

Paper accepted for publication in Science and Public Policy in April 2016

Beyond the Digital Divide: Towards a Situated Approach to Open Data

Louise Bezuidenhout, Sabina Leonelli, Ann Kelly and Brian Rappert

Abstract

Poor provision of information and communication technologies in low/middle-income countries represents a concern for promoting Open Data. This is often framed as a “digital divide” and addressed through initiatives that increase the availability of information and communication technologies to researchers based in low-resourced environments, as well as the amount of resources freely accessible online, including data themselves. Using empirical data from a qualitative study of lab-based research in Africa we highlight the limitations of such framing and emphasize the range of additional factors necessary to effectively utilize data available online. We adopt the ‘Capabilities Approach’ proposed by Sen to highlight the distinction between simply making resources available, and doing so while fostering researchers’ ability to use them. This provides an alternative orientation that highlights the persistence of deep inequalities within the seemingly egalitarian-inspired Open Data landscape. The extent and manner of future data sharing, we propose, will hinge on the ability to respond to the heterogeneity of research environments.

Funding

The research informing this article was supported by a grant from the Leverhulme Trust title 'Beyond the Digital Divide' (RPG-2013-153). Sabina Leonelli was also funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ ERC grant agreement n° 335925. Ann Kelly's contribution was also supported by UK Economic Social Research Council Urgency Grants mechanism (ES/M009203/1).

1. Introduction

“The “digital divide” – a term that refers to the gaps in access to information and communication technology (ICT) – threatens the ICT “have-nots”, whether individuals, groups or entire countries” (OECD 2015).

“The digital divide keeps billions of people, including millions of serious scholars, offline” (Suber 2014).

“Without open data policies, it is not possible for developing countries to close the digital divide” (CODATA-PASTD 2015).

The growing interest in harnessing ICTs to create new forms of data dissemination has precipitated the development of activities focused on realizing the ideals of “Open Science”, and particularly Open Access and Open Data. These activities have considerably increased the amount of publications and information freely available online (Suber 2014; Royal Society 2012). Despite these achievements, disparities in ICT provision between low/middle-income countries (LMICs) and high-income countries (HICs) are widely recognized as presenting a considerable challenge for researchers attempting to engage with resources available online, and particularly with the diversity of file types and infrastructures involved in the sharing of data. As exemplified by the quotes above, the cumulative effect of such disparities on research is often portrayed as a “digital divide”, a term that has a long history within ICT discussions.

In this paper, we take issue with the fruitfulness of this framing for the conceptualization and effective promotion of Open Data (OD). While usefully drawing attention to ICT inequalities, the idea of a digital divide emphasizes the overall provision of online resources, and the extent to which researchers based in different locations can access them. This focus on access to material assets leads to a binary opposition between those who “have” and those who “have not”, and thus runs the risk of obviating more subtle questions regarding what researchers want to achieve through engagement with data, and what kinds of resources are needed, by whom and for which purposes within and across specific research settings (c.f. Duque et al. 2005; Shrum 2005).¹ In contrast, we propose to consider the conditions under which research inputs and outputs are not only accessed, but also interpreted and used to generate new insights and productive collaborations.²

This line of inquiry requires careful consideration of research conditions in low-resourced settings, and the generation of a multifaceted picture of how scientists in these settings engage with data in daily practice (c.f. Davidson, Sooryamoorthy & Shrum 2002). To contribute to such an understanding, our analysis builds on in-depth interviews and participant observations of laboratory work carried out with chemists and biochemists based in South Africa and Kenya in 2014 and 2015. By providing a window on the conditions under which these researchers produce knowledge and handle data, this study acts as a counterpoint to the vast majority of contemporary scholarship on data practices in the sciences, which is focused on research activities carried out in HICs such as the US and UK (e.g. Bowker 2005; Evans 2010; Leonelli 2010; Edwards et al 2011; Whyte et al 2011; Acord and Harley 2012; Mauthner et al. 2012; Stevens 2013; Borgman 2015). Like that scholarship, we stress the infrastructural, social, institutional, cultural, material and educational elements necessary to ensure the realization of openness, and in particular effective data sharing and re-use. At the

same time, our focus LMIC context-specific dimensions further highlights discrepancies between ideals and practices that animate international discussions on OD and the sheer diversity of considerations that shape research practices around the world (c.f. Frandsen 2009).

Much of the written evidence on laboratory conditions in LMICs comes from the numerous research consortia or large-scale collaborations that include both HICs and LMICs, and are highly influential in the developing world (Malaria Genomic Epidemiology Network 2008; Tierney 2013; Tindana 2014). This literature provides excellent insight in collaborative activities and formats across countries, and yet these activities typically include only the most privileged laboratory environments in LMIC settings. By contrast, our study specifically aims to tell the story of researchers based in laboratories that are not part of international consortia and work under low-resourced conditions. This is of considerable importance to future discussions on OD, especially when one considers that the vast majority of academic researchers in LMICs are not affiliated to research consortia or large-scale collaborations, and yet a considerable amount of the publications from LMICs are in conjunction with HIC collaborators (United Nations 2015).

This paper builds upon an emerging, and increasingly rich, social scientific literature exploring the research conditions in LMICs, particularly those belonging to the life sciences. Such studies, which also include reports from research bodies such as the Association for Commonwealth Universities and the International Foundation for Science, consider a wide range of factors including culture and politics (e.g. Pollock 2014), responsible conduct of research (e.g. Bezuidenhout 2014), the relationship between science and society (e.g. Kelly 2012), perceptions of local research cultures (e.g. Gaillard and Tullberg 2001) and issues

surrounding Open Access (Harle 2010). Nonetheless, studies linking laboratory conditions to perceptions (and enactments) of OD remain scarce. This paper is conceptualized as a step towards addressing that gap.

The issues that can constrain researchers working in low-resourced research laboratories from benefiting from, and contributing to, OD activities, are manifold. We highlight how the binary framing of the digital divide risks to replicate, rather than challenge, existing assumptions about the distinctions between the developed and developing world, and could ultimately hamper, rather than promote, equality. Simply put: more contextuality is needed in OD discussions. As an alternative, we suggest using the “Capabilities Approach” to reframe the challenges of OD and foster a more holistic and situated approach to data engagement and thus a critical re-evaluation of the efficacy of current data sharing policies and discussions. In closing, we question what is necessary to gauge the realization of the global equity aspirations of the Open Science movement as it relates to OD, and how the current discourse may be modified in order to take these issues into account.

2. Open Data as Resource Access

2.1 Open Data for all

Whether focused on Open Access or OD, key pronouncements on Open Science such as the Bethesda (2002) and Berlin (2003) statements and the Panton Principles (2009) articulate a common aspiration: that (publicly funded) scientific results should be made available in a variety of forms (ranging from publications to data) and that they should be freely available to all. With regard to OD, the Panton Principles state that: “[f]or science to effectively

function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made open.”³ Similar sentiments were expressed by the US National Committee for CODATA (1997: 10) in their report *Bits of Power*, which declared that: “[t]he value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research”. What unites the varied discussions of OD is not only the recognition of the importance of unimpeded “collection, analysis, publication, reanalysis, critique and reuse” (Molloy 2011) of data but also the expectation that any researcher should engage in these practices, regardless of nationality, discipline, or place of work.

ICTs have provided new avenues for sharing, storing and reusing data quickly and widely (Hey et al. 2009), and indeed an increasing amount of research data is becoming accessible through databases and other sharing platforms such as personal websites, e-books, discussion forums, email lists, blogs, wikis, videos, audio files, RSS feeds and P2P file-sharing networks (e.g. Kitchin 2013; Suber 2014). Moreover, there are a wealth of initiatives that seek to ensure adequate recognition for those who disclose data, ranging from the launch of “data journals” designed specifically to publish datasets in a citable and recognizable format to institutional efforts to promote a culture of data sharing amongst scientific communities (Editorial 2013; Borgman 2015).

Many concerns have been identified as potential challenges to making data available to all, including the financial paywalls of for-profit research publishers, restrictions on data reuse by publishers and producers, concerns about appropriate credit and responsibility, confusion over what platforms to use and technical complications arising when collecting, storing and disseminating objects produced in large quantities, different formats and from a wide variety

of sources (e.g. Acord and Harley 2012; Calvert 2012; Caufield et al. 2012; Leonelli et al. 2013). These concerns are typically conceived as ‘barriers’ to be overcome, and a considerable amount of effort has gone into interrogating the means by which copyright and ownership requirements can be balanced against the benefits of placing data, as well as published articles, in the open domain (e.g. Tenopir 2011).

2.2 LMICs infrastructure and Divide rhetorics

While these initiatives are intended to serve the global scientific community, it has been recognized that researchers in LMICs may require additional assistance in order to utilize online resources, particularly given the low penetration of LMIC Internet users in comparison to their HIC counterparts.⁴ Whether through lack of finance, investment, political will, national and regional instability or social and educational barriers, ICT usage in LMICs continues to lag.

Universities in LMICs often occupy relatively privileged in-country positions with regards to ICT provision, and yet sustained challenges vis-à-vis access to online resources remain. Thus, many of the initiatives aimed at LMICs are focused on “unlocking” Internet resources for scientists in less-resourced settings. It is in framing these issues that OS policies and discussions use the concept of a “digital divide” to describe the discrepancies between North and South, as highlighted above. This terminology harks back to the early 1990s, when worries were first raised about the impact that heavy reliance on ICT technologies would have in locations where access to necessary equipment, training and resources could not be guaranteed (e.g. Molla 2000; Obijiofor 1998).

The IT access gap has also been highly influential in driving many technology-based initiatives (such as frugal design) that focus on sustainable technology development in LMICs. For instance, the explosion of mobile medical technologies, or mhealth, has been championed as a game-changer for health care delivery in resource-poor settings—particularly in the African region—precisely as it ‘leapfrogs’ the kinds of infrastructures and human infrastructures conventionally believed essential to a functioning public health system (DeRenzi et al. 2011; Lester et al. 2010). As observers have noted, mhealth poses considerable regulatory challenges, particularly concerning the security, management, and ownership of data. But perhaps more pressing is that as a substitute for infrastructures, these platforms, do little to address the underlying social determinants of health; indeed, social scientists have demonstrated the limitations of technological solutions to health by illuminating the wider social-political and infrastructural systems needed in order for ‘magic bullets’ to work (Vélez et al. 2014).

Wider initiatives to enhance the participation of LMIC scientists in OD are similarly focused on the legal and financial barriers that limit access to data online—in effect the divide between access and no access.⁵ Without wishing to diminish the significance of these achievements, it is also important to ask what happens once access has been achieved, how research environments shape scientists’ engagement with these online resources, and how, conversely, scientists can use OD to further develop their research environment.

Such concerns are not new. Many scholars in broader ICT discussions have questioned the use of a terminology that partitions up the world into “have” and “have nots” by considering what people had access to rather than at the conditions under which those resources could be effectively used (DiMaggio 2001; Shrum 2005). Despite such warnings, the concept of

digital divide continues to play a significant role in framing discussions around Open Science and particularly Open Data (e.g. Ford 2007). It is in translating these concerns to the OD discussion – particularly in relation to LMICs - that this paper contributes to current discussions.

That scientific knowledge production depends on the material and social conditions under which research is performed and data are situated and analyzed is well established (see Bowker 2005; Leonelli 2014; Kitchin 2014). However, framing these considerations within OD discussions remains challenging. In particular, more systematic empirical research is needed on how practices of data engagement, including data generation, curation, storage and dissemination, can be represented in a manner that considers the conditions that enable the use of data. What aspects of research environments facilitate the movement of data onto and off the Internet? And how do these vary depending on disciplinary cultures, community ethos and geographical location?

With regards to LMICs these questions remain difficult to answer for a variety of reasons. First, there is little in the way of systematic empirical investigations into material and social research environments in LMICs, with minimal consideration given to the working environments of laboratories that are not affiliated to international research networks. Second, few studies have investigated social attitudes of LMIC scientists to data and data sharing (Carr and Littler 2015: 315). Although this situation is improving with regard to researchers who donate their own data (Bull et al. 2015), rather than *using others'*, 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 including the life sciences (e.g., Pisani and Abou-Zahr 2010; Tindana 2014). Third, while a growing number of educational

initiatives have focused on building OS capacity in LMICs—particularly with regards to OA and altmetrics—these initiatives have not extensively investigated attitudes to data and data sharing. In order to address these issues we now examine the day-to-day research environment confronted by LMIC scientists, whose characteristics risk being made invisible within approaches focused largely on access and resource provision.

3. Research Design

The data presented in this paper was gathered during embedded visits to four university departments in Kenya (KY) and South Africa (SA). This study employed a qualitative approach, involving semi-structured interviews and participant observations.

3.1 Sampling and Site Descriptions

The four fieldsites were chemistry/biochemistry departments in national research institutions, with common research themes such as malaria, water quality and medicinal plants. These commonalities guaranteed that interviewees shared a minimal set of disciplinary commitments. All four departments could be viewed as examples of “homegrown” research in Africa. They had a range of research projects underway, some of which were funded from a number of national and international sources, but were not directly affiliated to large, international research networks or consortia.

They all had access to the Internet - either through wireless or cable connection. Staff and graduate students had access to a computer to be able to engage with online resources, and all four institutions had libraries with some level of online access to journals and other academic

resources. Thus, from an external perspective, these departments were on the “have” side of the “digital divide”, and would be assumed to have the capability not only to use online data, but also to disseminate the data that were generated in the course of their own research.

While the institutions in which these departments were based varied in terms of longevity, financial provision and size, they were nonetheless united by certain commonalities. Staff members had high teaching and supervision workloads, particularly in comparison to peer faculty in European countries and North America. As a result, researchers reported difficulties in finding time for research—particularly time to conduct experiments or supervise within the laboratory themselves. Thus, the majority of data generation was done by graduate students. Interestingly, however, at these institutions the promotion of staff was directly linked to publication outputs in the form of journal articles. All staff interviewed expressed a feeling that a lot of pressure was placed on publishing, despite varying degrees of support for research, patenting and the development of ICT skills.

Interviewees received little in the way of core funding from their institutions (or in the case of Kenya from their national government). Indeed, the vast majority of funding was provided by international funding sources for specific, one-off, projects.⁶

3.2 Data Collection and Analysis

Data collection was primarily carried out over a period of five months between November 2014 and March 2015 with 56 interviews were conducted with staff and graduate students. Participation in the interviews was voluntary and subject to written consent, with promises of anonymity given. Interviewees were recruited through departmental emails and personal

communication. They were questioned on what data they used in their research, what they shared, and what challenges they perceived to fully exploiting the increasing openness of science online. Our analysis of the issues raised therein - particularly the perceived barriers to data usage and contribution - were further informed by observations of the laboratories and the working practices of staff within these facilities. Each lab was revisited in again in 2015. All interviews took place in English and were audio-recorded by the field researcher (**). The interviews were then transcribed and analysed manually using a thematic approach, as exemplified in table 1.

<i>Theme</i>	<i>Relevant conversion factors</i>
Being online	Sharing computers Lack of institutionally-provided ICT equipment Lack of proxy access for off-campus access to resources Power cuts and low bandwidth Cost of personal data bundles Use of self-funded older ICT equipment and software
Learning IT competence	Lack of training and development Lack of mentors
Generating data	Lack of trained technicians Lack of equipment Lack of core funding High teaching and supervision loads Power cuts
Storing data	Lack of training and development Lack of directed funding for data storage
Curating data	Lack of training and development Lack of directed funding for data curation

	Lack of standards
Sharing data	Lack of institutional policies Power cuts Lack of awareness of sharing options
Using data	Lack of mentors High student turnover causes data loss or inefficiencies Access payments
Owning data	Absence of coherent IP rights Lack of institutional policies
Speed of research	High supervision loads Slow delivery of reagents Lack of technicians and postdocs Lack of skilled technicians for equipment maintenance Need to outsource analyses
Cultures of data usage	Lack of mentors Lack of peer endorsement

Table 1: Thematic grouping of conversion factors identified in interviews. Each theme was singled out as significant in most interviews, with interviewees related them directly to each conversion factor.

The identification of conversion factors within the interviews was cross-referenced to the written observations made by ** during the laboratory visits. Key topics identified in both interviews and observations are tabulated below.

Lack of ...	SA1	SA2	KY1	KY2
Core funding for laboratory maintenance and upgrading	*	**	***	***
Technicians for equipment maintenance	*	*	***	***
Project-specific funding	*	*	**	***
Delivery of reagents or slow delivery (over 3 weeks)	***	***	***	***
Technicians in laboratory	**	**	***	***

Postdoctoral researchers	***	*	***	***
Assistance with supervision and high number of postgraduate students	***	***	***	***
High speed internet connection	NA	**	**	**
Wireless access	NA	***	**	**
Proxy server	NA	***	***	***
ICT hardware provided by institution	*	**	***	***
ICT technicians	*	**	***	***
Stable power provision (power cuts less than once every six months)	**	**	**	**
Sufficient back-up generators	*	**	***	***
Ability to conduct all methods in house (need to outsource analyses)	*	**	***	***

Table 2: Key observations relating to fieldsites. The prevalence of each issue is indicated by number of stars – three being highly prevalent. The assignation of stars is recognized to be subjective, but is based on the frequency of reports in the interviews, the correlation with observations and the previous experiences of the researcher who has worked in laboratories in both the UK and Africa.

3.3 A Capabilities Approach framework for Analysis

Many of the themes emerging from the analysis of the interview transcripts were closely related to what we termed “data engagement activities”. We took this to include data generation and research practices, data storage, curation and analysis, data dissemination and the reuse of online data. When considering these data engagement activities it was unsurprising that it was not possible to separate support, motivation and endorsement of these activities from the conditions under which they were occurring. This raised difficulties in integrating our data into the current, “provision focused” OD discussions.

In attempting to frame the relation between access to and use of research data, we turned to theories of economic development and social justice, and in particular, Amartya Sen’s Capabilities Approach (CA). This theory is the critical reference point for contemporary discussions on poverty and inequality and the theoretical foundation for the UN’s Human

Development Index. The CA suggests that the “freedom to achieve well-being is matter of what people are able to do and to be, and thus the kind of life they are effectively able to lead” (Robeyns 2011: 2). By focusing on the availability of human agency to transform resources into desired utilities, CA represents a shift away from traditional ways of measuring inequality that focus on “having or not having” access to resources and recognizes that individuals differ in their ability to convert existing resources into valuable opportunities or outcomes (Sen 1999). Thus, measuring resources or assets only reflects part of the situation and is fundamentally different from measuring functionality – where people don’t have the capability to use their resources in the means they advances their desired states of being and doing. CA acknowledges the overwhelming diversity of human capabilities and goals, as well as the interdependencies between such capabilities and the material and social environment in which humans operate.

When applied to data engagement discussions, a CA framework inspires a basic reframing of the conditions under which data is made “open”. Rather than focusing on data access by asking *What online resources are available?* the emphasis shifts to data use within specific research settings, and thus to questions such as *How can these scientists effectively utilize online resources to realize their research goals?*. In this manner, when applied to the openness of research, CA directs attention to those factors in a research setting - from the presence of basic laboratory materials to the hierarchical structures of professional advancement – that can influence the data engagement capabilities available to individual scientists. Moreover, it supports the need to recognize that the agency to address these multifarious factors is a vital precursor to effective data engagement. By characterizing data sharing in this way it becomes apparent not only that the provision of online resources cannot automatically lead to data engagement utilities, but that the research environment - and the

ability of the scientists to tailor these environments – are important in the realization of research utilities.

In effect, the CA highlights how important it is to critically examine the specificities of research environments in order to identify what Sen calls *conversion factors*. These are the considerations that frustrate or enable individuals' ability to effectively use resources available to them in order to pursue their sought after 'beings' and 'doings'. In relation to the topics of this paper, we conceptualize conversion factors as the characteristics of research settings that influence the degree to which the provision of a given resource (e.g. data themselves, particular kinds of ICTs) can be converted in a functional ability of scientists to pursue their research interests; and in particular their ability to undertake data dissemination, retrieval and analysis. A detailed list of conversion factors – of relevance to data engagement is provided below.

Although relatively new in discussions on general ICTs, the CA is also gaining traction as a means of analyzing – and designing– ICT structures, particularly for LMICs (Oosterlaken 2009; Alampay 2010). In this way, the CA draws attention not only to the conditions in which the scientists work, but also the way that data are presented to them online. It also highlights the need to critically assess the design of the technologies and the policies that govern them for social inclusivity and contextual propriety (Alampay 2010). From such framings it becomes evident that the: “divide in access and use of ICTs for development [is] more complex than just the absence of the needed infrastructure (ITU 2003)” (Alampay 2010: 5).

Adapting the CA for use in OD discussions enables considerable scope to consider contextuality in a substantive manner. It requires us to consider whether conversion factors exist within data infrastructures as well as within laboratory environments, and what impact they have on the success of improving access to online resources.

4. Disparities in Research Environments

Based on the fieldwork, in this section we seek to identify conversion factors within the research settings observed that inhibited the effective engagement of researchers with OD initiatives, and, also, to trace the ways in which these conversion factors influenced how researchers discussed data, its uses, and openness as part of their daily practice.

4.1 Identifying Conversion Factors in Data Engagement Activities

Within CA, conversion factors are typically categorized as ‘personal’, ‘social’ or ‘environmental’. In order to highlight matters particularly related to the conduct of science, we propose a modified classification scheme. As part of the following scheme, we name sub-factors identified on the basis of the fieldwork and provide an initial suggestion of how they stymied the effective utilization of data (points that are then elaborated further below):

- Personal
 - Data management and curation skills: Interviewees reported the absence of training and development in data management and curation.
 - Technical servicing: All sites were characterized by a lack of trained technicians to service and repair laboratory equipment, and the absence of functioning

relevant equipment.

- Communal
 - Mentorship: Mentors with ICT skills and knowledge on data engagement activities were widely seen as missing for both staff and research students.
 - Endorsement: Support for data sharing from peers and supervisors was reported as not present.
 - ICT sharing: Some researchers reported a need to share existing computers which curtailed the time each individual could spend online.
 - Ownership: Absence of clear intellectual property rights lead to data hoarding in some cases.

- Organizational
 - Policies: Dearth of institutional policies such as data sharing guidelines.
 - Procurement: Complicated and restrictive procedures for procuring and reimbursing ICTs (e.g. paying for software).
 - Discretion: Because of the reliance on project based funding, researchers had a limited flexibility in how to spend research income.
 - Workplace demands: The extent of teaching loads reduced the time available to do research.

- Infrastructural
 - Remote access: The absence of proxy servers reduced researchers' ability to make use of university resources when off-site.
 - Basic provisions: Irregular power supply caused breaks in down/uploads as well

as limited functional time with ICTs.

- Transfers: Border controls slowed down data generation and analysis.

- Epistemic
 - Research continuity: High graduate student turnover resulted in data loss, inefficiencies and a diminished ability to develop cohesive research streams.
 - Dependency: The lack of equipment availability and technical support meant that standard forms of analysis elsewhere were outsourced or not possible; leading to lowered self-perceptions of peers' assessment of the quality of science undertaken in these labs.
 - Lack of standards: Diversity in data formats and labels made it difficult for researchers to assess the compatibility and significance of data classified and formatted by others, and thus to reuse them.

- Economic
 - Access payments: Especially for graduate students, the need to purchase personal data bundles proved expensive.
 - Personal provisions: The self-funded purchasing of computers and software led to the use of older machines with older software versions.

Within the interviews, researchers regularly linked these conversion factors to their data engagement and, more specifically, their inability to generate and manage data and thereby their disinclinations towards sharing data. Some of these factors are recognized in current OD discussions. For instance, the need for better data management skills, for enhanced research community sharing norms, and for the availability of affordably priced

data have been noted as important in ensuring openness can flourish in developing countries (e.g., ICSU et al. 2015). Other factors however – such as those related to the lack of discretion in spending or the high reliance on (often short term) students – are not routinely acknowledged.

Unsurprisingly, some conversion factors were more pronounced in some fieldsites over others, highlighting not only the heterogeneity of research environments, but also the extremely contextual nature of data engagement discussions. With the Kenyan sites, as evidenced in table 2, the sheer extent of teaching demands, the absence of research dedicated staff, and the costs of data and equipment were pronounced. But even in the South African sites where access to national research funding sources comparatively reduced such concerns, pressing factors not often acknowledged in OD discussions were reported. Regular power cuts and the time required for sending samples across borders, for instance, meant research was slowed in comparison to relatively well resourced competitor labs elsewhere.

In many respects though, it was not the presence of one, two, or three of these inhibiting conversion factors that limited researchers engagements with data. Rather, it was the accumulation of them that proved so taxing. Thus even if individual conversation factors could be addressed, doing so in an isolated manner would be unlikely to result in transformative improvements in data engagements.

At this point we would signal two general implications that follow from the points in the previous paragraphs. One, efforts to promote data openness need to acknowledge and address a diversity of conversation factors. Along these lines, it is important to acknowledge that the factors identified as pertinent through our research stem from vibrant examples of

‘home grown’ research in sub-Saharan Africa. Two, the relevance of inhibiting factors is poorly conceived in terms of a simple distinction of whether they are present or not. Instead, they exist along a continuum and their relevance depends on comparisons made about their prevalence elsewhere.

In the sub-sections that follow we elaborate how the absence of certain factors contributed to a set of (largely negative) outcomes, namely termination of research aspirations, lowering the speed of research activities, or changing the direction of research activities.

4.2 Conditions for Data Usability

In the laboratories visited during the fieldwork, all interviewees had access to the Internet as well as a laptop or desktop - placing them firmly “online”. Nonetheless, the interview data consistently raised additional challenges that the material environment presented to the scientists’ daily data engagement activities. A picture emerged of a state of a lowered ability to engage with data that was directly linked to conditions within the research environment.

To elaborate, while all interviewees had access to a computer, many of them (particularly in Kenya) had been required to purchase both the hardware and the software. As a result, they were using older computers and software - issues which shaped their online activities. These factors impacted on their speed of browsing, uploads and downloads; the range of online tools that they were able to access (such as plugins); as well as many other issues relating to the format of data and its presentation online. One discussion in particular framed this issue, saying: “[t]he University has really tried to make the internet available to all of us. So online we can always connect. And now it depends on the individual person – do you have a

computer, or what computer you have” (KY1/2: staff member). This situation was particularly severe amongst the graduate students that conducted the majority of research. For them the expense of new hardware, software and software updates represented a considerable expense and often meant that they resorted to working on older machines with earlier software versions.

The difficulties of working effectively online were also aggravated by the absence of necessary ICT infrastructures at the sites. Although all the universities that were visited provided online access to certain journals through their libraries, 3 out of the 4 sites had no working proxy server. Thus, while the staff and students technically had access to a range of digital resources, they were not able to access them when they were off campus. Furthermore, organizational conversion factors such as teaching being the primary focus of departments meant that “all other things come by the side” (KY1/1: staff member). As a result, most of online activities needed to be performed out of office hours - and thus predominantly off campus. These concerns linked to physical conversion factors so that: “no, from home you can’t [work]. You see, from here [at the university] I’m using wifi, so the moment you step out of the college you’re shut off and again in the estates where we stay as of now the Internet is expensive. It’s not affordable. So I do as much as I can here so that when I go back home I’m going to rest” (KY1/3: staff member). In such statements it is also apparent that the high cost of personal Internet provision played a key role in determining the times and style of interviewees working patterns.

Similarly, erratic power and Internet provision were commonly cited as challenges to daily data engagement. As one participant at SA2 put it: “so when there’s internet it’s fast enough, but now – I don’t want to say most of the times – but there’s times when we don’t have

access to the internet” (SA2/7). Another participant at the same university said: “you need a lot of patience – waiting when the Internet is not strong enough to allow you do download things” (SA2/12).

These issues affected the ways in which the interviewees could interact with the data available online, as well as their ability to generate, store and disseminate data. Issues such as the time taken to upload data in low bandwidth conditions, the time and expense of cleaning data and the lack of IT support for storage and curation solutions were commonly cited as barriers to data sharing. Additional challenges that were mentioned in the Kenyan sites were related to data being in the wrong format. As one participant said: "I have just seen somebody requesting about four papers that I published. But now the problem is that I could be having that paper but it is not in pdf form or it is not digital. The titles I have put there, the titles and abstract I have put online, but it is now the full text that they want. And sometimes it is not easy for me to send that because I may not have full text in digital" (KY1/2: staff member). This epistemic factor was identified by a number of other participants, and was – at least in part – related to the publication of articles in local journals and university publications that did not maintain an online archive.

Particularly in relation to data generation, the issue of equipment was also understandably linked to issues of time – the time it took to do research - to collect and process data. Such considerations were invariably linked to issues of data sharing, ownership and dissemination by interviewees. Stories such as: “we have limited lab facilities. Our equipment is not running or idle. We have an AS (automated peptide synthesizer) 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: staff member). As a result, researchers were often obliged to send

samples away for analysis in HICs – something that considerably slowed down the research process. This influenced the speed at which data were disseminated: data were produced slowly and often privately stored until the research project was completed, which could take years.

By relating these (and other) conversion factors to the material issues of daily data engagement activities scientists clearly demonstrated that the issues experienced by participants could not be summarized as a lack of access to online resources. The innocuous aspects of their research environment played important roles in how they created, chose to share and accessed data for re-use – despite all of the participants being “officially online”. Indeed, the current framings of data sharing discussions leave little room for these issues to be gainfully recognized and explored.

4.3 Cultures of Data Engagement Reflecting Laboratory Environments

If one considers the actions of scientists from a “digital divide” perspective, it would seem reasonable to expect that once the “divide” is overcome, scientists would enthusiastically participate in data sharing - both as data contributors as well as users. The fieldwork highlighted the limitations of such assumptions, and showed how the influence of a low-resourced laboratory environment affected how ‘openness’ was discussed by the interviewees. While interviewees demonstrated a high degree of support for the Open Science movement in principle (“I think it leads to better science”: SA1/3), the expectations and practices of their laboratory environments significantly impacted on their interpretation of what it means to be a member of their scientific community, what responsibilities it entails and what it means in terms of ownership of the data. Nonetheless, while there was

widespread theoretical endorsement for the idea of openness in research, when it came to actual practices of openness, the responses were markedly different. As one participant succinctly said: “[P]eople are just locking it [data] away in their computer”: SA2/7).

The issue of the time taken to gather and process data played an important role in many interviewees’ perceptions of trust in sharing - particularly when sharing unpublished data - and their willingness to share data. The idea of “being scooped” came up regularly in most interviews – linked to a variety of conversion factors. Although this has been internationally reported as a common concern amongst scientists (Ferguson 2014), issues arising in their research environments vis-à-vis to their geographic position undoubtedly exacerbated this concern amongst interviewees (see as well Bull et al. 2015). Interviewees widely offered variations on statements such as: “[B]ecause it takes us so long to complete our research, other people have a lot of opportunity to steal our data. We must keep it secure until we publish” (KY1/10), thereby linking the time for research to their lack of openness. The idea of keeping data secure was further reiterated in statements such as: “[E]ven when you’re hiding your data, anyone can run away with it”.

This tense distinction was linked to the difficulties of doing research in these departments, as the high teaching loads and insufficient provisions undermined research activities. As one participant succinctly said: “[T]he research agenda itself is struggling for survival in a lot of African institutions” (SA1/12). These worries were compounded by the fact that in two of the institutions visited the lack of national and international funding had caused researchers and graduate students to fund their projects with their personal money. Three researchers explicitly mentioned using their own money to buy reagents and equipment necessary for research - something that would not be allowed in most HIC institutions. This personal

investment was reflected in the manner that data ownership was discussed. Interviewees talked about the need to accrue some sort of personal benefit - either in the form of a promotion or as a patent - to justify the investment. Sharing these data was not an option until the some concrete benefit had been realized.

These concerns were linked to the absence of institutional or governmental support for data management (see Tangcharoensathien et al. 2010), and the complete lack of awareness of licensing options such as the Creative Commons. These deficiencies fueled fears that data would be appropriated by others without recognition: “with the size of [SA2] we don’t have the same legal power like a university in Australia or America. If someone steals their idea they will go for them. But we are small and who is going to believe me when I say “this was my idea”. So there is that fear” (SA2/11). This lack of agency to actively counter misappropriation of data was perceived by many interviewees to be a serious barrier to participating in OS initiatives - from altmetric engagement to the dissemination of data sets post-publication.

The reluctance to share data was further exacerbated by individual interpretations of what types of data were valued by the OD movement. While the interviewees did not raise concerns about the *quality* of their data, they felt that the conditions under which they were produced (using older equipment and methodologies) - and where they were produced – were substandard with respect to laboratory work in HICs, particularly with respect to using older equipment and methodologies. They thus feared that sharing data would cause them to be judged negatively by the online scientific community. As one interviewee in South Africa highlighted: “[I]f you’re a reviewer you’re going to be harsher on the [LMIC] guys and a little more lenient on the others [from HICs]. You’re going to be, like, from the onset, these

guys seem to know what they're on about" (SA1/2: graduate student). Such quotes draw attention to concerns that the data they created would not be gainfully used if shared. As one scientist mentioned: "...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: staff member). The influence of conversion factors thus had epistemic implications beyond the engagement of scientists in OD initiatives.

Perceptions that data shared would be undervalued were exacerbated by the choices of research subjects. Common themes included studies on medicinal plants or studies on water purification, which have often been described as "niche areas" for African scientists to exploit. These choices were often the result of strategic financial- and equipment-related decisions by the scientists (enabling them to do the most with little), however they also deepened the gap in interests and expertise separating these researchers from their HIC counterparts. Exploiting a niche helps researchers by diminishing the risk of competition with richer groups, enabling a better use of local resources and a heightened sensitivity to the needs of the local population. At the very same time, this choice makes it harder to find relevant data online, and interviewees were thus unable to use existing data to inform their own research. The lack of formal training about online data or the use of altmetric tools compounded this problem, heightening perceptions of isolation and marginalization.

Conversion factors present in the laboratory and institution thus play a fundamental role in shaping scientists' data engagements. It is of particular importance to note that the provision of access to online resources should not be taken as a causal precursor of support for data engagement initiatives. Indeed, understanding the conditions in which data are generated and re-used is vital for understanding attitudes to calls research openness.

5. The Implications of Capabilities for Open Data

The fieldwork thus highlights some important considerations for OD discussions. First, that the conversion factors present in the research environments had significant effects on how research was conducted in these laboratories. This had far-reaching implications linked not only to the data that were generated, but also to how scientists understood responsibilities to share and disseminate it. Current OD models that assume a linear progression between increased openness and increased research outputs cannot appropriately model these concerns.

These observations echo current criticisms of the “digital divide” that have appeared in general ICT studies literature, which condemn this dichotomous approach as “simplistic, formalistic and thus idealistic” (Burgelman 2000: 56). In particular, DiMaggio and Hargittai (2001) proposed that problems with the utilization of ICT should be viewed in terms of a “digital inequality”. This includes not only the differences in access, but also (1) inequality among persons with formal access to the Internet, which can be manifested in varying access to equipment, restrictions to autonomy of use, skill of users, lack of social support and differing purposes of use; and (2) inequality in economic and political conditions under which individual resources can be expressed and used. Redefining “access to ICT” in social as well as technological terms (DiMaggio 2001: 3) makes it possible to interrogate what “complex mixture of social, psychological, economic and, above all, pragmatic reasons” (Selwyn 2004: 348) might affect the reuse of data available online.

Our fieldwork clearly supports these observations and highlights the limitations of perpetuating the “divide” rhetoric. The empirical study of this issue raises another important consideration: while OD scholars have long recognized that it is impossible to separate the scientists’ perceptions of data, openness and sharing from the structures of their research environment, the lack of discussion on the heterogeneity of research environments masks the significance of this observation. The research environments – and the conversion factors present in them - of the departments we visited had a marked effect not only on the scientists’ ability to participate in their data engagement activities, but also on their perception of their own abilities and opportunities in relation to their HIC colleagues. For instance, a South African academic discussing the systemic issues present within their research environment stressed out difficult it was to raise these issues within international discussions: “I find it difficult [that] people don’t understand our situation – it’s not bad will, it’s just not being able to figure it out” (SA2/12) – thus drawing attention to the dearth of information about research conditions in non-HIC laboratories.

In line with a number of studies that have examined ICT diffusion in LMIC (e.g., Duque et al. 2005) and in particularly Africa, our research points out that access to data does not necessarily lead to data use and thus, increased scientific outputs. Indeed, the fieldwork emphasizes more pervasive, nuanced challenges to the application of data—a more insidious form of inequality that can be exacerbated by an emphasis on ‘universal service and universal access’ (Selwyn 2004: 345). Our list of conversion factors, without pretending to be exhaustive, illustrates the variety of factors influencing data engagement activities in LMICs. It is impossible to provide a stable ranking of the relative importance of those factors, or the order in which they should be tackled, since the significance of each factor and relevant priorities change depending on the specific situations at hand. Nevertheless, considering the

multiplicity of conversion factors involved in data engagement indicates the difficulties with applying stark access/no access distinctions to these situations, and the advantages of viewing the problem of access as a continuum, thus attempting to increase access by gradually addressing each relevant factor and considering the wider implications of each intervention.

Moreover, our study emphasizes the additional problems associated with getting buy-in for OD by scientists in LMICs who continually work in low-resourced settings. While theoretical support for OD was highly prevalent amongst the interviewees, none reported robust OD activities within their daily research. The distinction between the “ideal” and “real” should be of serious concern for discourse promoting a global approach to OD.

Ultimately, this leads to a couple of serious problems – including the fact that some scientists do not have the capabilities to exploit the online data resources to the same degree or in the same way as their colleagues in high-income settings has considerable implications, and that they often do not share the data that they do generate to the same degree as their HIC colleagues. Such concerns strike at the heart of the commitment to egalitarianism that drives most contemporary OS discussions.

5.1 Laying the Foundations for New Divides?

As outlined above, the continued use of the “digital divide” framing in OD discussions influences how initiatives are designed to build research capacity in LMICs – hence the focus on resource provision instead of capability strengthening. Perpetuating this approach has far-reaching consequences, the most extreme being that that science in these regions may continue to progress at a slower speed than in HICs – a phenomenon exacerbated by the

accelerating speed of North American and European research. It is possible that the already-marginalized research communities in LMICs will be further disadvantaged by not being able to effectively take advantage of the growing “data deluge”. In addition, the adoption of global standards for data dissemination and re-use may end up hindering, rather than fostering, the diversity of conditions under which research can be successfully performed. A failure to address these issues may therefore increase the gap between research in ‘the North’ and ‘the South’ – instead of closing it as assumed. Moreover, eliminating the awareness of these issues from the design of future data engagement initiatives may continue to exacerbate these problems.

In relation to the online research environment, a number of other considerations must be raised. These include the linguistic and cultural hurdles experienced by users whose first language is not English (Wilson 2000; DiMaggio 2001). Linguistic and social choices have already been suggested as elements that shape how corporations and governments make strategic choices about website developments – thus transforming access and use (DiMaggio 2001: 17), but how these concerns are reflected in OD remain largely unexamined. Furthermore, it has been noted that the make-up of the Internet is culturally informed and the algorithms, website structures and search tools predominantly reflect elements of Western culture (Wacjman 1991; DiMaggio 2001). Similarly, the graphics and design of the websites themselves may be inappropriate for use in low-bandwidth areas. As a result, it may be possible that the very tools available to search for, access and reuse data may be problematic to non-Western scientists.

In addition to conversion factors influencing data usage, a further concern speaks to the resources that are available for reuse. If these data are predominantly collected by – and for –

scientists in highly resourced settings, it is possible that these data will be less useful and reusable to those in alternative contexts – particularly, as discussed above, when scientists (due to resource constraints) opt to work on more marginal topics of research. Such observations lead into a wider discussion about the significance of data practices in shaping the autonomy, directions and social utility of research. Moreover, the curation of data and the design of databases has considerable cultural content, as curators select, define, and annotate based on their own perceptions of what is happening in a field and what is necessary (Hine 2006, Leonelli 2010).

Challenges are particularly acute when the *relative* prestige, visibility, and outputs of data are considered. Many of the policy discussions about openness in science casts the benefits to society, science and scientists in general terms, without regard for the diversity of conditions under which research can be carried out. With a differential capability to partake in the demands of openness, though, it could be said to benefit those researchers more able to partake in developments, and may well increase their standing vis-à-vis those less able. As noted by the Royal Society, “The greater the strength of the home science base, the greater its capacity to absorb and benefit from science done elsewhere” (Royal Society 2012: 17). The corollary is that the weaker the home base (e.g. skillsets, infrastructure, networks), the weaker the capability to take advantage of circulating data, which then sets the conditions for capacity differentials in the future.

5.2 Future Implications

The perpetuation of inequalities in the OD movement not only undermines the egalitarian commitments central to it, but also risks to skew knowledge accumulation and dissemination.

These epistemic consequences relate to inaccessible data that refers to the selection, choice of dissemination and curation of data alluded to above. For instance, with regard to the introduction of ICTs in general, others have asked how their introduction can be understood as resulting in ignorance. Roberts and Armitage argued the growing importance of ICTs within ‘the knowledge economy’ should be understood as also resulting in a growth of ignorance. This is so through the manner in which tacit forms of knowledge that cannot be codified are disregarded; how what counts as knowledge gets skewed towards the agendas of those that are able to codify; and the demands of managing information as well as managing information about information (Roberts and Armitage 2008). Information, therefore, runs the risk of remaining in the hands of those who create it due to the subtle influences of selection and creation.

The possibility that scientists in LMICs not to be able to make the most out of the data revolution has significant consequences beyond the development of research in these regions. A number of studies have suggested that the new “data centric” model of science has made data into objects of market exchange, and thus market logics apply when considering investment in research and the potential for returns. As stressed by Leonelli (2013):

[t]o make it at all feasible for data to travel, market structures and political institutions need to assess not only their scientific value but also their value as political, financial and social objects: The increased mobility of data is unavoidably tied to their commodification.

It is necessary to ask, then, whether certain researchers are being left behind not only epistemically but also relatively to finance, infrastructure development and competition in the global market. Sharing between unequal partners can quickly become exploitative even when the purpose of such partnerships purports to be the ever-widening inclusion of publics

(c.f. Shrum 2005). The ethical furor over transnational clinical trials, for instance, hinges on the multiple orders of value that exist between access to patient populations (and their data) and access to health care. Beyond the vast discrepancies that exist between the benefits that accrue to research subjects and those to foreign scientists, the latter's ability to rapidly convert source materials into scientific outputs can effectively undercut the scientific capacities of slower research partners (Crane 2011). A better understanding of the unintended consequences of opening data and the expectations and obligations it creates, may shed light on the current problems identified in data sharing discussions (Kelly 2011; Lezaun 2014).

A more refined view of the difference between access and utility of data can also provide a basis for rethinking the distinctions between research environments. To the extent that the mundane everyday challenges in low-resource research environments are noted within Open Science and Open Data discussions, the distinction is routinely made between LMICs and HICs, with attention focusing on what might be done to address the resource deficiencies in the former. Obscured through this thinking are the commonalities between labs across varied geographical settings. In highlighting the conversion factors that influence scientists' ability to engage with data as users and producers, the issues raised in this paper are also relevant for relatively poorly resourced labs located within HICs, which raises questions concerning the very significance of distinguishing between low and high income countries, rather than for instance rich and poor labs. This is a complex issue that we will not attempt to resolve here, particularly the many concerns relating to self-perceptions of identity, geographic isolation and the legacy of postcolonial relations in LMICs. What our study has hopefully highlighted is the importance of further detailed empirical study of what distinguishes and unites researchers in low/middle- and high-income countries.

7. Conclusion: Moving OD Discussions into a Context-Sensitive Framework

We have shown that the “digital divide” framework, despite its usefulness in highlighting basic inequality issues in the implementation of OD mandates, is of limited use when attempting to tackle those issues. Simply put: an emphasis on access fails to capture the social and material conditions under which data can be made useable, and the multiplicity of conversion factors required for researchers to engage with data. Our empirical investigation of these conditions in sub-Saharan Africa shows the challenges of designing data sharing approaches that are both internationally meaningful while of practical utility in differently resourced research environments.

We believe that current data engagement structures, by focusing on resource provision instead of resource utilization, inadvertently perpetuate marginalization, exclusion and “data poverty” amongst some communities of scientists. In keeping with this “poverty” framing, we have proposed to shift the debate from how to bridge the digital divide to the importance of identifying the *capabilities* necessary to share data and exploit those available online within any one research setting. Conceptualizing knowledge production in relation to specific environments and beyond issues of inclusion and exclusion banishes the notion that capacity building is simply a matter of making more data available. Not only are considerable resources and diverse expertise are needed to transform data into new knowledge, but those resources and expertise vary widely across disciplines and research settings around the globe.

To quote the Global Research Council Report, “[t]he structure of academia and the research communities, the landscape of publishers, and the funding of research and publications vary from country to country just as the interaction between the stakeholder groups also varies.

Taking into account these differences, specific approaches towards implementing open access that are well suited for country A might not be feasible in country B. [Furthermore], in implementing open access, issues of language and standardization need to be taken into account as well as differences which might arise from differences between scientific disciplines” (Global Research Council 2013: 2). We argued that a better understanding of how capabilities may be addressed and fostered by data sharing structures used across such a complex and diverse research landscapes can help to ensure that future initiatives bridge – rather than exacerbate – divides.

While the issues raised by the CA are undoubtedly more apparent in LMICs, it is important to remember that they are not solely the problem of the global South. Even within HICs the existence of a “sliding scale” of access must not be overlooked. Unpacking privilege – in terms of both research environments and data access – within discussions on scientific research is both a vital and urgent need. Revitalizing OD discussions through the framing of the CA may be an important contribution to future science policy and provide a counterpoint to existing discussions on data as objects of market exchange.

References

Acord, S.K. and Harley, D. (2012) 'Credit, time, and personality: the human challenges to sharing scholarly work using web 2.0.', *New Media & Society* 15, no. 3 (2012): 379-97.

Avgerou, C. (2008) 'Information systems in developing countries', *Journal of Information Technologies*, 23: 133 – 146.

Basnyat, B. (2000) 'Letter to the editor', *The Lancet*, 356: 1036.

Bezuidenhout, L. (2014) 'Moving life science ethics debates beyond national borders', *Science and Engineering Ethics*, 20(2): 445 – 467.

Bodenheimer, T. (2005) 'The political divide in health care', *Health Affairs*, 24 (6): 1426-1435.

Borgman, C. (2015) *Big Data, Little Data, No Data*. Cambridge, MA: MIT Press.

Bowker, G. (2005) *Memory Practices* Cambridge, MA: MIT Press.

Burgelman, J. (2000) 'Regulating access in the information society: the need for rethinking public and universal service', *New Media and Society*, 2(1): 51–66.

Bull, S., Nia Roberts and Michael Parker (2015) 'Views of ethical best practices in sharing individual-level data from medical and public health research', *Journal of Empirical Research on Human Research Ethics*, 10(3): 225– 238

Bull, S., Phaik Yeong Cheah, Spencer Denny, Irene Jao, Vicki Marsh, Laura Merson, Neena Shah More, Le Nguyen Thanh Nhan, David Osrin, Decha Tangseefa, Douglas Wassenaar, and Michael Parker (2015) '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) 302–313.

Calvert, J. (2012) 'Ownership and sharing in synthetic biology', *BioSocieties* 7, no. 2: 169-87.

Caulfield, T., Harmon, SHE and Joly, Y. (2012) 'Open science versus commercialization: a modern research conflict', *Genome Medicine* 4, no. 2: 17.

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): 314– 316
DOI: 10.1177/1556264615593485.

CODADA-PASTD (2015) *Data Sharing Principles in Developing Countries*. Statement at the WSIS Forum, 26 May 2015, Geneva.

Crane, J. (2011) 'Scrambling for Africa?' *The Lancet*, 377(9775): 1388-1390.

Davidson, T., Sooryamoorthy, R., & Shrum, W. (2002) 'Kerala connections: Will the Internet affect science in developing areas', *The Internet in everyday life*, 496-519.

DeRenzi, Brian, Gaetano Borriello, Jonathan Jackson, Vikram S. Kumar, Tapan S. Parikh, Pushwaz Virk, and Neal Lesh. (2011) 'Mobile phone tools for field-based health care workers in low-income countries', *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* 78, no. 3 (2011): 406-418.

DiMaggio, P. and Hargittai, E. (2001) 'From the "digital divide" to "digital inequality" studying internet use as penetration increases' Princeton University: Centre for Arts and Cultural Policy Studies. Working Paper Series no.15.

Dixon, A. et al. (2011) 'Is the NHS equitable? a review of the evidence', London School of Economics and Political Science: LSE Health and Social Care.

Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D. B. S., & Shrum, W. (2005). 'Collaboration paradox scientific productivity, the Internet, and problems of research in developing areas', *Social Studies of Science*, 35(5): 755-785.

Duque, R., Shrum, W., Barriga, O., & Henríquez, G. (2009) 'Internet practice and professional networks in Chilean science: Dependency or progress?', *Scientometrics*, 81(1), 239-263.

Editorial (2013) 'Announcement: launch of an online data journal', *Nature*, 502: 142.

Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C. and Borgman, C.L. (2011) 'Science friction: data, metadata, and collaboration', *Social Studies of Science* 41, no. 5 (2011): 667-90.

Evans, J.A. (2010) 'Industry collaboration, scientific sharing, and the dissemination of knowledge', *Social Studies of Science* 40, no. 5: 757-91.

Faber Frandsen, T. (2009) 'Attracted to open access journals: a bibliometric author analysis in the field of biology', *Journal of Documentation* 65 (1): 58-82.

Ford, Dwedor Morais (2007). 'Technologizing Africa: On the bumpy information highway', *Computers and Composition* 24 (3): 302-316.

Fine, J. C. (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.

Gaillard, J., Tullberg, A. F. (2001) *Questionnaire Survey of African Scientists*. Stockholm: International Foundation for Science.

Global Research Council (2013) *Action Plan towards Open Access to Publications*, http://www.dfg.de/download/pdf/dfg_magazin/internationales/130528_grc_annual_meeting/grc_action_plan_open_access.pdf (Accessed 30/03/2014).

Goddard, M. and Smith, P. (2001) 'Equity of access to health care services: theory and evidence from the UK', *Social Science and Medicine*, 53: 1149–1162.

Harle, J. (2010) *Growing Knowledge: Access to Research in East and Southern African Universities*. London: Association of Commonwealth Universities.

Hine, Christine. "Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work." *Social Studies of Science* 36(2) (2006): 269–298.

International Council for Science, InterAcademy Partnership, World Academy of Sciences, and the International Social Science Council (2015) *Open Data in a Big World: An International Accord* Paris: ICSU.

Kaye J. (2012) 'The tension between data sharing and the protection of privacy in genomics research', *Annual Reviews of Genomics and Human Genetics*, 13: 415 – 431.

Kelly, A.H. (2011) 'Will he be there? a meditation on immobility and scientific labour', *Journal of Cultural Economy*, 4/1: 65-79.

Kelly, A. H. (2012) 'The experimental hut: hosting vectors', *Journal of the Royal Anthropological Institute (N. S.)*: S145 – S160.

Kelly, A.H. & Geissler, P.W. (eds) (2012) *The Value of Transnational Research: Labour, Participation Care*. London: Routledge

Kitchin, R. (2014) *The Data Revolution*. London: SAGE.

Leonelli, S. (2010) 'Packaging data for re-use: databases in model organism biology'. In: Howlett, P. and Morgan, M.S. (eds) *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*, pp.325-348. Cambridge University Press.

Leonelli, S. (2012) 'Making sense of data-intensive biological and biomedical research', *Studies in the History and the Philosophy of the Biological and Biomedical Sciences, Part C* 43: 1 – 9.

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 & Society*, published online 20 August 2013 DOI: 10.1177/0270467613496768.

Leonelli, S. (2014) 'What difference does quantity make? on the epistemology of big data in biology', *Big Data and Society*, 1.

<http://bds.sagepub.com/content/spbds/1/1/2053951714534395.full.pdf>

Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, Jack W, Habyarimana J, Sadatsafavi M, Najafzadeh M, Marra CA. (2010) 'Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomized trial', *The Lancet*, 376(9755): 1838-45.

Levin, N., Weckowska, D., Castle, D., Dupré, J. and Leonelli, S. (under review) 'How do scientists understand openness? assessing the impact of uk open science policies on biological research'.

Lezaun, J., and Montgomery, C. M. (2014) 'The pharmaceutical commons sharing and exclusion in global health drug development', *Science, Technology & Human Values*, 0162243914542349.

Malaria Genomic Epidemiology Network (2008) 'A global network for investigating the genomic epidemiology of malaria', *Nature*, 456: 732 – 737.

Mauthner, Natasha Susan, and Odette Parry. (2013) 'Open Access digital data sharing: principles, policies and practices', *Social Epistemology* 27, no. 1: 47-67.

Molla, A. (2000) 'Downloading or uploading? The information economy and Africa's current status', *Information Technology for Development*, 9(3-4), 205-221.

Nguyen, V. (2005) 'Antiretroviral globalism, biopolitics, and therapeutic citizenship'. In: Ong A. and Collier S. J. (eds.) *Global Assemblages: Technology, Politics, and Ethics as Anthropological Problems*, pp. 124–144. Malden, MA: Blackwell.

Obijiofor, L. 1998. 'Africa's dilemma in the transition to the new information and communication technologies', *Futures* 30 (5): 453-462.

OECD (2015) *Bridging the Digital Divide*.

<http://www.oecd.org/site/schoolingfortomorrowknowledgebase/themes/ict/bridgingthedigitaldivide.htm> (accessed 21/07/2015).

Oliver, A., Healey, A., Le Grand, J. (2002) 'Addressing health inequalities', *The Lancet*, 360(9332): 565-567.

Oosterlaken, I. (2009) 'Design for development: a capability approach', *Design Issues*, 25(4): 91 – 103.

Petryna, A. (2002) *Life Exposed: Biological Citizens after Chernobyl*. Princeton: Princeton University Press.

Pisani, Elizabeth and Carla Abou-Zahr (2010) 'Sharing health data: good intentions are not enough', *Bulletin of the World Health Organization* 88:462-466.

Pollock, A. (2014) 'Places of pharmaceutical knowledge-making: global health, postcolonial science and hope in South African drug discovery', *Social Studies of Science*, 44(6): 848 – 873.

Powell, M., Exworthy, M. (2003) 'Equal access to health care and the British National Health Service', *Policy Studies*, 24(1): 51-64.

Power, M. (1999) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Roberts, J., Armitage, J. (2008) 'The ignorance economy', *Prometheus*, 26: 4, 335-354.

Royal Society (2012) *Science as an Open Enterprise*. London: Royal Society.

Selwyn, N. (2004) 'Reconsidering political and popular understandings of the digital divide', *New Media and Society*, 6(3): 341 – 362.

Sen, A. (1999) *Commodities and Capabilities*. New Delhi: Oxford University Press.

Shrum, W. (2005) 'Reagency of the Internet, or, how I became a guest for science', *Social Studies of Science*, 35(5): 723-754.

Suber, P. (2014) *Open Access Overview*. <http://bit.ly/oa-overview>, accessed 03/03/2014.

Tangcharoensathien, V., Boonperm, J., & Jongudomsuk, P. (2010) 'Sharing health data', *Bulletin of the World Health Organization*, 88: 468-469.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., Frame, M. (2011) 'Data sharing by scientists: practices and perceptions', *PloS One*, June 29: DOI 10.1371/journal.pone.0021101.

Tierney, W., Nyandiko, W, Siika, A., Wools-Kaloustian, K., Didle, J. E., Kiplagat, J., Bell, A., Inui, T. S. (2013) "'These are good problems to have ...': establishing a collaborative

research partnership in East Africa', *Journal of General and Internal Medicine*, 28: 625 – 628.

Tindana, P., Molyneux, C. S., Bull, S. and Parker, M. (2014) 'Ethical issues in the export, storage and reuse of human biological samples in biomedical research: perspectives of key stakeholders in Ghana and Kenya', *BMC Medical Ethics*, 15:76 – 86.

Torjesen, I. (2014) 'Access to some NICE approved drugs varies hugely across England', *BMJ: British Medical Journal*, 348.

Vélez, O., Okyere, P.B., Kanter, A.S. and Bakken, S., (2014) 'A usability study of a mobile health application for rural Ghanaian midwives' *Journal of Midwifery & Women's Health*, 59(2):184-191.

UNESCO (2010) *UNESCO Science Report 2010. The Current Status of Science around the World*. Geneva: UNESCO.

United Nations (2015) *Taking Stock of the Global Partnership for Development*. Geneva: UN.

US National Committee for CODATA (1997) *Bits of Power: Issues in Global Access to Scientific Data*. Washington DC: National Academies Press.

Wacjman, J. (1991) *Feminism Confronts Technology*. Cambridge: Polity Press.

Walport, M., & Brest, P. (2011) 'Sharing research data to improve public health', *The Lancet*, 377: 537-539.

Wellstood, K., Wilson, K. Eyles, J. (2006) 'Reasonable access to primary care', *Health & Place*, 12(2): 121-130.

Whyte, A., and Pryor, G. (2011) 'Open science in practice: researcher perspectives and participation' *International Journal of Digital Curation* 6, no. 1: 199-213.

Wilson, E. J. (2000) *Closing the Digital Divide: An Initial Review*. Briefing the President. Washington: The Internet Policy Institute. <http://www.internetpolicy.org/briefing/ErnestWilson0700.html>, accessed 03/03/2014.

Ynalvez, M., Duque, R. B., Mbatia, P., Sooryamoorthy, R., Palackal, A., & Shrum, W. (2005) 'When do scientists "adopt" the Internet? Dimensions of connectivity in developing areas.' *Scientometrics*, 63(1): 39-67.

Ynalvez, M. A., & Shrum, W. M. (2011) 'Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country.' *Research Policy*, 40 (2): 204-216.

Notes

¹ As pointed out by the anonymous reviewer of the article, an alternative reading of the “Digital Divide” criticism is that this literature treats access as a distinct state although it is actually a variable. This is an interesting alternative perspective that requires further investigation.

² Here we follow the lead of critical scholarship that has questioned the assumption that, regardless of context, ICT infrastructures facilitate scientific collaboration and thereby enhance research productivity (e.g. Duque et al. 2009; Ynalvez et al. 2009; Ynalvez and Shrum 2011). Rather than “much needed ‘elixir’ that will free Third World science from its relative isolation and integrate it successfully into the global scientific community”, by following the utilization of online information this work has pointed to the ways in which ICT technologies can ‘pool’ in particular research environments, exacerbate socio-geographic inequalities and entrench the dependency of developing world scientists on powerful Northern counterparts (Duque et al 2005: 757).

³ <http://pantonprinciples.org/#sthash.iPPTcuI0.dpuf> (accessed 11/08/2014).

⁴ <http://data.worldbank.org/indicator/IT.NET.BBND.P2/countries?display=map> (accessed 11/07/2014).

⁵ <http://twas-old.ictp.it/links/open-access-scientific-information> (accessed 11/07/2014).

⁶ Indeed, most funding agencies do not fund core running costs, and assume a level of support from the institutions receiving the grants in terms of facility maintenance and upkeep.