

*Editorial to Special Collection “Open Data and Africa” published in Data Science Journal in February 2018*

## **Introduction: Open Data and Africa**

Sabina Leonelli, University of Exeter, [s.leonelli@exeter.ac.uk](mailto:s.leonelli@exeter.ac.uk)

Brian Rappert, University of Exeter

Louise Bezuidenhout, University of Oxford

Open Data is widely regarded as the greatest challenge in the pursuit of openness in science, given the vastly different data practices – and related ethos and cultures of ownership, curation, storage and dissemination – which characterise each area of research. Initiatives such as OpenAIRE, Elixir and, most recently, the European Open Science Cloud are investing some of their resources in assessing the extent of the differences among disciplinary approaches to data sharing and re-use, and identifying the standards and related infrastructures that can foster communication and exchanges across fields while also respecting their diverse methodological traditions.<sup>1</sup> These efforts are crucial to making sure that data are reliable, appropriately curated and useful to future research – and thus, that it makes sense to invest time and effort in making them widely accessible in the first place.

However, the current focus on diversity in methods and subject matter is taking attention away from another important type of diversity within science: the variation in research environments around the globe. Depending on the country and type of institutions in which they work, researchers can be confronted with significantly different research conditions, ranging from very high-resourced environments guaranteeing access to the latest equipment, reagents and

---

<sup>1</sup> See the related websites at [www.openaire.eu](http://www.openaire.eu), [www.elixir-europe.org](http://www.elixir-europe.org) and [eosc-pilot.eu](http://eosc-pilot.eu); also Leonelli (2017) for an overview of these initiatives and their impact.

computational tools, to low-resourced environments with intermittent access to a broadband connection and inexpensive or outdated instrumentation. Such variation does not necessarily hamper the excellence of the research being conducted, but it does affect the choice of research goals and collaborators at different locations, and the ways in which outputs are disseminated. The diversity of research environments is particularly visible in the African continent, where excellent research is carried out in a wide variety of settings, but some of those settings are characterized by limited or no access to highly expensive equipment.

This collection aims to shed some light on what the goals and aspirations associated with Open Data means in Africa today: what opportunities they offer, what challenges they pose and what the implications follow from the increasing political and institutional support for this concept. The collection is by no means comprehensive and touches only on specific issues and cases, yet we hope it constitutes a step towards an improved understanding of diversity in research environments as a key component of implementing Open Science. In this sense, this collection reflects the spirit of initiatives such as the African Open Science Platform, which seek to foster openness in African science while also highlighting the distinctive challenges and goals of researchers working in this continent.<sup>2</sup> It explores *how*, *why*, and *to what end* scientists working in African research environments share and re-use data, and the extent to which these activities relate to the priorities, practices, and policies of data management and scientific discovery that are associated with research elsewhere. Our goals are to document the significant impact – positive or negative - that a shift towards openness can have in the African context, and to underscore the need for Open Science policies and infrastructures to learn from diverse research conditions, resources and goals around the world.

In their paper on weather forecasting in Uganda, Shuaib [Lwasa, Ambrose Buyinza and Benon Nabaasa](#) elaborate the challenges of constructing big data models in

---

<sup>2</sup> For an overview of the remit and structure of the African Open Science Platform initiative, see: <https://www.slideshare.net/AfricanOpenSciencePlatform/african-open-science-platform-77971130> (accessed 2 February 2018).

countries where public and civil organizations continue to hoard data. They suggest creative ways through which this challenge can be addressed using multiple sources of data to address the challenges facing pastoral communities in light of climate change. In a related article, [Josiline Chigwada, Blessing Chiparausha and Justice Kasiroori](#) draw attention to the lack of research data management in many African institutions. Using empirical data gathered in Zimbabwe they suggest that the majority of researchers continue to develop their own personal data management plans. Lack of guidelines on good practice, together with inadequate human resources, technological obsolescence, insecure infrastructure, use of different vocabulary between librarians and researchers, inadequate financial resources, absence of research data management policies and lack of support by institutional authorities and researchers negatively impact on research data management. The authors advocate for the establishment of research data repositories as well as overarching guidance and oversight to foster responsible data management. Tewodaj Mogues and Leonardo Caceres further contribute to this theme by investigating the “black box” of public expenditure data in Mozambique, focusing specifically on efforts and strategies to quantify agricultural spending on the basis of publicly available data.

The papers by Bezuidenhout and Leonelli both highlight the need for more nuanced discussion relating to data sharing and Open Data against the diversity of research environments in Africa. [Louise Bezuidenhout](#) draws attention to the critical interrelation between the availability of laboratory technologies and scientists’ perceptions of data sharing. She suggests that an expanded understanding of laboratory equipment and research speed will be important when advocating for data sharing amongst researchers. [Sabina Leonelli](#) focuses on data quality standards. She identifies an unequal power relation in the setting of standards for what counts as ‘good science’ worldwide, and suggests this can make researchers based in the Global South resistant to sharing data and/or describing their provenance and methods. To counter this, Leonelli advocates that debates around Open Data need to include critical reflection on the criteria used to evaluate data quality, and the extent to which that evaluation requires a localised assessment of the needs, means and goals of each research environment.

Nicola Mulder and her eighteen co-authors provide a wide-ranging perspective on the challenges of sharing genomic data gathered in African countries. They highlight how African researchers continue to work mainly on study recruitment, determination of phenotypes and collection of biological samples end of the genomic research spectrum, rather than contributing to the generation of genomic data. This leads to a concerning separation between data sharing practices as designed and implemented by non-African collaborators and the evaluation of what constitutes adequate safeguards for primary data generators based in Africa. As an alternative, Mulder and colleagues discuss recent initiatives such as H3Africa and H3BioNet as examples of capacity building in large-scale genomics projects in Africa where ethical data sharing has been prioritized.

Finally, [Brian Rappert](#) draws attention to the limitations of current funding structures in addressing the low-resourced nature of many African research institutions. The lack of flexible funding to improve infrastructures and daily research environments can inhibit research efficiency and affect data sharing practices. Rappert offers a model of 'micro-funding' that seeks to address the day-to-day demands in low-resourced environments and offer a new approach to promoting data sharing.

Overall, this collection provides insights into the diversity of practices, requirements and working conditions of researchers located in different institutional settings and various parts of the African continent. The last decade has witnessed some international initiatives to highlight concerns and needs from African researchers, such as the African Open Science Platform and a recent report by the Global Young Academy that documents access to research software in Sub-Saharan countries (Vermeir et al 2018). Much more can and should be done. Collaboration and feedback across African countries is urgently needed. Empirical research is required to document how data can be made findable, accessible, interoperable and re-usable in these settings, thus following the FAIR principles for the effective management of data (Wilkinson et al 2016). Perhaps most notably, encouraging data re-use is not necessarily the same as making all data freely and widely available. Well-informed ethical, legal and institutional considerations need to be attached to each choice to release data in an open format, and relevant training in Open Science practices and tools is required for researchers, research-performing institutions,

funding bodies and governmental agencies to fully embrace the opportunities and challenges of Open Science.

### **Acknowledgments**

We thank all the authors who contributed to this special collection, the referees who kindly lent their expertise to improving the papers, Sarah Callaghan for generously shepherding us through the editorial process at the *Data Science Journal* and the participants to the workshop “Pacing Science” held in Exeter in May 2016 for their feedback on this project, particularly Ann Kelly, Linsey McGoev and Simon Hodson for their insightful comments, support and help in co-organising the event. This work was funded by the Leverhulme Trust Research Grant “Beyond the Digital Divide” (award number RPG-2013-153) and the European Research Council Starting Grant “The Epistemology of Data-Intensive Science” (award number 335925).

### **References**

Leonelli, S. (2017) *Incentives and Rewards to Engage in Open Science Activities (Thematic Report No 3)*. Report for the Mutual Learning Exercise of the European Commission: Open Science – Altmetrics and Rewards.

<https://rio.jrc.ec.europa.eu/en/library/mutual-learning-exercise-open-science-%E2%80%93-altmetrics-and-rewards-incentives-and-rewards-engage>

Vermeir, K et al (forthcoming 2018) *Global Access to Research Software: The Forgotten Pillar of Open Science*. Report of the Global Young Academy.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018. DOI: 10.1038/sdata.2016.18