Council for the Development of Social Science Research in Africa
Conseil pour le développement de la recherche en sciences sociales en Afrique
Conselho para o Desenvolvimento da Pesquisa em Ciências Sociais em África
مجلس تنمية للبحوث الإجتماعية في أفريقيا

CODESRIA

Fourth CODESRIA Conference on Electronic Publishing

Quatrième conférence du CODESRIA sur l'édition électronique

The Open Access Movement and the Future of Africa's Knowledge Economy

Le mouvement pour le libre accès et l'avenir de l'économie africaine du savoir

Dakar, Senegal, March 30 - April 1, 2016

Dakar, Sénégal, 30 mars - 01 avril 2016

What Hinders Data Sharing in African Science?

Dr. Louise Bezuidenhout & Professor Brian Rappert

CODESRIA
Avenue Cheikh Anta Diop X Canal IV
BP 3304, CP 18524, Dakar, Senegal
Phone: (+221) 33 825 98 22 / (+221) 33 825 98 23
Fax: (+221) 33 824 12 89
http://codesria.org/
https://www.facebook.com/CODESRIA-181817969495/
https://twitter.com/codesria

open.access@codesria.sn

Abstract

The Open Science movement has gained considerable standing in recent years. As part of this, activities such as data dissemination and its re-use are becoming positioned as integral aspects of research. Nonetheless, despite the enthusiasm for global access to online data, data sharing discussions recognize that many scientists in low/middle-income countries have struggled to make full use of online resources due to financial and technological constraints. Despite the unquestionable value of efforts to alleviate such conditions, this paper contends that they are not sufficient in themselves to counter asymmetries between the data practices between scientists in high- and low/middle-resource settings. This argument is based on embedded fieldwork in four (bio)chemistry laboratories in sub-Saharan Africa. Our research identified a range of systemic issues present in low-resourced research environments that hinder scientists from engaging with online resources – both as producers and users of data. On the basis of this identification, we suggest complementary strategies based on addressing the day-to day demands of undertaking research to ensure that scientists in low resource settings are not unintentionally marginalized from Open Science.

In the last decade, computing has revolutionized scientific research. Within the life sciences, for instance, advances in processing have ushered in the '-omic age' of genomics, proteomics, and metabolomics; while fields such as epidemiology, genetics are able to generate and analyze data sets of ever increasing size. 'Big data' has become a by-word for cutting edge science.

The recognition of both the exponentially increasing amount of information available and its immense potential for re-use has reinvigorated long standing discussions about openness and data sharing. 'Open Science' has served as prominent label for efforts commonly associated with the suggestion in the Panton Principles that: '[f]or science to effectively function … it is crucial that science data be made open'.[1] The Open Science movement represents a range of different initiatives that have been crucial in facilitating openness in scientific research. The Open Access and Open Data movements have been pre-eminent aspects of Open Science, addressing issues relating to access to published articles and data respectively.

The potential and challenge of Open Science is evident in calls made as part of it that '[f]ull and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research' (CODATA 1997: 10). There are thus growing expectations that research data will be, as the Wellcome Trust put it: 'managed and used in ways that maximize public benefit.'[2] Indeed, key pronouncements such as the Principle of Universality of Science (Statute 5 of the International Council for Science), require: 'freedom... of communication for scientists, as well as equitable access to data, information and other resources for research'.[3] Towards this end, many of the efforts to improve openness in research have gone into ameliorating barriers to ensure that data and published articles are readily available (Tenopir et al. 2011). Amongst these achievements have been the re-negotiation of financial paywalls of for-profit research publishers and the restrictions on data reuse by publishers and producers (Molloy 2011).

Discussions about openness and data in particular have focused on the universal importance of data sharing, the responsibility of data producers to facilitate access to resources, and the need for international standards to oversee the various aspects of data pathways. Pledges towards data openness (such as CODATA 1997) and international codes of conduct for data producers (Knoppers et al. 2011; 2014) are examples of this trend. Indeed, it may be suggested that the Open Data initiatives are frequently premised on the idea that data that is made available *will* be re-used, and that this re-use *will* benefit humanity. Efforts have focused on increasing the amount and availability of online data through increasing the number of accessible institutional repositories, databases, and sharing platforms.

In Open Data, while it is commonly recognized that not all data can be made immediately available, there is a frequent expectation that scientists are expected to 'maximize the availability of research data with as few restrictions as possible.'[4] As an example, BioMed Central (BMC: a for-profit scientific publisher specializing in OA) issued the following statement: '[s]ubmission of a manuscript to a BioMed Central journal implies that readily reproducible materials described in the manuscript, including all relevant raw data, will be freely available to any scientist wishing to use them for non-commercial purposes. Well established and widely supported databases exist for certain types of data such as nucleic acid sequences, protein sequences, and atomic coordinates; information on which can be found below and in journal instructions for authors and 'about' pages. An increasing number of research funding agencies also now support data sharing in the life sciences.'[5] Scientists are thus placed under broad obligations to 'provide links in a consistent place within an article to supporting data - regardless of the location or format of the data - and to make it clear to readers when they can also access the data as well as the article.'[6]

Moreover, 'data sharing' is routinely associated with the *provision of open online resources* – in other words, sharing should extend far beyond an individual researcher's formal collaborations or informal networks. For example, Biomed Central states that: '[b]y open data we mean that it is freely available on the public internet permitting any user to download, copy, analyze, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. We encourage the use of fully open formats wherever possible.'[7]

Challenges to Openness

The Open Science movement, and the policies that are starting to define data interactions, thus focus on the notion of science as a universal good that should be readily accessible. While the Open Science movement has met with widespread enthusiasm, many challenges continue to exist. How to adequately address issues of ownership and credit, privacy and harm are some of commonly recognized ones.

In relation to the themes of this paper, we wish to flag two issues in particular that raise challenges for the aspirations voiced today: the conditions of 'developing countries' as well as the underlying concepts of using data.

Developing Countries

With regard to the first, Open Access and Open Data debates routinely acknowledge that financial constraints mean that researchers in some countries will be more pressed than others to make their research open. This focus has had significant influence on initiatives designed to assist scientists in what are often referred to as developing countries. The Open Access community has been highly proactive in addressing the need to '[remove] price barriers (subscriptions, licensing fees, pay-per-view fees) and permission barriers (mostly copyright and licensing restrictions)' that potentially hamper scientists' ability to engage with online published resources (Suber 2014). Commitments to assist researchers have also been made by many funding bodies. The Global Research Council, for instance, stated 'funds to cover expenses for open access publications might not be readily available. However, in transitioning to open access undue publication barriers must be avoided. It will be necessary to look for solutions that assist those authors in openly sharing their research results and thus making impact' (COAR 2013: 14). With regard to Open Data, attention has been drawn to improving the skills of and infrastructure available to researchers in developing countries so that they can fulfill the universal responsibility towards openness (CODATA 2014; International Council for Science 2015).

Compounding the recognition of difficulties is a lack of social research into the factors that enable and inhibit data sharing in resource constrained laboratories (Carr and Littler 2015: 315). There is little systematic understanding of how Open Data may be understood differently across research contexts, as well as a lack of empirical investigation into how data sharing is understood within low-resourced environments. Further complicating this matter is the observation that the existing literature on data sharing in developing countries focuses on comparatively well-funded and well-connected research networks or consortia (de Vries et a. 2011; Parker and Bull 2015; de Vries et al. 2015) that deal predominantly with clinical research (Bull et al. 2015; Pisani and Abou-Zhar 2010). As such, discussions about both the motivations to share data and the ways in which data are shared are tied to specific understandings of resource distribution, infrastructure provision, and governmental involvement.

In light of the increasing amount of non-clinical research occurring in regions such as Sub-Saharan Africa, it is becoming vital that data sharing practices be more broadly examined, and related to questions about priorities and resources of science, what science is, and where science should lead. Scientists in this region often confront a variety of issues related to their research environment, including lack of open data sharing traditions, governments treating publicly generated or publicly funded research data as secret or as a commercial commodity, and lack of local data centres or digital repositories, all of which present potentially problematic barriers to effective data management.

Thus it becomes important to question whether the value attached to data sharing is informed by Western presumptions that underplay contextual issues associated with research in relatively resource-constrained settings because they are not recognized problems within many Western contexts.

Conceptualizations of Data

Following on from these points, much of the Open Data discourse maintains a 'universalist' perspective, highlighting the importance of global sharing and access to data online. In effect, the Open Data discourse makes use of a strong causal link between the availability of data, their re-use by the global scientific community, and the benefits that humankind can accrue as a result of this re-use. Thus we have asked whether it is sufficient to simply provide online resources or whether there is more that needs to be considered.

As part of this it is worth noting that the vast majority of discussions related to Open Data focus exclusively on the act of 'data sharing'. Indeed, within these discussions data sharing is often taken to be synonymous to ensuring openness of data and research. In practice though it is worth keeping in mind that in practice data sharing refers to a range of activities including the 'collection, analysis, publication, reanalysis, critique and reuse' (Molloy 2011: 1) of data. Moreover, what constitutes data now can include: 'writings, films, sound recordings, pictorial reproductions, drawings, designs or other graphic representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing or computer programs (software), statistical records and other research data' (NIH 2006: section 8.2.1). Thus it becomes evident that the requirements for and the pathways to openness are manifold. Beyond the traditional avenues of journal publication and database contributions these could include: '… innovative sharing platforms such as personal websites, e-books, discussion forums, email lists, blogs, wikis, videos, audio files, RSS feeds and P2P file-sharing networks' (Suber 2014), as well as novel social networking forums such as Figshare.

Research

In the absence of much relevant research on data engagement, as part of a Leverhulme Trust funded project the authors and other project collaborators decided to investigate the day-to-day experiences of those undertaking research in low-resourced environments.[8] We conducted qualitative fieldwork consisting of extended visits (three to six weeks) to four bio/chemistry laboratories in sub-Saharan Africa. That research combined semi-structured interviews with staff and post-graduate students (n. 56) with laboratory observations and other activities (e.g., attendance at department meetings). Interviewees were asked to discuss their data generation, storage, curation, dissemination and re-use activities within the context of their work and wider professional development.

The depth of extended observational research was opted for in our research design in order to allow for the collection of detailed descriptions of scientists' research environments. Without these descriptions, we reasoned, it may be difficult for those outside of highly resourced constrained environments to envision the factors at in play resource constrained research contexts and their influence on data engagement activities. The decision about what labs would be visited was made according to the following selection reasoning:

> *Continent*: We decided to focus on Africa because experiences from researchers from this continent are largely missing from Open Science conversations (with the exceptions

noted above), and yet, it is the continent most often identified as requiring additional support to take advantage of open science.

Country: South Africa and Kenya were chosen because they have relatively robust national research activities.

University/department: Examples of vibrant, 'homegrown' research labs were sought. In that respect, all the labs visited undertook and published research internationally recognized. While some of the scientists collaborated with others in Europe and North America as part of funding grants, none of the sites were formally affiliated to large internationally research consortia or networks.

Specialization: To reduce differences related to areas of study, all four laboratories conducted chemistry/biochemistry research and had an interest in medicinal chemistry.

Thus the fieldsites represent productive and relatively well-resourced laboratories in Africa, but not ones that benefited from being members of international consortium or other formal collaborations that provided stable and ample sources of research funding.

In conducting this research we sought to address a number of inter-related questions, including:
- How do researchers convert online resources into useful utilities for their research?
- What roles do research environments play in scientists' engagement with data and involvement in Open Science, particularly with regard to Open Data?
- How are current Open Science initiatives both succeeding and failing to meet the objectives of increased openness and heightened research capacity?

Findings

In-depth analysis of the fieldwork has been given elsewhere (AUTHORS *under review*; AUTHORS *under review*). For the purposes of this paper, we wish to focus on one of the central recurring themes of the interviews: namely the day-to-day difficulties of doing research that shaped what science could be done; frustrated interviewees' ability to collect, analyze, publish, assess and reuse data; and downgraded the self-relevance they perceived in calls for data openness. These day-to-day demands included:

- Physical Infrastructure:
    - Limited connectivity – Low bandwidth and varying wifi signal were regularly identified as daily challenges to working online. This resulted in tasks that might otherwise be straightforward in more ideal situation (e.g., uploading data to an on-line platform, conference registration) proving taxing in practice;
    - Personal internet provision – Many staff and nearly all the research students incurred burdensome costs associated with the self-purchasing of data bundles, ICT equipment, and software updates;[9]
    - Off campus access – Three of the four universities visited did not have proxy servers and thus researchers could not make use of library facilities off campus;
    - Disruption – Power outages, service delivery and border controls all negatively affected the speed at which research was conducted.

- Funding
    - Base funding – Core funding to maintain and upgrade physical facilities and service equipment was limited to the most basic provisions (especially in Kenya);
    - Constraints: Staff research funding was largely (in South Africa) and wholly (in the Kenya) supplied by organizations other than the university or government. Project specific funding by (international) external bodies came with budgetary strings though which meant staff were curtailed in how they could spend it. This point and the previous one meant that vital tasks could not be funded. For instance, even if possessed, funding limitations meant that laboratory equipment could not be repaired or serviced such that it was in functioning order;
    - Procedures for procuring and reimbursing resources were slow and cumbersome.

- Support:
    - The ability to undertake research was diminished because of other substantial commitments (e.g. teaching loads);
    - The lack of hardware or software ICT support and other technical support meant researchers had to outsource such support, with resulting financial and time burdens;
    - Training: The sites visited lack training in data management. This situation was exacerbated by the absence of research dedicated staff (there was only one dedicated research post-doc in any of the labs visited) which in turn resulted in a lack of data mentorship to junior colleagues;
    - Institutional policies such as data sharing guidelines were not in place.

So, while all the researchers interviewed had access to a computer, internet and other ICTs, when discussing data engagement it became quickly apparent that factors existed within their working environment that significantly shaped their research activities and contributed to a state of sustained lowered ability to engage with data – online or otherwise.

It was hardly surprising then that those interviewed consistently voiced a stark distinction. Across all four sites, interviewees distinguished between the theoretical value of openness and the practical involvement in these activities. When discussing sharing and openness from a theoretical perspective, it was widely agreed that Open Data was an important development in modern science. Nonetheless, when asked to position *themselves* within the Open Data paradigm, interviewees responded in a markedly different manner. Despite their recognition of the theoretical benefits of the emerging Open Data movement, when talking about their daily research activities participants repeatedly highlighted their marginalization from the benefits of the movement. Openness might be laudable in the abstract, but through speaking to their positions within highly stratified global fields of science, those we interviewed repeatedly spoke of its perils. Interviewees accounted for their lack of involvement in sharing data by citing various considerations. The majority of these involved time – the lack of time to conduct research, the time needed to upload data in infrastructure-challenged environments, the time taken to produce data and the need to control it, and the difficulties of sharing data in environments with no support structures.

While our fieldwork was limited to four specific sites, the day-to-day conditions noted above are likely to be shared by other scientists in relatively resource-constrained environments in Africa.

In helping to understand the wider relevance of these empirical arguments, we have drawn on the on Amartya Sen's (1999) Capabilities Approach (CA) to poverty. CA offers a shift away from traditional ways of measuring inequality that focus on 'having or not having' a given good. Instead, it suggests that: '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). Importantly, the CA recognizes that individuals differ in their ability to convert resources into valuable opportunities (capabilities) or outcomes (functionings).

When applied to data engagement discussions, by characterizing data sharing in this way CA helps make it apparent that not only can the provision of online resources not automatically lead to data sharing utilities, but that the research environment plays an important role in the realization of utilities. In this CA offers a basic reframing. The question is not *What online resources are available?*, but instead *Can scientists effectively utilize these resources to realize their research goals?* The listing of day-to-day demands above describes some of the many factors prevalent in resource-constrained environments that frustrate turning available data in research utilities.

The use of CA to analyze data engagement activities in low-resourced research environments thus suggests how situations of data poverty can easily persist regardless of increasingly open online resources. Moreover, it cautions that initiatives to promote data openness and engagement in the conditions that characterize research environments in many developing countries are likely to be of limited effect if they simply aim to get researchers online or proselytize the benefits of being open. It is possible that without meaningful and contextually sensitive discussion as to how Open Science aspirations can be applied in low-resource research environments that the data openness requirements coming down via funding, collaboration and publication agreements will lead to scientists to dissociate themselves from openness initiatives. Instead, efforts must be direct toward tackling those considerations that result in *lowered access* in practice.

What Needs to be Done?
In this final section we want to turn to what actions might help alleviate the factors that frustrate data openness. To start with, our research would suggest that it is of considerable importance that Open Science discussions recognize the link between data engagement and research environments and the heterogeneity of these workspaces. Toward this acknowledgement, more descriptive information – both qualitative and quantitative – is needed on the diverse research environments around the world so as to better inform such policies. This needs to include research into developing country laboratories that are not in research networks or recipients of large-scale foreign grants.

In addition, we also strongly advocate for more discussion between scientists in developing and developed countries about the challenges experienced in data engagement. We believe that scientists in developing countries who face highly resource constrained conditions need to be better drawn into data engagement discussions and encouraged to discuss not only their current problems, but their opinions on possible solutions. Without this, it is thus a distinct possibility that the enthusiasm for making online resources available will continue to unintentionally

marginalize scientists working in low-resourced environments as they do not directly address how research poverty can exist in conditions of data surplus.

Also of considerable importance is the need to assist scientists to alter the factors in their environment that impede research and data engagement. As many of these factors are as innocuous as they are insidious, they are often poorly addressed through grants and international aid. Nonetheless, evidence from well-funded collaborations in Africa suggests that directed attention to these issues is invaluable in building research capacity in low-resourced regions.

For instance, the MalariaGen network involves researchers from twenty-one countries who collaborate to build a malaria genome database. As discussed by Lang (2011: 715): 'the project has addressed some of the fundamental challenges that are inherent to data sharing, such as the ethical challenges of recruiting participants and setting out clear agreements on data linking and release'. More specifically relating to data engagement, MalariaGen has implemented policies and schemes that speak to a number of the functional capabilities involved in data engagement. These include training and assistance initiatives developed to build data expertise within the research community. Moreover, the network has developed software specifically for use in Africa that includes a range of web interfaces and user-friendly tools to facilitate downloading and analyzing genotyping and sequence data. All these initiatives are supported by a highly engaged and supportive network with a group identity and united research objectives. Participants receive considerable academic and social support from network members and network policies that guide their data engagement activities. As a result, the MalariaGen network - through continual dialogue with community members (as well as considerable investment of resources) - has been able to address a number of factors impeded research and data management that may exist within some of their member sites.

Micro-Credit: A Viable Way of Moving Toward more Openness?
Of course, the support provisions of the MalariaGen network are exceptional by international standards and thus not widely reproducible given the investment demands entailed. As a more modest but much more widely applicable initiative, we propose that a system of *microcredit* that would allow scientists to tailor their research environment and remove data engagement barriers.

To expand, as highlighted by so many of our interviewees, the very innocuousness of the factors frustrating research also meant that they were the unlikely to be addressed through official grants or funding schemes. While absolute funding levels were no doubt important, the issues at stake went far beyond this. Resources were sometimes available, but could not be used in a manner aligned with the necessities faced. Those interviewed highlighted a set of financial obstacles relating to the minor amounts of money – usually well below $100 – required for such matters as professional membership, off campus access to internet and papers, the provision of necessary ICT hardware and software, and so forth that played a crucial role in how they understood their research and the benefits to be accrued therefrom. Not being able to fix an NMR machine, obtain liquid nitrogen, rewire the plugs in the lab or get buy-out from teaching to write a grant all become rate limiting steps in the production of data and, consequentially the sharing of data and the acquisition of credibility.

What is further evident from our fieldwork was the reported difficulties that scientists have to ameliorate these frustrations. The lack of core infrastructure funding for the laboratories, the difficulty of securing (often foreign) project grants, and the strict limits on which grant funding can be used all contributed towards these perceptions of lack of capabilities and agency.

Thus, addressing the disjunction between research and publication is more complicated than simply an absolute lack of research funds – it is also about *what the funds can be used for as well as how the funds are spent.* It thus becomes apparent that one way of promoting data engagement by scientists in Africa working in resource-constrained environments is by *enabling them to shape their research strategies in a manner that makes strategic sense to themselves with regards to data interactions.*

In seeking to ask how to redress such varied concerns, we take inspiration from efforts in recent decades to promote forms of 'micro-credit' for those with inadequate access to financial services (Barry 2012). Microcredit loans offer funding – typically small amounts – to those starting up new activities, expanding existing ones, or seeking to overcome difficulties that develop. Micro-credit differs from charity because: 'rather than assuming that poverty is the result of personal failings on the part of the poor, micro-credit [agencies] believe that poverty is created through social processes that deprive the poor of their rightful access to social resources, including monetary credit' (Elahi 2006: 477). Similarly, we suggest that extending types of micro-credit support to scientists could well be valuable in allowing them to address the often-innocuous aspects of their research environments that they understand as hampering their practice. Providing a flexible and responsive support that could take the form of research grants or loans would be of significant value to the scientists we interviewed.

Bayulgen (2008) elaborates a number of wider benefits from microcredit. For instance, through enabling individuals or groups to make choices for themselves about their future directions, micro-credit financing can foster self-confidence, self-reliance, and self-esteem. They can also promote shared senses of identity. These were matters of concern in our interviews; scientists repeatedly expressed perceptions of low (self-)esteem due to their geographic and professional positions. Finally, when micro-credit scheme promote interaction within communities they offer the possibility of building trust between participants as well as extending forms of networking and co-ordination. These, in turn, can foster future coordinated and cooperative interactions. We think this would be valuable in the case of the scientists examined because of the limited networking and sharing, in no small part due to the lack of trust in those whom were not directly personally known.

To continue development literature parallels, enabling scientists to shape their research environments through micro-credit loans would also be an important way of minimizing the influence of the dominant Western perspective that currently exists in most Open Science discourse. Allowing African scientists working in resource demanding situation to define what they need, instead of prescribing what they should do would not only strengthen capacity in these regions, but also potentially minimize any dependency relations (Somers and Block 2005).

The in-practice advantages of microcredit schemes depend on their particulars and how they are conducted. On these points, both the diversity and potential regressive dimensions of

microcredit have been topics of long standing commentary (e.g. Fernando 1997). Our proposing the potential of microcredit forms of support for science is not based on a whole scale endorsement of microcredit. Instead, we wish to contend such forms of support offer the prospect for engaging with the day-to-day challenges faced by scientists.

References

Avgerou, C. 2008, 'Information Systems in Developing Countries: A Critical Research Review', *Journal of Information Technologies*, 23, pp. 133 – 146.

Barry, J. 2012, 'Microfinance, the Market and Political Development in the Internet Age', *Third World Quarterly*, 33, 1, pp. 125 – 141.

Bayulgen, O. 2008, 'Muhammad Yunus, Grameen Bank and the Nobel Peace Prize', *International Studies Review*, 10, 3, pp. 525 – 547.

Bull, S., Cheah, P. Y., Denny, S., Jao, I., Marsh, V., Merson, L., Shah More, N., Nhan, L. N. T., Osrin, D., Tangseefa, D., Wassenaar, D., Parker, M. 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, pp. 302 – 313.

Carr, D., and Littler, K. 2015, 'Sharing Research Data to Improve Public Health', *Journal of Empirical Research on Human Research Ethics*, 10, 3, pp. 314 – 316.

COAR (Confederation of Open Access Repositories) 2015. *Promoting Open Knowledge and Open Science Report of the Current State of Repositories* 21 May Göttingen: COAR.

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

CODATA. 2014, *Data Sharing Principles in Developing Countries* (The Nairobi Data Sharing Principles) Nairobi: CODATA https://zenodo.org/record/22117/files/Data_Sharing_Principles_in_Developing_Countries.pdf (accessed 04/02/2016).

De Vries, J., Bull, S. J., Doumbo, O., Ibrahim, M., Mercereau-Puijalon, O., Kwiatkowski, D., Parker, M. 2011, 'Ethical issues in human genomics research in developing countries', *BMC Medical Ethics*, 12, pp. 5 – 15.

De Vries, J., Tindana, P., Littler, K., Ramsay, M., Rotimi, C., Abayomi, A., Mulder, N., Mayosi, B.M. 2015, 'The H3Africa policy framework: negotiating fairness in genomics', *Trends in Genetics*, 31, 3, pp. 117 – 119.

DiMaggio, P and Hargittai, E. 2001, *From the "Digital Divide" to "Digital Inequality": Studying Internet Use as Penetration Increases.* Working Paper Series no.15 Princeton: Centre for Arts and Cultural Policy Studies.

Elahi, K.Q. and Rahman, M.L. 2006, 'Micro-credit and micro-finance', *Development in Practice,* 16, 5, pp. 476 – 483.

Fernando, J.L. 1997, 'Nongovernmental organizations, micro-credit, and empowerment of women', *Annals of the American Academy of Political and Social Science,* 554, pp. 150-177.

International Council for Science. 2015, *WDS Data Sharing Principles*. https://www.icsu-wds.org/files/WDS_Data_Sharing_Principles_2015.pdf (accessed 04/02/2016).

Knoppers, B. M., Harris, J. R., Tasse, A. M., Budin-Ljosne, I., Kaye, J., Seschenes, M., Zawati, M. H. 2011, 'Towards a data sharing code of conduct for international genomic research', *Genome Medicine*, 3, pp. 46.

Knoppers, B. M., Harris, J. R., Budin-Ljosne, I., Dove, E. S. 2014, 'A human rights approach to an international code of conduct for genomic and clinical data sharing', *Human Genetics*, 133, pp. 895 - 903.

Lang, T. 2011, 'Advancing global health research through digital technology and sharing data' *Science*, 331, p. 714.

Molloy, J. C. 2011, 'The Open Knowledge Foundation: open data means better science', *PLoS Biology,* 2, 12, e1001195.

NIH. 2006. *Grants Policy Statement,* Bethesda: NIH
 http://grants.nih.gov/grants/policy/nihgps_2013/nihgps_ch8.htm#_Toc271264947
(accessed 04/02/2016).

Parker, M., 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 – 224.

Pisani, E., and Abou-Zahr, C. 2010, 'Sharing health data: good intentions are not enough', *Bulletin of the World Health Organization*, 88, pp. 462 – 466.

Robeyns, I. 2011, 'The Capability Approach' in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy,* http://plato.stanford.edu/

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

Somers, M. R., and Block, F. 2005, 'From poverty to perversity: ideas, markets, and institutions over 200 years of welfare debate', *American Sociological Review*, 70, 2, pp. 260-287.

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

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*, 6, 6, pp. e21101.

Notes

[1] See www.pantonprinciples.org (accessed 04/02/2016).

[2] http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm

[3] http://www.icsu.org/general-assembly/news/ICSU%20Report%20on%20Open%20Access.pdf (Accessed 22/09/2014).

[4] http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm (accessed 04/02/2016).

[5] http://www.biomedcentral.com/about/supportingdata (accessed 04/02/2016).

[6] http://www.biomedcentral.com/about/supportingdata (accessed 04/02/2016).

[7] http://www.biomedcentral.com/openaccess/ (accessed 04/02/2016).

[8] The research informing this article was supported by a grant from the Leverhulme Trust title 'Beyond the Digital Divide' (RPG-2013-153). Our thanks to Ann Kelly and Sabina Leonelli for their contributions as other team members.

[9] For a discussion of how the level to which an individual has access to ICT equipment – including permission rights on software and user control on individual machines – all affect engagement with online data, see DiMaggio and Hargittai (2001). Furthermore, pragmatic issues such low numbers of personal computers in laboratories (and thus sharing of computer time and the autonomy of computer use), the positioning of computers in work areas (Avgerou 2008), how long they are able to spend on the Internet during the working day, and so forth all influence whether they are able to truly engage with the available data.