
Data Driving Sustainability—the African Open Science Platform Project

Ina Smith and Susan Veldsman

Introduction

- 1 Research across disciplines are increasingly driven by data, based on the enormous growth and capacity of digital systems to acquire, store, analyse and instantaneously communicate data and information, and to do so at continuously reducing cost. Collaboration underpins the sharing of data, but also the sharing of resources where there is resource scarcity. On a resource scarce continent such as Africa, collaboration and sharing of research and the underlying data are creating enormous potential for scientific discovery, and new and powerful ways for addressing issues such as agricultural productivity, disease tracking and control, and the production of bio-fuels.
- 2 An example of the need for collaboration is when there is an outbreak of a disease such as Ebola, where data can transform the response of scientists and health workers to these outbreaks. According to Yozwiak (2015) rapid sequencing, combined with new ways to collect clinical and epidemiological data, are critical to quickly respond to outbreaks. Also of importance is that data is well managed. When an outbreak happens, there is no time to establish whether patient consent was obtained, whether the data was collected in an ethical way, or whether there are gaps in the data.
- 3 In addition to outbreaks and diseases, open data can help achieve the SDGs by providing critical information on natural resources, government operations, public services, and population demographics (Gurin 2015). Insights into critical areas can inform government, continental and international priorities, and at the same time help determine the most effective paths for action on these issues.
- 4 The Science International Accord on Open Data in a Big Data World (Science International, 2016) presents an inclusive vision of the need for and the benefits of open

data for science internationally, and in particular for less affluent and least economically developed countries (LEDCs). In addition to benefiting from data from the international community, African countries have much to contribute in terms of FAIR (findable, accessible, interoperable, re-usable) data (Wilkinson, 2016) for all to benefit in making progress towards implementing the SDGs (United Nations, 2016). Although there is a huge global movement towards making data openly accessible, concerns remain—particularly among researchers from LEDCs, characterised by limited capacity and a deep mistrust, digital illiteracy, inadequate infrastructure, and minimal investment in data science training (Serwadda *et al.* 2018).

- 5 The African Open Science Platform—which is an outcome of the Science International Accord on Open Data in a Big Data World—is a major intervention towards creating an understanding of the importance of open data to achieve the SDGs, and to address the concerns experienced by researchers in terms of open data specifically.

Methodology

- 6 A literature review was conducted, supported by findings from the African Open Science Platform landscape study, webinars and presentations by various experts. Data was also collected during project interventions, including workshops in Ghana, Madagascar, Ethiopia and Botswana. These workshops were attended by African stakeholders across the continent. Existing and established conferences were used to further advocate and create an awareness for the need for data and research to be open, and meetings with governments were successful in kick-starting conversations around open science policies on national level.

About the African Open Science Platform

- 7 The African Open Science Platform (AOSP)—now in its second year—is managed by the Academy of Science of South Africa (ASSAf), through the International Council for Science Regional Office for Africa (ICSU ROA) being hosted by ASSAf. The project is funded through the SA National Research Foundation (SA Dept. of Science and Technology), with direction from the International Council for Science (ICSU) Committee on Data for Science and Technology (CODATA). Although it is funded by South Africa, it is expected to benefit the whole continent, not only accelerating data sharing from South Africa, but also from other interested African countries, at the same time leapfrogging and benefitting from what has been achieved by higher income countries in terms of data sharing. The project team has its office in Pretoria, South Africa, and the project is further governed by an Advisory Council, and a Technical Advisory Board, both guided by Terms of Reference. Members of both the Council and Board are from across the African continent.
- 8 The project aims to develop an open science and innovation dialogue platform in order to increase awareness, accessibility and visibility of African science and data, at the same time reflecting on progress made on the African continent in terms of the following areas:
 - Open science/data policy and strategy;
 - Open science/data information technology (ICT) infrastructure to support data sharing;

- Capacity building/training to support and apply open science/data when conducting research; and
 - Incentives for sharing science output and specifically the underlying data sets, in an open and transparent way.
- 9 Expected deliverables include frameworks and roadmaps for each of the focus areas (open science policy, ICT infrastructure, capacity building and incentives), as well as a landscape survey, and represented as an indexed database. In addition to providing an overview of data related initiatives happening on the African continent, it is expected that this database will serve as a resource to identify research partners, opportunities for collaboration, funding opportunities, research gaps, and many more.
- 10 Data remains an important part of the research lifecycle, and in addition to be shared, it needs to be managed. Good research data management practice allows reliable verification of results and permits new and innovative research to build on existing research. More and more, research councils/funders also require from researchers to create research data management plans, and to deposit research data in recognized or institutional data centres/repositories. The following funder initiatives are examples:
- The South African National Research Foundation (NRF) policy requires underlying data supporting the funded research output, to be made accessible on a trusted repository—sooner rather than later.
 - The Science Granting Councils Initiative (SGCI) in Sub-Saharan Africa aims to strengthen the capacities of 15 science granting councils in Sub-Saharan Africa in order to support research and evidence-based policies that will contribute to economic and social development.

Rationale behind Open Data

- 11 According to the FOSTER Project, open science—of which open data is one component—is: “the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.”
- 12 In order for collaboration to happen, data needs to be open, well managed (trusted) and accessible, so that other researchers can build on existing research. Some of the benefits of open data and the sharing of data are listed below:
- Without data as evidence, research is merely an opinion. Data should be linked to final research output, so that others can verify the outcome, where needed, applying academic rigor. In a world of fake news, fake research, fake journals, it is expected that fake data will also future more and more, potentially damaging the integrity of research. Therefore it should be well managed, as open as possible, and as closed necessary only.
 - Open data can help advance science: the more diverse the scientists involved in finding solutions to addressing societal problems, the more it will accelerate findings, towards innovation.
 - The more open data, the more it can be used to predict trends so that scientists and the community alike can make informed decisions.
 - Access to data—similar to access to completed peer-reviewed research—can help drive development, creating opportunities for people to become entrepreneurs, empowering themselves.
 - A data set can potentially have far more outcomes, beyond one single research article only.

- More outcomes can lead to increased impact, and thereby a higher return on investment.
 - Open data can be used to train new scientists (Mulder *et al.* 2017).
 - Open data can help increase national autonomy and equity, across the globe (Mulder *et al.* 2017).
- 13 To benefit from research data, African countries need to take ownership of it—especially where funded with public money. Data needs to be managed and metadata needs to be added to make sure it is findable, accessible, interoperable and re-usable—aligned with the FAIR principles of data sharing, supported globally. Each country has to oversee the ethical use of data from that country, and where researchers from outside the LEDCs collect data, the ideal would be for them to engage local researchers and communities to ensure that their science is not harmful to local communities and to avoid misinterpretation of results (Serwadda *et al.* 2018).
- 14 Quite a number of data intensive research activities are currently ongoing on the African continent. An example of a managed collaborative in the biomedical sciences is the Malaria project, in which Malaria-GEN, H3Africa, INDEPTH and SATuRN collaborate. H3ABioNet—the bioinformatics leg of H3Africa (Human Heredity and Health in Africa) — in turn consists of a network of 30 institutions, from 15 African countries, with 2 partners outside Africa (Mulder *et al.* 2017).
- 15 Bioinformatics is just one area which relies heavily on data and collaboration to advance discoveries towards eliminating diseases on the African continent. Mulder *et al.* (2017) describes Africa as currently undergoing an epidemiological transition, with endemic infections and a rapidly growing burden of cardiometabolic and other non-communicable diseases. Genomics research holds great promise for medical and health care research, and a greater understanding of underlying data or genomic variation can accelerate the pace towards catching up with outbreaks. Some of the outbreaks recorded by the Centers for Disease Control and Prevention includes the Ebola virus, Susan virus, and Salmonella.
- 16 The promise to find cures and to make progress can however only be fulfilled if data is well managed and preserved. Snow (2017) in another malaria study indicated that a team of researchers spent 21 years tracking down malaria survey reports done across Africa. Most of it was hidden in old government archives or curated by the World Health Organisation. Most of the records were either poorly stored, burnt or were missing. In some countries like Kenya, Senegal, Tanzania, South Africa, Botswana, Namibia and Burkina Faso the surveys dated back to the 1950s, before even typewriters were used. Conversely, recent surveys have been easier to locate through more modern web based searches.
- 17 The next part of this paper will briefly discuss progress made in terms of the four AOSP focus areas.

Open Science Open Data Policy

- 18 Although there is no existing open science policies on national level this far, advocacy and awareness creation efforts among government officials—in Madagascar, Botswana, South Africa, Kenya, Uganda and Ethiopia—have brought together stakeholders and decision makers on national level to discuss this very important topic towards formulating national policies. The fundamental role of data for research, for economic development and ecological sustainability are recognised as important and are supported

by the selected governments, as well as research institutions and the community. Selected institutions, such as the ICT Centre of Excellence and Open Data in Kenya, has institutional level policies to govern data and the outcomes of science.

- 19 Also to be addressed as part of policy is copyright (part of intellectual property). According to Terroir (2016), “in many African countries, intellectual property protection is undeveloped, ineffective, expensive and unenforced and in some African countries there exists uncertainty on protection of IP and the threat of innovation being stolen away from inventors.” African countries have to be accountable, and have to learn to manage its intellectual property in a responsible way, with digital licenses indicating conditions of use where research and data are made digitally available.

Open Science Open Data Incentives

- 20 Throughout the project and this far, the issue of incentives was tabled many times, with a call for a paradigm shift in how we recognise and credit research contributions, incl. published/peer-reviewed research articles in recognized high quality scholarly journals.
- 21 This far, we have identified the following reasons why researchers are hesitant to share their data:
- They are afraid of their ideas and research getting scooped, somebody else doing better than them.
 - They feel that they have invested a lot of time and effort, and why should they share their hard work? What is the benefit in it for them? What will they get in return?
 - They are worried that someone else might find a path-breaking application which—being the original researcher—they have not considered.
 - Fear of embarrassment, and mistakes/errors in the data collection or measurement being exposed.
 - Often ethical clearance has not been done properly.
 - Privacy of participants might be at stake, when different data sets are compared and parallels are drawn.
 - Little understanding about intellectual property rights (incl. copyright and author rights), and licensing.

Open Science Open Data Capacity Building

- 22 Developing new skills, expanding existing skills, and building capacity are required across all research disciplines and institutional support services. At the same time, professions that will probably have to look at further skills development include Engineers, Statisticians, Data Scientists, Librarians, Data Curators, Researchers, System Administrators, Policymakers, Auditors, Data Centre Managers, and Data Architects.
- 23 A broad spectrum of skills are listed below, which will also inform the AOSP Capacity Building Framework, to be delivered by the end of Year 2:
- Data Use Principles.
 - Data Storage & Sharing Repositories/Databases.
 - Command Line Interpretation.
 - Software Development, Programming.
 - Data Transfer.

- Data Organisation & Management.
- Data Cleaning.
- Data Analysis/Interpretation.
- Data Translation & Visualisation.
- Data Curation & Standardisation.
- Machine Learning.
- Artificial Neural Networks.
- Research Data Management Planning.

ICT Infrastructure for Open Science Open Data

- 24 Reliable ICT Infrastructure is fundamental to collaboration as part of the research process, and without it, research cannot have the impact required to address the objectives of the 2030 SDGs. Research intensive universities in Africa are connected through NRENs, and the investment in African NRENs are bearing much fruit, making research collaboration and exchange of research possible with the North, through the European GEANT and American InternetConnect2 networks.
- 25 The following NRENs already allow for running data-intensive applications and sharing of high end computing assets, bio-modelling and computation: KENET (Kenya), TENET (South Africa), RENU (Uganda), and ZAMREM (Zambia) (Adam 2016). Great progress has also been made in Algeria, Egypt, Kenya, Morocco, Senegal, Tunisia, South Africa, Uganda, and Zambia, connecting universities and research institutions.
- 26 Examples of data services provided by NRENs (Adam 2016) are listed below:
- The Tunisian Computing Center el Khawarizmi (CCK) established the Tunisian National Academic Network to interconnect the campus networks of the various Tunisian universities, and also manages a data centre and provides a set of Internet and application services.
 - The Kenya education network (KENET) is both the physical network and the organization delivering access to shared services like domain names, data center, cloud computing and science gateways, capacity building, security services.
 - The UbuntuNet Alliance regional network, the network latency between Kenya and SA, has been cut 6 fold from 360 milli seconds (2011) to 60 milli seconds (2015), to share data quickly and collaborate more efficiently.
- 27 Foley (2016) in the World Bank funded report, lists numerous other services that are provided by NRENs, and that can potentially be provided by NRENs. It includes security, data storage and archives, connection of e-resources such as electron microscopes, astronomical telescopes, medical imaging, simulators, sensor networks (oceans, sky, rivers, forests, space, remote areas), accelerators, supercomputers, state-of-the-art affordable bandwidth on demand, computing power, capacity-building workshops, grid computing middleware, and dedicated point-to-point Internet Protocol (IP) circuits for special applications. Data storage using open source data repository software such as Dataverse that can facilitate the scholarly citation of data, should further be considered as a possible service.
- 28 To move forward from merely providing basic connectivity services to supporting a range of collaborative tools, gateways and the “middleware services” mentioned, African NRENs will have to invest in capacity building (Foley 2016), and align further development with

the needs experienced by researchers—also in terms of exchanging big volumes of data across disciplines, in a trusted and secure way—across the globe. High-speed Internet will always remain essential for collaboration with global peers and for access to high-cost instrumentation and for the exchange of large data sets (Foley 2016).

Conclusion

- 29 Research across disciplines are becoming more and more data driven. The African Open Science Platform is a major intervention by the South African Department of Science and Technology, trying to better understand the African landscape in terms of research collaboration for sharing. It tries to better understand progress made in terms of open science policy, ICT infrastructure in support of data sharing and collaboration, skills required and capacity building towards data science, and incentives for sharing data.
- 30 A better understanding of the landscape is expected to inform a next phase, working towards a virtual space which can be used by African researchers to access/upload research instruments and data sets across disciplines in a cloud—possibly the African Research Cloud. The project further expects to stimulate collaboration among researchers, make funders aware of research conducted, avoid duplication of research, stimulate interdisciplinary use of data, and help identify research areas where there might be a need for further research.

BIBLIOGRAPHY

References

- Academy of Science of South Africa. (2009). *African Open Science Platform*. Retrieved February 14, 2018, from <http://africanopenscience.org.za/>
- Adam, L. (2016). *Riding the National Research and Education Networking Train in Africa*. Retrieved February 14, 2018, from <http://africanopenscience.org.za/wp-content/uploads/2018/01/AAU-Riding-the-NREN-Train-final-draft-.pdf>
- Foley, M. (2016). *The Role and Status of National Research and Education Networks (NRENs) in Africa*. Retrieved February 14, 2018, from <http://africanopenscience.org.za/wp-content/uploads/2018/01/WB-Report-Role-and-Status-NRENs-Africa.compressed.pdf>
- FOSTER Open Science. Retrieved February 14, 2018, from <https://www.fosteropenscience.eu/>
- Gurin, J. (2015). *Sustainable Development Goals and Open Data*. Retrieved February 14, 2018, from <http://blogs.worldbank.org/ic4d/sustainable-development-goals-and-open-data>
- Mulder, N. *et al.* (2017). “Development of Bioinformatics Infrastructure for Genomics Research in H3Africa.” In *Global Heart*, vol 11, issue 2. Retrieved February 14, 2018, from <http://dx.doi.org/10.1016/j.gheart.2017.01.005>

- Open Data in a Big Data World. Science International, 2015. Extended and Abbreviated Versions.* Retrieved February 14, 2018, from <http://www.icsu.org/scienceinternational/accord>
- Royal Society. (2012). *Science as an Open Enterprise*. Retrieved February 14, 2018, from <https://royalsociety.org/topics-policy/projects/science-publicenterprise/report/>
- Science International. (2016). *Science International Accord on Open Data in a Big Data World*. Retrieved February 14, 2018, from <http://www.science-international.org/>
- Serwadda, D. et al. (2018). “Open data sharing and the Global South—Who benefits?” In *Science*, vol 359, issue 6376. Retrieved February 14, 2018, from <http://10.1126/science.aap8395>
- Snow, B. (2017). “What 115 years of data tells us about Africa’s battle with malaria past and present.” In *The Conversation*. Retrieved February 14, from <https://theconversation.com/what-115-years-of-data-tells-us-about-africas-battle-with-malaria-past-and-present-85482>
- Terroir, P. (2016). “A new look at Intellectual Property and Innovation in Africa.” In *IPStrategy.Com*. Retrieved February 14, 2018, from <https://ipstrategy.com/2016/12/05/a-new-look-at-intellectual-property-and-innovation-in-africa/>
- Torto, N. (2017). *Why the African Academy of Sciences is creating AAS Open Research*. Retrieved February 14, 2018, from <https://aasciences.ac.ke/updates/blogs/aas-open-blog/>
- United Nations. (2016). *Transforming our world: the 2030 Agenda for Sustainable Development*. Retrieved February 14, 2018, from <https://sustainabledevelopment.un.org/post2015/transformingourworld>
- Wilkinson, M.D. et al. (2016). “The FAIR Guiding Principles for scientific data management and stewardship.” In *Sci. Data*, vol 3, issue160018. Retrieved February 14, 2018, from https://www.researchgate.net/publication/298345883_The_FAIR_Guiding_Principles_for_scientific_data_management_and_stewardship
- Yozwiak, N.L., Schaffner, S.F. & Sabeti, P.C. (2015). “Data sharing: Make outbreak research open access.” In *Nature*, vol 518, issue 7540. Retrieved February 14, 2018, from [doi:10.1038/518477a](https://doi.org/10.1038/518477a)

ABSTRACTS

Exploitation of the digital revolution offers great potential for less affluent and least economically developed countries (LEDCs) and for the attainment of the UN Sustainable Development Goals. However, LEDCs typically have poorly resourced national research systems. If they cannot participate in research based on big and open data, the gap could grow exponentially in coming years. They will be unable to collect, store and share data, unable to participate in the global research enterprise, unable to contribute as full partners to global efforts on climate change, health care, and resource protection, and unable to fully benefit from such efforts, where global solutions will only be achieved if there is global participation. Thus, both emerging and developed countries have a clear and direct interest in helping to fully mobilize LEDC science potential and thereby to contribute to achievement of the UN Sustainable Development Goals. The initiative described here (African Open Science Platform or AOSP) is directed towards minimising a divide between emerging and developed countries in what is arguably the most important current opportunity to enhance the power and efficiency of the scientific enterprise and its contribution to societal benefit.

INDEX

Keywords: Open science, open data, African Open Science Platform

AUTHORS

INA SMITH

Academy of Science of South Africa (ASSAf), South Africa

ina@assaf.org.za

(corresponding author)

SUSAN VELDSMAN

Academy of Science of South Africa (ASSAf), South Africa

susan@assaf.org.za