

Advancing African Open Science through a Network for Education and Skills

Joseph Mwelwa, Ina Smith & Patrick Molutsi

Abstract

This paper is part of a broader strategy of the African Open Science Platform to convene and coordinate the interests, ideas, people, institutions and resources needed to advocate and to advance Open Science in and for Africa. It is predicated on two complementary sets of activity namely; enabling activities consisting of three strands and applications activities comprising four strands. Specifically, the paper's spotlight falls on strand five which advocates creation of understanding and awareness in African citizens and professionals about the dynamics of a data- and information-intensive world through an Education and Skills network.

Keywords

Open Science Platform, Education and Skills Network, National Research and Education Networks (NRENs)

1. Introduction

Agenda 2063 envisions an African continent that is peaceful, prosperous and integrated. The emerging assumptive view is that the continent would have developed elaborate intra and inter information systems and processes to exploit the digital revolution for the attainment of Agenda 2063 and its inherent benefits for the African people. The Africa of the future would be of countries whose institutions are well resourced and well-coordinated with operational infrastructural capabilities to facilitate efficient flow and exchange of information for development and decision making. However, a peak into current realities in Africa reveals typically ill-funded national research institutions, fragmented collaborative efforts between institutions at country level and internationally within the African continents, let alone the outside world. In the absence of a well-coordinated framework that promotes and facilitates collaboration and sharing of information among African countries, the goals and objectives of agenda 2063 will remain elusive. We argue that a key driver to achieving agenda 2063 is the Open Science and Open Data platform where the African science community freely shares their research thinking, outputs and data in a well-coordinated framework.

If Africa is to attain the envisaged peace, prosperity and integration in the short and long term, it has to embrace Open Science and Open Data with the development benefits that come with it. In this dispensation, all African societal actors are engaged with science in considering societal futures. Current demographical predictions on Africa show an explosion of youthful populations that are digitally alert but wafting in a maze of information from heterogenous sources to which Africa contributes little. However, these African citizens need to be equipped with contextually relevant knowledge and skills that would enable them to exercise their responsibilities as citizens in making informed choices about all spheres of their lives.

Given this imperative, the implication is to regard science as a public enterprise rather than one conducted behind closed laboratory and library doors. It is a science that engages actively with business, policymakers, governments, communities and citizens. It is a dimension of Open Science that is not simply a matter of Open Data and Open Access publishing, that merely represents science talking to itself, albeit more efficiently. This Open Science is also about engaging with other societal actors as knowledge partners in jointly framing questions and jointly seeking solutions.

We argue that these expectations on Open Science can only be fulfilled and met in an elaborate, systematic and well rationalized eco-system that feeds off a network for Education and Skills. This means involving the entire African educational eco-system – primary, secondary, colleges and universities – into thinking about Open Science. It entails crafting, developing, designing and implementing African Open Science curricular that will enable Africans to manage Open Science on their own terms. This will help to create a shared understanding and awareness in African citizens and professionals about the dynamics of a data and information-intensive world through an Education and Skills Network hosted through the African Open Science Platform. A framework to enhance accessibility and impact. The Platform (AOSP) will play a critical role in assisting African countries to develop necessary capacities to manage and exploit scientific data for the benefit of society.

2. Literature Review

African countries have elaborated school systems and relatively well developed educational infrastructures that can be supported by the Network of Education and Skills hosted of the African Open Science Platform (AOSP). The 'Platform will play a critical role in assisting African countries to develop necessary capacities to manage and exploit scientific data for the benefit of society' (<http://africanopenseience.org.za/>). Furthermore, African countries have abundant learners/students – a key resource – who are both the recipients and drivers of the innovation. Learners across the educational spectrum need to be trained in Open Science and Open Data. An exposure to training and transfer of skills aligned with what is happening in the rest of the world would develop common skills and a common understanding and adherence to international standards, aligned to best practice, across all disciplines. African science is more and more required to form part of a global Open Science ecosystem, where all are connected. Nardi & O'Day (1999) present the view that an eco-system is one in which 'the subsystems are inter-linked and interdependent, where there is continuous co-evolution, where change is systemic and where complementarity encourages niches for different roles and functions'.

The eco-system is also characterised by interactions of 'actors and organizations linked by flows of resources and information' (Mars, Bronstein & Lusch, 2012: 277). For the global science ecosystem to function in tandem, the individual ecosystem identities need to align basic elements of structure and function. To avoid the digital divide, learners in the school system must be prepared on fundamentals and complexities of Open Science and Open Data. Knowledge and skills on Open Science, gained systematically in the school system, will assure learners of confidence to participate in the science economy of the future. If basic building blocks in the form of Open Science curricular are not in place, graduates enrolling for undergraduate and post-graduate studies will lag behind, hence creating barriers towards advancing science. Effectively, these scientists will be excluded from finding solutions for global problems. This strengthens the case for a network of education and skills through the African Open Science Platform (AOSP) to provide an advisory role and focus the AOSP's strategic intents and efforts on developing awareness and skills development in Open Science and Open Data among African

citizens.

3. Methodology

A systematic review of literature on Open Science and desk review on a network of Education and Skills in Open Science was undertaken to generate insights for the paper.

4. Proposed Network for Education and Skills

The proposed *Network for Education and Skills* (NES) will enhance understanding, skills development and awareness in African citizens and professionals about the evolving dynamics of a data- and information-intensive world and basics of data science, across disciplines. The NES proposes exploration of three levels of activity in the African educational ecosystem namely schools, universities and short courses. These are briefly discussed below:

Schools. Children are growing into an increasingly complex world of data and information. The capacity to negotiate that complexity and to avoid its pitfalls is increasingly important for individual well-being and for development of the judgements required of responsible citizens. The NES will be a source of advice and curriculum development in this domain, preparing learners towards future studies and empowering them with the necessary skills towards a smooth transition between school and university, but also the world of work, privately enterprises increasingly encouraged to take hands with the research sector to implement new innovations.

Universities. It is important that universities increasingly take over the role of providing data science education as an integral part of bachelors programmes as well as at more advanced levels for researchers and research trainees. In this regard, African universities can work in collaboration with other partners ^[1] who have experience and developed expertise in delivering training workshops in data science from a well-developed curriculum. Another valuable source to guide the curriculum is the AOSP Framework towards Capacity Building in Open Science and Open Data - currently in development, and to be tested during International Data Week 2018 3-8 November. This framework will guide respective African countries as well as research organisations as to which data skills need to be developed, along with available tools relevant to respective disciplines.

The role of libraries should not be underestimated. Together with librarians from Africa, a framework is being developed towards the implementation of a service for Research Data Management (RDM) planning. The framework recommends the integration of RDM training as part of Information Literacy (IL) training, already provided by libraries. AOSP is working with AfLIA (African Library and Information Associations and Institutions) on a roadmap to implement this service across African research libraries.

Short courses for professionals. There is an increasing demand for in-service short courses in data science and management for professionals from a wide variety of sectors. Such courses will build on existing curricula, such as the CODATA Schools, the FOSTER Open Science online courses, the series of Carpentry courses, and many more, and could be a significant source of income that would contribute to self-sustenance of the Network and could even support schools and universities in the long term. For certain sectors it will also contribute to Continuing Professional Development (CPD), the process of tracking and documenting the skills, knowledge and experience that one gains both formally and informally as one work, beyond any initial training. It's a record and reflection of what one experience, learn and then apply.

4.1 Rationale for inclusion of schools

Preparation of learners on Open Science and Open Data should be part of the science curriculum. Additionally, the principles of good science and the ability to research in an ethical way needs to be embedded in African learners across the educational spectrum at a very young age. Learners ought to learn about coding, robotics, Artificial Intelligence, and more. Inclusion of learners in the Open Science discourse ensures that learners acquire knowledge and skills about Open Science in a systematic way through an Open Science curriculum. However, connectivity to promote openness is a challenge on the African continent. Scholars note that African countries typically have 'poorly resourced national research systems/[institutions]' (Smith & Veldsman, 2018)). Unless African countries develop effective and sound networks and collaborations, Smith & Veldsman (2018), warn that, these nation states 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 critical areas such as 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' (Smith & Veldsman, 2018:1).

However, learners would require skills to facilitate collaboration and exchange of research data. Furthermore, for Open Science and Open Data sharing to occur, some key factors such as infrastructure and curriculum are crucial: - For example, schools need to be well connected. In Africa, (NRENs) are expected to play an increasingly important infrastructure role. KENET is a good example where it has been successfully implemented. The curriculum is another factor that needs serious attention. A context specific and African needs driven African Open Science and Open Data curriculum needs to be developed. Africa could look to international organizations within and outside Africa for expertise and to benchmark curriculum development and implementation. Examples of *Curriculum that has already been developed are*:- Data Science as part of curricula e.g. University of the Witwatersrand (South Africa), Bachelor of Science in Data Science, Sol Plaatje University (South Africa), Masters Program in Biodiversity Informatics, Prof Jean Ganglo, University of Abomey-Calavi (Benin). What this demonstrates is leadership, strength and progress in Open Science on the African continent

Although this discussion cites important milestones in an Open Science and Open Data curriculum on the African continent, these milestones are largely anecdotal. However, NES through the AOSP would systematise planning and implementation of Open Science curricular to maximise impact among African citizens although enormous challenges exist. Chief among these challenges are: lack of infrastructure, hardware, trained teachers and facilitators. Additionally, standardising curricula across the continent and providing curricular in accessible linguistic platforms would be issues that would require urgent attention if NES is to succeed.

4.2 Suggested Solutions

The critical building blocks towards a systematic network of education and skills on the African continent would be through harnessing the National Research and Education networks (NRENs) that support needs of education and research. In Southern, Eastern, Central, West and North Africa, some NRENs have been set up which the NES will tap into to broaden and expand the support eco-system on research and education services to schools, universities and the community to promote citizen science & engagement. Mapping data from the AOSP on AU initiatives also provide insights to build understanding on infrastructure and hardware. AOSP2 working through NES would provide advisory information on curricular on learning and training/facilitation in Open Science and Open Data.

5. Curriculum Content towards Capacity Building

Curriculum content - on school, university and short-course levels - should be aligned with the skills required as part of the data life cycle, also part of Research Data Management (RDM) planning, demonstrated below:

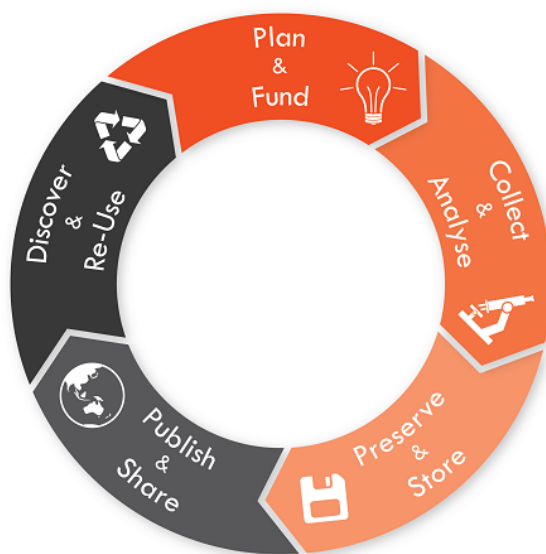


Figure 1. The Research Data Lifecycle ([The University of Sydney](#), 2018)

Different tools are being used by disciplines during the various stages of the data lifecycle, and the authors would like to recommend each discipline to explore which tools to use to advance Open Science and Open Data for collaboration to be effective. Often the tools will also be project dependent, and the decision as to which tools to use will depend on funding, functionality and more. Projects such as H3ABioNet refers to “pipelines” being set up, connecting different Open Source applications, programming it to work together as part of an automated workflow.

Following is a summary of the proposed data related skills to be taught to successfully collaborate in sharing data:

Data Activity/Skill	Description	Example Applications
Data Planning	Explore how research data will be managed in a holistic way, at the start of a research project. Especially important in the context of collaborative research projects. Researchers may identify areas of potential difficulty or conflict, and these can be resolved with colleagues and collaborators before they escalate into issues. Clarifying ownership of data, and ensuring early agreement on technical standards and frameworks across institutions, are an important part of establishing trust and ensuring that a project runs smoothly (Monash University).	http://www.dcc.ac.uk/resources/data-management-plans The RDM checklist can be in a wordprocessing format, online form format, or machine-actionable format. Example of a specific online form format RDM tool: DMPonline DMPRoadmap can be installed on servers at local institutions.

Data Collection	Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes (Wikipedia).	Surveys, Questionnaires, Telescopes
Data Analysis	Data analysis (data computing) is a process of inspecting, cleansing, processing, transforming, visualising and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making (Wikipedia).	R, Python, ggplot2, MATLAB, C, C++, Git, GitHub, Mercurial, OpenRefine Incl. batch/grid computing: Condor, HTCondor, SGE, PBS/Torque, LSF, SLURM Incl. computing workflow: DAGMan, Pegasus, Makeflow Incl. command line interface interpretation: UNIX Shell, bash (Nano, Emacs, Vim)
Data Preservation	Data preservation is the act of conserving and maintaining both the safety and integrity of data. Preservation is done through formal activities that are governed by policies, regulations and strategies directed towards protecting and prolonging the existence and authenticity of data and its metadata (Wikipedia).	SQL, MySQL, PostgreSQL, Hadoop (Big Data) Incl. data repositories: Figshare, Zenodo, Dataverse, DSpace
Data Publishing and Sharing	Data publishing (also data publication) is the act of releasing research data in published form for (re)use by others. It is a practice consisting in preparing certain data or data set(s) for public use thus to make them available to everyone to use as they wish. This practice is an integral part of the open science movement (Wikipedia).	Data repositories e.g. registered on re3data.org Data Journals
Data Mining, Discovery, Access and Re-Use	Data mining or discovery is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems (Wikipedia). Data Access refers to data being discoverable - following it being collected - on the World Wide Web through metadata assigned to datasets. Distinguish between Open Access data and Restricted Access data. Depending on the license that applies (see	Data repositories e.g. re3data.org Data Journals

	Creative Commons Licenses), data can be re-used in different ways.	
--	---	--

Table 1. Overview of Research Skills/Activities towards Capacity Building

Network and system engineers (also those employed by NRENs) will further need to be trained to scale the network, make provision for the sharing of big data sets, and offer services similar to those mentioned as part of the data lifecycle.

7. Conclusion

This paper provided an overview of three levels to benefit from an AOSP *Network for Education and Skills* (NES), namely schools, universities and professionals. A rationale for developing skills was provided, as well as a rationale for NRENs to expand connectivity beyond just universities. The skills towards capacity building were tabled, and will need to be adapted according to the needs experienced within specific disciplines.

A successful AOSP *Network for Education and Skills* (NES) will depend on policies, infrastructure, incentives and skills development to be in place, towards playing a critical role in assisting African countries to develop necessary capacities to manage and exploit scientific data for the benefit of society. Africa is to take ownership and responsibility for its research data, which is at the same time its intellectual property, but this can only be achieved once the building blocks are in place and there is the political will from governments to further support this platform.

8. References

- About research data management, University of Sydney Library. [online] Available at: <https://library.sydney.edu.au/research/data-management/research-data-management.html> [Accessed 6 Aug. 2018].
- African Open Science Platform. [online] Available at: <http://africanopenscience.org.za/> [Accessed 6 Aug. 2018].
- FOSTER Open Science. [online] Available at: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition> [Accessed 6 Aug. 2018].
- Mars, M.M, Bronstein, J.L. & Lusch, R.F. (2012). The value of a metaphor: organizations and ecosystems. *Organizational Dynamics*. 41: 271-280. Available at: http://www.sdlogic.net/uploads/3/4/0/3/34033484/mars_bronstein_lusch_og_dyn.pdf [Accessed on 6 Aug. 2018].
- Nardi, B.A. & O'Day, V.L. (1999). Information ecologies: using technology with heart. *First Monday*. Vol 4. No1. Available at: <http://firstmonday.org/ojs/index.php/fm/article/view/672/582> [Accessed on 6 Aug. 2018].
- Smith, I. & Veldsman, S. (2018). Data Driving Sustainability: The African Open Science Platform Project. Leslie Chan; Pierre Mounier. *ELPUB 2018*, Jun 2018, Toronto, Canada. <10.4000/proceedings.elpub.2018.25>. <hal-01816711>

Sol Plaatje University (2018). Bachelor of Science Degree in Data Science.
<https://www.spu.ac.za/index.php/programs-3/>

[1] Partners are Research Data Alliance (RDA), Committee on Data for Science and Technology and CODATA, RDA and the Organisation for Economic Cooperation (and Development (OECD).