, reputation, arguments, credentials, prestige, etc. – are converted into one another.

Scientists' behaviour is remarkably similar to that of an investor of capital. An accumulation of credibility is prerequisite to investment. The greater this stockpile, the more able the investor to reap substantial returns and thus add further to his growing capital. ... The essential feature of this cycle is the gain of credibility which enables reinvestment and the further gain of credibility. Consequently, there is no ultimate objective to scientific investment other than the continual redeployment of accumulated resources. It is in this sense that we liken scientists' credibility to a cycle of capital investment. (Latour and Woolgar, 1979, pp.197–98)

The notion of credibility cycles applies to various aspects of science, such as the status of facts, material resources and institutional reputation. Within credibility cycles, the factual status of claims made by scientists and their reputational standing are tied to one another. When deemed credible, researchers can use their associations and credentials to access forms of material infrastructure that can enable the refinement of skills that then lead to the production of data and papers. This provides the means for enhancing prestige that is then converted into high profile presentations that lead to invitations to collaboration, that A prime aim of scientists was 'to extend and *speed up the credibility cycle as a whole*' (Latour and Woolgar, 1979, p.207 (emphasis in original)).

In line with previous research making use of credibility cycles (Packer and Webster, 1996; Hessels and van Lente, 2011), our fieldwork highlighted the heterogeneous range of skills required for daily research activities; for example, how to organize research programmes without dependable power or without reliable access to basic inputs. These frustrated scientists' ability to do research, but also their ability to secure non-material forms of symbolic capital that could be converted into other forms of capital as part of credibility cycles.

Most studies using the notion of credibility cycles are concerned with conditions in high-income Western settings (e.g. Rip, 1994; Leisyte *et al.*, 2008; Hessels *et al.*, 2009). Consequently, certain assumptions have gone unchallenged. For instance, such studies have adopted Latour and Woolgar's basic starting point, giving primacy to credibility rather than reward. In contrast, pressed with the difficulties of conducting and sustaining research in the sites studied in Kenya and South Africa, scientists regularly adopted strategies that made reward the goal. These strategies, in turn, affected what data were generated and whether the data were shared.

To expand, one of the ways in which this took place was by the termination of active research careers. At all four sites, given the demands of doing research, some faculty members sought to increase their professional status through bureaucratic advancement rather than through experimental achievement. By being head of department or participating in internal university governance, they were able to gain a degree of visibility that could be translated into membership of national and international funding and policy committees without the burden of maintaining an active research career. Similarly, as a number of high-level faculty members at the Kenyan sites made clear, participation in research fell off with the promotion of individuals to professor: 'They have been promoted as far as they can go and there are no financial rewards' (KY1/8).

The pursuit of credibility also gave way because of the manner in which faculty often paid for their research and patent applications. At times, this self-funding supported the

accumulation of forms of scientific capital. At other times, though, the monetary rewards from doing science (e.g. personal salary) were used to obtain other forms of reward (e.g. royalties from licensing, or a higher salary from promotion directly linked to publication that might then lead to the cessation of research careers). Complex judgements were made by some about whether to seek credibility or reward in inter-twined economic and science capital cycles operating at the level of universities and individuals. In such ways, long-term credibility strategies often took second place behind the pragmatic needs of immediate and near-term career and organizational demands associated with doing research in resource-poor environments. Rather than ‘status, rank, award, past accreditation, and social situation [being] merely resources utilised in the struggle for credible information and increased credibility’ (Latour and Woolgar, 1979, p.213), they were often the ends of science – even for faculty still notionally designated as research active.

What can be done?

Taken together, the previous sections indicate how the exchange of data can be stifled and the prospects for promoting greater exchange diminished. Concerns about the asymmetrical conditions of research suggest something of a vicious cycle of practice and action. Rather than engaging in informal peer exchange, those interviewed repeatedly approached data sharing through market transaction and reward-centred forms of reasoning; with corresponding uncertainties in deciding how to assess their data (see Macdonald, 1998).

What, then, can be done to promote data sharing by those organizations wishing to support science in resource-challenged environments? Recognition of global inequalities in clinical and public health research has led to proposals to enhance data sharing by those in LMICs. Policies advocated to bolster sharing include improving professional recognition through standards for citations (see Pisani and Abou-Zahr, 2010), fostering international trust-based collaborations (Tangcharoensathien *et al.*, 2010; Bull *et al.*, 2015b) and improving data management capacities through greater resourcing (Alter and Vardigan, 2015; Bull *et al.*, 2015a). From the perspective of everyday research environments, our analysis suggests an alternative approach. The mundane, day-to-day demands of conducting research in resource-constrained environments mean scientists are not inclined toward data sharing, and may not be inclined to continuing to conduct research at all. Addressing the sources of what undermines scientists’ ability and willingness to do science could be one way of fostering conditions more conducive to data sharing.

Open data initiatives could benefit from re-orientation. Attempts to impose sharing requirements or extol the importance of data openness are in conflict with the personal experience of those in low-resourced settings. Such efforts are likely to meet with suspicion, estrangement or token participation. An alternative approach would be to hold that greater openness can shed light on the circumstances that make performing research so difficult.

While absolute funding levels are no doubt important, this is not the only aspect of funding that needs to be addressed. In the labs visited, resources were sometimes available, but they could not be used. The absence of even the small sums – usually below US\$100 – required for professional membership, equipment servicing, off-campus access to the Internet and papers, IT hardware and software, and so forth played a crucial role in how well those interviewed could do their science, as well as what they thought about data sharing (see Bezuidenhout *et al.*, 2016). Not being able to fix an NMR machine, obtain liquid nitrogen,

rewire the plugs in the lab, buy software or update, or get buy-out from teaching to write a research proposal limited the production of data and, consequentially, the sharing of data.

What is further evident from the cases is not only this link, but also the difficulties scientists experience improving their situation. The lack of core infrastructure funding for the laboratories, the difficulty of securing (often foreign) project grants, and strict limits on what such project funding can be used for, all contribute towards lack of capability and agency. That many of these problems are so mundane only contributes to the frustration they cause. It follows that one way of engaging scientists in data engagement activities is enabling them to shape their research in a manner compatible with their day-to-day demands. More specifically, our analysis suggests that very small sums, easy to apply for, but flexible in how they can be spent, would go some way toward meeting the challenges of doing science. Such support would enable researchers to cope with the obstacles associated with converting scientific capital in their credibility cycles. It would also offer an alternative compact for integrating those in low-resourced settings into the open data agendas of high-income settings. Enabling research, rather than compelling adherence to rules or norms, could be the basis for promoting data sharing.

In this basic recommendation, we take inspiration from efforts in recent decades to promote forms of micro-credit for those without access to conventional financial services (Barry, 2012; Mahmuda *et al.*, 2014). In the case of research, rather than monetary return, micro-funding could be tied to forms of repayment associated with greater research efficiency in producing outputs or, specific to the concerns of this paper, requirements to share data. To extend development literature parallels, enabling scientists in resource-constrained conditions to shape their research environments could have a wide range of benefits. For instance, it could enable them to define what they need instead of this being prescribed elsewhere, thus promoting agency and self-confidence. It could also encourage engagement with the international research community, thus decreasing the sense of isolation of these scientists.

The applicability of these recommendations needs further consideration. For instance, who exactly should be able to access such micro-funding is a critical issue. Our experience suggests there are scientists in low-resource settings in South Africa and Kenya who are managing (though struggling) to undertake high quality research despite all of the additional demands they experience compared with their colleagues in well-resourced labs. Additional flexible support for such individuals could go some way toward promoting their research. This proposal might also apply to those in HICs who, despite the overall conditions in their countries, are hindered in their ability to do science and thus to engage in data sharing.

Concluding comments

Much of the policy promotion of open data and data sharing today starts not only with belief in the importance of openness, but also with an assumption that all scientists will benefit from releasing data, no matter where they are based. The highly variable nature of laboratory environments around the world is often overlooked. So, too, is the key role they play in mediating the behaviours of scientists working in them. Our study suggests the importance attributed by individual scientists to concerns about credit and recognition should be understood as the product of a dynamic and continuous interplay within the research environment.

Such points are of critical importance to open data initiatives. Diverse science communities are not well served by generalized expectations designed with a limited range of environments in mind. In response to the day-to-day conditions that frustrate both science and data sharing, we have suggested that flexible support, designed to address the factors scientists identify as hindering their research, could go some way toward enabling it. As this is achieved, it will be possible to think about how to promote sharing and openness.

Acknowledgements

The authors would like to thank Sabina Leonelli and Ann Kelly for comments and corrections, and the paper's reviewers.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research for this paper was supported by the Leverhulme Trust under the award entitled 'Beyond the digital divide' [grant number RPG-2013-153].

References

- Alter, G. and Vardigan, M. (2015) 'Addressing global data sharing challenges', *Journal of Empirical Research on Human Research Ethics*, 10, 3, pp.317–23.
- Ankeny, R. and Leonelli, S. (2015) 'Valuing data in postgenomic biology: how data donation and curation practices challenge the scientific publication system' in Richardson, S. and Stevens, H. (eds) *Postgenomics*, Duke University Press, Durham NC, pp.126–49.
- Barry, J. (2012) 'Microfinance, the market and political development in the internet age', *Third World Quarterly*, 33, 1, pp.125–41.
- Bezuidenhout, L., Leonelli, S., Kelly, A. and Rappert, B. (2016) "'\$100 is not much to you': open access and neglected accessibilities for data-driven science in Africa", *Critical Public Health*, 27, 1, pp.39–49.
- Borgman, C. (2012) 'The conundrum of sharing research data', *Journal of the American Society for Information Science and Technology*, 63, 6, pp.1059–78.
- Bull, S., Cheah, P., Denny, S., Jao, I., Marsh, V., Merson, L., Shah More, N., Nhan, L., Osrin, D., Tangseefa, D., Wassenaar, D. and Parker, M. (2015a) 'Best practices for ethical sharing of individual-level health research data from low- and middle-income settings', *Journal of Empirical Research on Human Research Ethics*, 10, 3, pp.302–13.
- Bull, S., Roberts, N. and Parker, M. (2015b) 'Views of ethical best practices in sharing individual-level data from medical and public health research: a systematic scoping review', *Journal of Empirical Research on Human Research Ethics*, 10, 3, pp.225–38.
- Carr, D. and Littler, K. (2015) 'Sharing research data to improve public health: a funder perspective', *Journal of Empirical Research on Human Research Ethics*, 10, 3, pp.314–16.
- CODATA (1997) *Bits of Power: Issues in Global Access to Scientific Data*, National Academies Press, Washington DC.
- CODATA (2014) *Data Sharing Principles in Developing Countries (The Nairobi Data Sharing Principles)*, report from CODATA Workshop on Open Data for Science and Sustainability in Developing Countries, August, Nairobi, available from <https://www.rd-alliance.org/sites/default/files/attachment/NairobiDataSharingPrinciples.pdf> [accessed January 2017].
- de Vries, J., Bull, S., Doumbo, O., Ibrahim, M., Mercereau-Puijalon, O., Kwiatkowski, D. and Parker, M. (2011) 'Ethical issues in human genomics research in developing countries', *BMC Medical Ethics*, 12, 1, p.5.

- de Vries, J., Tindana, P., Littler, K., Ramsay, M., Rotimi, C., Abayomi, A., Mulder, N. and Mayosi, B. (2015) 'The H3Africa policy framework: negotiating fairness in genomics', *Trends in Genetics*, 31, 3, pp.117–19.
- European Commission (2011) *Open Data: An Engine for Innovation, Growth and Transparent Governance*, COM (2011) 882 final, Brussels, available from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF> [accessed January 2017].
- European Science Foundation (2008) *Sharing Responsibilities in Sharing Research Data: Policies and Partnerships*, report of an ESF–DFG workshop, September 2007, Padua, Italy, available from http://archives.esf.org/fileadmin/Public_documents/Publications/SharingData_01.pdf [accessed January 2017].
- Fecher, B., Friesike, S. and Hebing, M. (2015a) 'What drives academic data sharing?', *PLoS ONE*, 10, e0118053, available from <http://journals.plos.org/plosone/paper?id=10.1371/journal.pone.0118053> [accessed January 2017].
- Fecher, B., Friesike, S., Hebing, M., Linek, S. and Sauermann, A. (2015b) *A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing*, discussion paper, February, German Institute for Economic Research, Berlin, available from <http://www.hiig.de/wp-content/uploads/2015/02/dp1454.pdf> [accessed January 2017].
- Ferguson, L. (2014) 'How and why researchers share data (and why they don't)', *Wiley Exchanges*, available from <http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/> [accessed January 2016].
- Hagstrom, W. (1965) *The Scientific Community*, Basic Books, New York.
- Hayden, C. (2010) 'The proper copy: insides and outsides of domains made public', *Journal of Cultural Economy*, 3, 1, pp.85–102.
- Hessels, L. and van Lente, H. (2011) 'Practical applications as a source of credibility', *Minerva*, 49, 2, pp.215–40.
- Hessels, L., van Lente, H. and Smits, R. (2009) 'In search of relevance', *Science and Public Policy*, 36, 5, pp.387–401.
- International Council for Science (2015) *Open Data in a Big Data World*, ICSU, Paris.
- Kelly, A. and Geissler, P. (2012) *The Value of Transnational Research: Labour, Participation and Care*, Routledge, London.
- Lamberton, D. (2001) 'An information infrastructure for development', *Prometheus*, 19, 3, pp.223–30.
- Latour, B. and Woolgar, S. (1979) *Laboratory Life*, Princeton University Press, Princeton NJ.
- Leisyte, L., Enders, J. and de Boer, H. (2008) 'The freedom to set research agendas – illusion and reality of the research units in the Dutch universities', *Higher Education Policy*, 21, 3, pp.377–91.
- Leonelli, S. (2012) 'When humans are the exception: cross-species databases at the interface of biological and clinical research', *Social Studies of Science*, 42, 2, pp.214–36.
- Leonelli, S. (2013) 'Why the current insistence on open access to scientific data? Big data, knowledge production, and the political economy of contemporary biology', *Bulletin of Science, Technology and Society*, 33, 1–2, pp.6–11.
- Leonelli, S., Spichtinger, D. and Prainsack, B. (2013) 'Sticks AND carrots: incentives for a meaningful implementation of OS guidelines', *Geo: Geography and Environment*, 2, 1, pp.12–15.
- Macdonald, S. (1998) *Information for Innovation: Managing Change from an Information Perspective*, Oxford University Press, Oxford.
- Mahmuda, I., Baskaran, A. and Pancholi, J. (2014) 'Financing social innovation for poverty reduction: a case study of microfinancing and microenterprise development in Bangladesh', *Science, Technology & Society*, 19, 2, pp.249–73.
- Mauthner, N. and Parry, O. (2013) 'Open access digital data sharing: principles, policies and practices', *Social Epistemology*, 27, 1, pp.47–67.
- OECD (2007) *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD, Paris.
- OECD (2015) *Making Open Data a Reality*, OECD Science, Technology and Industry Policy Paper 25, OECD, Paris.
- Olmos-Peñuela, J., Benneworth, P. and Castro-Martinez, E. (2015) 'What stimulates researchers to make their research usable? Towards an "openness" approach', *Minerva*, 53, 4, pp.381–410.

- Open Knowledge International (2016) *Open Definition 2.0*, available from <http://opendefinition.org/od/2.0/en/> [accessed September 2016].
- Oyelaran-Oyeyinka, B. (2006) 'Systems of innovation and underdevelopment', *Science, Technology and Society*, 11, 2, pp.239–69.
- Packer, K. and Webster, A. (1996) 'Patenting culture in science: reinventing the scientific wheel of credibility', *Science, Technology, & Human Values*, 21, 4, pp.427–53.
- Parker, M. and Bull, S. (2015) 'Sharing public health research data: toward the development of ethical data-sharing practice in low- and middle-income settings', *Journal of Empirical Research on Human Research Ethics*, 10, 3, pp.217–24.
- Pisani, E. and Abou-Zahr, C. (2010) 'Sharing health data: good intentions are not enough', *Bulletin of the World Health Organization*, 88, 6, pp.462–66.
- Pisani, E., Whitworth, J., Zaba, B. and Abou-Zahr, C. (2010) 'Time for fair trade in research data', *Lancet*, 375, 9716, pp.703–5.
- RCUK (2013) *RCUK Policy on Open Access and Supporting Guidance*, Research Councils UK, London.
- Research Information Network (2009) *Patterns of Information Use and Exchange: Case Studies of Researchers in the Life Sciences*, Research Information Network/British Library, London.
- RIN/NESTA (2010) *Open to All? Case Studies of Openness in Research*, Research Information Network and National Endowment for Science, Technology and the Arts, London.
- Rip, A. (1994) 'The republic of science in the 1990s', *Higher Education*, 28, 1, pp.3–23.
- Royal Society (2012) *Science as an Open Enterprise*, The Royal Society, London.
- Tangcharoensathien, V., Boonperm, J. and Jongudomsuk, P. (2010) 'Sharing health data: developing country perspectives', *Bulletin of the World Health Organization*, 88, 6, pp.468–69.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., Manoff, M. and Frame, M. (2011) 'Data sharing by scientists: practices and perceptions', *PLoS ONE*, 6, 6, pp.1–21.
- Wessels, B., Finn, R., Linde, P., Mazzetti, P., Nativi, S., Riley, S., Smallwood, R., Taylor, M., Tsoukala, V., Wadhwa, K. and Wyatt, S. (2014) 'Issues in the development of open access to research data', *Prometheus*, 32, 1, pp.49–66.
- Zeitlyn, D. (2003) 'Gift economies in the development of open source software: anthropological reflections', *Research Policy*, 32, 7, pp.1287–91.