African Open Science Platform

PART II: FRAMEWORKS

PREPARED BY THE ACADEMY OF SCIENCE OF SOUTH AFRICA

FUNDED BY THE NATIONAL RESEARCH FOUNDATION, DEPARTMENT OF SCIENCE AND TECHNOLOGY, SOUTH AFRICA

12 AUGUST 2019

African Open Science Platform (AOSP) Capacity Building Framework

Fostering a culture of Open Data within African National Systems of Innovation

Developed by Joseph Mwelwa[1], Ina Smith[2] and Susan Veldsman[2]

[1]Joint Minds Consult
[2]Academy of Science of South Africa

Rationale for a Framework on Capacity Building in Open Data Science

The world is catapulting towards the Fourth Industrial Revolution (4IR) (the 'data revolution') whilst in the midst of the Third Industrial Revolution (the 'digital revolution'), the latter resulting in an explosion of Big and Broad Data with potential profound implications on societal, economic and scientific areas (AOSP, 2018). This data revolution is shaping the thinking and actions of humanity to make decisions and solve problems in more innovative ways than ever before. In advanced and emerging economies, data is the key ingredient to solving humanity's problems. A society's ability to solve complex social and economic problems is predicated on that society's data analysis capabilities and deployment of skill sets that deliver information at the points of need. This process enables efficient decision making and response mechanisms based on analysis drawn from the appropriate data sets (AOSP, 2018).

For Africa to excel in the new data dispensation, commitment and investment are required towards: 1) Open Data policies on governmental level, 2) enabling Information and Communications Technology (ICT) infrastructure environments, 3) commitment to train system architects, system support staff, user support staff, data engineers, data architects, data stewards and data scientists to practice data science, but also to design, implement and maintain an e-infrastructure and 4) efforts to incentivise FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**e-usable) data sharing. Only if there is commitment from African governments, would it be possible to realise the African Open Science Platform's mission, which is "*[T]to put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for a modern society*" (AOSP, 2018). In contrast with countries in Europe, the USA, Australia, Russia and parts of Asia and the Middle East where 'government bodies and organisations everywhere are digitising their business [thus] producing huge amounts of data', the majority of African countries seem to be hesitating on the edge of the digital revolution. Even where Africa produced data, data mining and analysis have not been sufficiently progressive, for example using data to predict future trends or inform policy and decision makers. Research literatures note that the lack of data science (incl. data curation) and software engineering skills is a problem worldwide, especially in Africa. Africa has not been able to train and produce enough data analysts and scientists and other support staff to process large data sets (Big and Broad data) and to identify patterns and establish relationships to solve problems. With the data explosion emanating from activities in both the private and public sectors in Africa, the challenge presented by a shortage of skilled workers in these areas will only get worse.

The assumption is that the AOSP mission can only be 'fulfilled [in] an elaborate, systematic and well rationalised [Open Data Science] ecosystem that feeds off a network for education and skills. This means involving the entire African educational ecosystem – primary, secondary, graduate and post-

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

graduate – into thinking about Open Science (including Open Data as a sub-set of Open Science). It entails crafting, developing, designing and implementing African Open Science curricula that will enable Africans to manage Open Science on their own terms' (Mwelwa, Smith & Molutsi, 2018). A framework on capacity building in Open Data – feeding into Open Science - is therefore proposed to guide African systems of innovation on the required training to develop the requisite knowledge, attitudes, skills and values to participate in global Open Data activities.

Although training institutions are trying to mitigate the need for data specialists through teaching computer science, statistics and informatics, they are unable to meet the increasing demand for data scientists, data stewards and the necessary support staff. Other challenges faced include: curricula between institutions are not aligned and/or coordinated; the use of resources are not optimised; training institutions function in a fragmented way and operate in silos, and African students are only exposed to aspects of data science and software engineering on a higher level education. To add to this - the need to provide training to African citizens increases with a growth in population. For these reasons, the AOSP's strategy to 'create and sustain high level, internationally competitive research capacity in data analytics and Artificial Intelligence (AI)' will have a huge impact on Africa.

Through rationalised and coordinated training, the AOSP aims to create understanding, awareness and capacity among citizens and professionals in dealing with a data and information-intensive world. This can be done through the creation of priority programs that place African scientists at the international forefront in the application of cutting-edge technologies to major research domains, and as a fundamental resource for modern society.[1] Given this background, there is a need for a *Capacity Building Framework* that champions the values and aspirations of the AOSP and, which advances Africa's science priorities as outlined in the Science, Technology and Innovation Strategy for Africa (STISA 2024) and Agenda 2063.[2]

Focus and Scope

The *AOSP Capacity Building Framework* acknowledges the multiple stakeholders involved in capacity building in data science, including but not limited to policy and decision-makers (governmental and institutional), curriculum developers, researchers, system architects, system support staff, system engineers, user support staff, data engineers, data architects, data stewards, data scientists statisticians, informaticians, computer scientists, facilitators, lecturers, teachers, librarians, and citizens within society, training institutions, funding bodies, publishers, NRENs, national governments, etc. The framework will focus mainly on the competencies (incl. skills) of 'data scientists' and 'data stewards' (sometimes referred to as 'data curators' or 'data managers').

Governments and Funding Agencies (policy/decision-makers, funders)

Policy and decision-makers in governments and funding agencies should have a good understanding of what Open Science entails, the need for Open Data, and the vast array of competencies required to practice Open Science and Open Data. Data scientists play an important role in producing the information for decision making by governments and industry, and a better understanding of the importance of data scientists and data stewards can assist with making informed decisions regarding policy and funding allocation.

[1] The African Open Science Strategy document articulates Application activities summarised as Strand 3, 4 and 5

[2] STISA 2024 sets out 8 important priorities for the African research community and recognises that long-term success in achieving the priorities will depend upon increased rates of wealth creation in Africa and its states. This is why data and training in data science becomes an important driver in meeting these priorities. Agenda 2063 on the other hand aspires to see a continent of prosperous Africans. The implication for data and use of data to create wealth and prosperity cannot be understated.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049
This work is licensed under a Creative Commons Attribution 4.0 International License.

Data science training is often a huge component of funded projects, due to the lack of skilled data scientists and data stewards. Policies and funding should therefore make sufficient provision for training, towards developing 'relevant individuals and institutional capacities in Open Data Science and [which] will help to create national [Open Data] intellectual infrastructure: in education, in the national science base, in public administration and in the commercial sector' (AOSP, 2018).

Higher Education Training Institutions (data scientists, data stewards)

The *AOSP Capacity Building Framework* aims to enable African training institutions to draw from it and design courses for study by African data scientists and data stewards, to develop curricula to govern research data (Open Data), at the same time implementing Open Science practices and FAIR principles. The ultimate goal will be to produce data scientists and stewards who will provide the 'much-needed [curatorship], insights, knowledge and analysis for more informed strategic and inclusive decision-making across the continent'.[3]

Primary and Secondary Schools (teachers, facilitators, learners)

The *AOSP Capacity Building Framework* supports the teaching of data skills and data science from primary school level, through secondary school, and ultimately on graduate and post-graduate level (research). To achieve a realistic curriculum framework that links to primary and secondary school levels, a review of Mathematics, Science and Social/Development Studies curricula across African education systems will be required to inform possible restructuring (not in the scope of this framework). In addition to understanding and applying all stages of the research and data lifecycles, '[A]all students [across the education spectrum] should understand how data contributes to advances in knowledge, be able to critique claims based on data and understand the role of data in the modern world'.[4] For the interim, existing data science courses developing data skills among school learners can be integrated as part of the school curriculum (see section on *Approaches to Training*).

Citizens

In an increasingly data-driven society, where everything is mapped, measured and recorded in digital bits, the ability to engage with, handle and produce data, are important life skills (not in the scope of this framework). Citizens should have a good understanding of their rights and responsibilities as far as data is concerned, in other words: when to share data, whom to share data with, how to responsibly use open data available to inform decisions, policies and laws governing society in general, data privacy, data security, and many more. Data can be used to benefit society and to innovate, but at the same time, it can also be exploited for not so good intensions. All citizens should be equipped with the necessary skills to be responsible digital citizens, with the ability to use online tools in the most accurate responsible way - also as far as data is concerned. Adding data skills as part of school curricula and advocating for responsible data engagement through community library programmes are two examples of how citizens can become responsible digital data citizens.

Benefits of Capacity Building in Open Data

Commitment and investment in Open Data training are expected to offer the following benefits for individual institutions and countries, but also for the continent and the world at large:

- Embracing Open Data practices and training in Africa will unleash enormous potential for innovation and job creation in Africa.
- Foundational skills in Open Data will open new frontiers in technological innovation that utilises Big and Broad Data towards Artificial Intelligence (AI) or Machine Learning (ML) technologies and robotics, combined with relatively cheap and accessible techniques such as 3D printing.
- Data can be collected 24 hours per day, through using remote sensors, satellites, and Unmanned Aerial Vehicles (UAVs), which in turn can be used to monitor plant health, soil

[3] Suchith (2019) Email comments on the benefits of Digital Earth Africa. These comments generally apply across all data domains in Africa.

[4] Vanessa Pittard (2018). The integration of data science in the primary and secondary curriculum p. 5

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

condition, temperature, humidity, and more. Such data on local conditions, categorised and correlated using Machine Learning algorithms, can identify optimal farming practice to maximise yields, plan production and predict outcomes. This will not only benefit farmers, but also consumers, food security being a huge challenge to overcome. This is just an example, and the benefits of utilising Big and Broad Data can be extended to many areas, such as infectious diseases, disasters, climate change, and more.

Challenges in Building Data Science Capacity

Factors hindering the development of data science skills are for example:

- Management unaware of the need/importance to invest in data science training.
- Lack of training opportunities.
- Lack of funding.
- Inadequate infrastructure to practice data science. This include: slow and unstable connectivity, unreliable power supply, continent-wide obsolete computer infrastructure that varies between medium-scale server infrastructures to a small number of workstations, lack of centralized and secure data storage.
- Negative attitudes/resistance towards data science, which can be the result of researchers not wanting to add to their existing workload, or wanting to continue to do science in traditional ways.
- Available training courses not acknowledged by national accreditation agencies.
- Lack of leadership and direction.
- Acquiring data science skills not being part of performance agreements and appraisals.
- The thought that integrating data science as part of the primary and secondary school curriculum will place a demand on the education systems in African countries.

Suggestions to overcome the mentioned challenges include for example:

- Institutions to adopt Open Data policies, and to create an enabling environment to practice data science.
- Create opportunities for training, for example workshops and online seminars (webinars), or encourage enrolment in Open Access online courses (MOOC's) on data science, available free of charge.
- Collaborate with partner institutions, share educational resources, and develop online courses, accessible through Learning Management Systems (LMSs).
- Share resources among institutions, optimising the potential of those resources.
- Funders to make provision for capacity building as part of grants allocated.
- Institutions to make provision for capacity building as part of institutional budgets.
- Include data science training as part of Continuing Professional Development (CPD).
- Integrate a data science module as part of all courses in higher education, and faculty should encourage students to enrol in these modules.
- Incentivise employees and recognise efforts as part of performance appraisal.

Open Data Science Competency Framework (Technical & Non-technical)

The focus of the following competency framework is on higher education level training institutions and discipline-specific project training initiatives, and applies to data stewards and data scientists specifically. It is not prescriptive, and should be regarded as a mere guideline. It further proposes both technical and non-technical foundational data science skills for the following categories of people:

- Data Stewards (can include data scientists, librarians, repository managers, data managers, etc.)
- Data Scientists (can include statisticians, informaticians, computer scientists, software engineers, data stewards, etc.)

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Although not in the scope of this framework, system architects, system support staff, user support staff, data engineers, data architects, software engineers, network engineers, system administrators, and citizens will also require some level of training.

Foundational competencies required by data stewards and data scientists, as well as example applicable technologies, are tabled below. It is broadly organised according to different stages of the research data lifecycle. The roles of the data steward and that of the data scientist are not exclusive of one another, and should complement one another – the roles are expected to be defined by the available resources, skills, discipline, expertise, and more. The definitions for the two roles can be used as a guideline to assign roles and responsibilities:

- A 'data steward' refers to a professional who handles and manages data and whose responsibilities include planning, implementing and managing (research) data input, storage, search, and presentation. [The] Data Steward creates data models for domain specific data, support and advice domain scientists/ researchers during the whole research and data management lifecycle.[5]
- A 'data scientist' is a practitioner who has sufficient knowledge in the overlapping regimes of expertise in business needs, domain knowledge, analytical skills, programming and systems engineering to manage the end-to-end scientific method process through each stage in the Big Data lifecycle – until the delivery of an expected scientific and business value to science or industry.[6]

This framework further acknowledges that – depending on the discipline/research focus - different approaches and different applications will inform the different stages of the research data lifecycle in *Figure 1*.

The *AOSP Capacity Building Framework* should further be viewed and interpreted alongside the *AOSP Open Data Policy Framework*[7] (providing direction to data stewards and data scientists), the *AOSP Incentives Framework*[8] (creating an enabling environment for data scientists and data stewards to engage in high performance computing and share data), the *AOSP Research Data Management Framework*[9] (planning how research data will be managed throughout the research data lifecycle), and the *AOSP e-Infrastructure Framework*[10] (making provision for ICT support services and an infrastructure conducive for collaboration and data sharing).

[5] The working definition of data steward adopted in this framework is the Edison definition for a data steward on p. 21 of the Data Science Framework document presented at the Malta workshop June 8-9 2017.

[6] The working definition of data scientist adopted in this framework is the Edison definition of a data scientist on p. 9 of the Data Science Framework document presented at the Malta workshop June 8-9 2017.

[7] http://dx.doi.org/10.17159/assaf.2019/0052

[8] http://dx.doi.org/10.17159/assaf.2019/0051

[9] http://dx.doi.org/10.17159/assaf.2019/0050

[10] http://dx.doi.org/10.17159/assaf.2019/0048

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

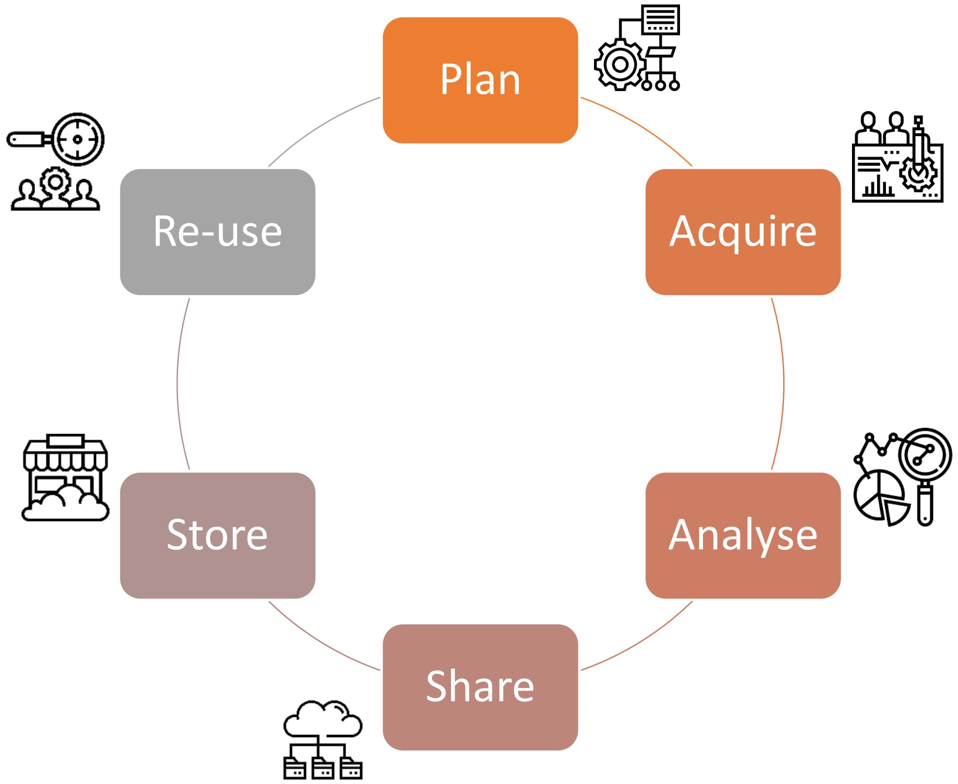


Figure 1. Research data lifecycle

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Open Data Science Competency Framework (Technical Skills)

Research data should be managed throughout the research data lifecycle, from the initial planning, up to the sharing and archiving stages. According to the AGUs DMM^{SM} Best Practices and the Data Lifecycle,[11] the following activities should span the research data lifecycle: measurement and analysis; managing processes; quality assurance; risk management; configuration management, and sustainability, consistency and resilience.

The following table classifies various technical competencies according to the different stages in the research data lifecycle. A description of the data stage and accompanying activities are provided, followed by example competencies required and activities to accomplish that specific stage. Example applications/sources are listed in support of the activities in each data stage. It contains both open source and proprietary applications. This table is by far not comprehensive, and is merely a guideline.

	Description	Example Competencies	Example Activities	Example Applications/Sources
Research Data Management	Plan **Planning** - Explore how research data will be managed, and document it as part of a data management plan, often required by funders. **Funding** - Identify possible funders/calls for funding (grants); Interpret funder policies, terms for funding; Prepare and submit a proposal/apply for funding. Report back to funder according to the funder contract/agreement, etc.	1. Understand what FAIR data entails, and understand the jargon associated with the research data lifecycle. 2. Manage data throughout the research data lifecycle. 3. Compile an application for a funding grant; report back to the funder. 4. Design a research study idea/concept. 5. Register an ORCiD and integrate as part of the scholarly workflow. 6. Apply and adhere to research related policies. 7. Set-up alerts to receive notifications on published resources. 8. Use a reference management tool to manage references. 9. Include DOIs as part of citations, and assign DOIs to new datasets. 10. Familiarise yourself with an appropriate text writing tool.	• Plan/strategise how the data will be managed from the start until the end of the research process. • Attend data management training. • Register a personal online scholarly identity using ORCiD. • Inform yourself and conform to institutional research related policies e.g. Intellectual Property Rights (incl. copyright), Data Policy, Ethics Policy, etc.	**Data management planning** The data management plan can be in a word-processing format, online form format, or machine-actionable format. DMPonline https://dmponline.dcc.ac.uk/ DMPTool https://dmptool.org/ DMPRoadmap https://github.com/DMPRoadmap NECDMC https://library.umassmed.edu/resources/necdmc/dmp Also see http://www.dcc.ac.uk/resources/data-management-plans **FAIR data sharing** FAIRsharing.org https://fairsharing.org/educational/

[11] https://dataservices.agu.org/dmm/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management	**Designing a research study idea/concept/problem** - Organise ethics approvals and other collaboration agreements.	11. Convert text format content to a web published version. 12. Generate active, online bibliographies. 13. Apply version control to different versions of datasets. 14. Identify Open Access Data Repositories, and explore which data have been collected in your field. 15. Upload data to a suitable, trusted data repository. Adhere to the funder and publisher policies for data sharing. 16. Adhere to Creative Commons License (CCL) conditions that apply to re-usable data, and assign a machine-readable CCL to your own published datasets. 17. Extract data from existing websites where applicable, using web scraping. 18. Monitor the impact of your published data using impact metrics.	• Set-up alerts to receive notifications on relevant data/research published. • Explore tools for collaboration. • Determine which reference style should apply within the specific discipline, at the research institution. Identify a reference management tool to use. Cite re-usable datasets using the assigned DOIs (digital object identifiers), and assign DOIs to your own datasets when published. • Decide which writing tool to use. • Convert text from a print to a web format. • Generate active, online bibliographies. • Convert textual data for republication and re-use. • Apply version control to different versions of datasets. • Identify Open Access Data Repositories, and explore which data have been	GO FAIR https://www.go-fair.org/fair-principles/ **Persistent identifiers (PIDs)** ORCiD https://orcid.org/ Crossref https://www.crossref.org/ DataCite https://datacite.org/ **Reference management** Zotero https://www.zotero.org/ EndNote https://endnote.com/ BibTeX http://www.bibtex.org/ WordPres/ZotPress https://wordpress.org/plugins/zotpress/ **Authorship attribution** CRediT https://www.casrai.org/credit.html OpenVIVO https://github.com/OpenVIVO **Web publishing** Jekyll https://jekyllrb.com/ GitHub https://github.com/ **Version control** Git https://git-scm.com/ Apache Subversion (SVN) https://subversion.apache.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management			collected in your field. • Adhere to Creative Commons License (CCL) conditions that applies to re-usable data, and assign a machine-readable CCL to your own published datasets. • Extract data from existing websites where applicable, using web scraping. • Monitor the impact of your published data using impact metrics.	Mercurial https://www.mercurial-scm.org/ **Document formatting** Markdown https://www.markdownguide.org/ Pandoc https://pandoc.org/ LaTeX https://www.latex-project.org/ **Funding** Crossref Funder Registry https://www.crossref.org/services/funder-registry/ Open Society Foundations https://www.opensocietyfoundations.org/ **Licensing** Creative Commons Licenses https://creativecommons.org/choose/ CC Zero (0) https://creativecommons.org/share-your-work/public-domain/cc0/ GNU General Public License (GNU GPL) https://www.gnu.org/licenses/gpl-3.0.en.html **Web scraping** Web Scraper https://webscraper.io/ Scrapy https://scrapy.org/ **Collaboration** Confluence https://www.atlassian.com/software/confluence Syncplicity https://www.syncplicity.com/en Google Apps

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management

Research Data Management

			https://get.google.com/apptips/apps/#!/all
Description	**Example Competencies**	**Example Activities**	**Example Applications/Sources**
Acquire Also see *Re-use* **Collecting/Capturing** - Data collection/ capturing refers to all activities involved in collecting information from all the relevant sources (experiment, measure, observe, simulate) to find answers to the research problem, test the hypothesis and evaluate the outcomes. Data collection methods can be divided into two categories: 1) Primary methods of data collection (capture new data using devices such as surveys, questionnaires, focus groups, interviews, telescopes, electron & astronomical microscopes, satellites, digital surveillance cameras, drones, Internet of Things (IoT), sensors & sensor networks, computer simulation models, synchrotrons, medical	1. Search existing data repositories to establish if there are quality secondary data that can be re-used. 2. Organise and structure data files. 3. Validate the data collected to make sure the collected data adhere to requirements. 4. Save files in open formats. Proprietary formats are acceptable if they are well documented and an API or conversion tool exists. 5. Clearly document how the data were collected. 6. Adhere to all research related policies e.g. Intellectual Property Rights (incl. copyright), Data Policy, Ethics Policy, etc. 7. Adhere to policies related to sensitive data (e.g. personal data, health and medical data, ecological data that may place vulnerable species at risk). 8. Acknowledge sources from which data were obtained.	• Collect data using one/more suitable instrument/s. • Apply a community/discipline-specific metadata standard. • Assign standardised keywords part of an existing community/discipline-specific vocabulary. • Follow institutional policies and discipline procedures when assigning metadata to the dataset/s. • Link metadata to the dataset to ensure provenance and citation accuracy. • Ensure metadata can support reuse: e.g. date, parameter settings, calibrations, software used, and computing environment etc.	**Collecting/Capturing** Surveys, questionnaires, focus groups, interviews, telescopes, microscopes, satellites, digital surveillance cameras, drones, Internet of Things (IoT) etc. REDCap https://www.project-redcap.org/ Open Data Kit https://opendatakit.org/ **Metadata (controlled vocabularies, glossaries)** See http://www.dcc.ac.uk/resources/ metadata-standards **Taxonomies (categories, ontologies, tags)** WordPress https://developer.wordpress.org/themes /basics/categories-tags-custom-taxonomies/ WebProtege (Biomedicine) https://webprotege.stanford.edu/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management		imaging, etc.) 2) Secondary methods of data collection (aggregate data from pre-existing resources; re-use existing data). **Assign metadata** – Do so at the point of data capturing. This will expedite data sharing, publishing and citation. Metadata can be captured in a manual or automated way.			
		Description	**Example Competencies**	**Example Activities**	**Example Applications/Sources**
Research Data Management	Analyse	Data are analysed through workflows, algorithms, software and models. **Integrating** – Combine data (incl. metadata) from different sources into a single, unified view. Integration begins with data ingestion, followed by cleaning, ETL (extraction, transformation, loading) mapping (accessing and manipulating source data and loading it into the target database, thereby	1. Combine data from different sources into a single, unified view. 2. Clean data using either a wrangling tool or scripting. 3. Screen data for further statistical analysis. 4. Analyse collected data through applying appropriate analytical processes. 5. Apply data mining, clustering, segmentation, synthesis, etc. 6. Profile data – examine the data and make interpretations. 7. Model the data before storing it in a database. 8. Store active data on an appropriate database. 9. Query a database containing datasets. 10. Monitor the quality of the data throughout the data lifecycle. 11. Visualise the findings from the data analysis through storytelling. 12. Secure a database containing datasets.	• Use a wrangling tool to clean data collected. • Use scripting to batch clean a dataset. • Integrate different datasets from different databases. • Analyse data using different tools. • Model the data collected in preparation of uploading it into a database. • Visualise the findings from the data through applying storytelling.	The categories below are broad, and the tools are to be used in an intertwining manner, with some of the tools spanning across all categories, not necessarily organised according to the classification below. **Batch/grid computing/high throughput computing** HTCondor https://research.cs.wisc.edu/htcondor/ SGE http://genomics.princeton.edu/support/grids/sge.shtml TORQUE https://www.adaptivecomputing.com/products/torque/ LSF https://www.ibm.com/support/knowledgecenter/en/SSETD4/product_welcome_platform_lsf.html SLURM https://slurm.schedmd.com/overview.html

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management	creating data element mappings between two distinct data models), and transformation. **Cleaning** - Detect and correct (or remove) corrupt or inaccurate records from a record set, table, or database. Identify incomplete, incorrect, inaccurate or irrelevant parts of the data and replace, modify, or delete the dirty or coarse data. 'Data screening' ('data screaming') is the process of ensuring your data is clean and ready to go before you conduct further statistical analyses. Data must be screened in order to ensure the data is useable, reliable, and valid for testing causal theory. **Analysis** – Inspect, clean, transform and model data, which can lead to new and useful information and conclusions to support decision-making. 'Machine Learning'	13. Document all data process diagrams, workflows, tools/applications and other data analysis activities. 14. Prepare data for long-term preservation and sharing.		**Command line interface interpretation** UNIX Shell https://swcarpentry.github.io/shell-novice/ Bash (Unix shell) (Nano, Emacs, Vim) https://medium.freecodecamp.org/how-to-parse-command-line-arguments-using-bash-case-statements-42d5c307d1c2 **Programming** PHP https://www.w3schools.com/php/ Raspberry Pi https://www.raspberrypi.org/ Ruby https://www.ruby-lang.org/en/ Perl https://www.perl.org/ **User Interface (portal/science gateway)** Future Gateway https://www.indigo-datacloud.eu/future-gateways-programmable-scientific-portal **Transmission/Execution (Pipelines/Workflows)** Cloud Pub/Sub (Google) https://cloud.google.com/pubsub/docs/overview JSON https://www.json.org/ DAGMan https://research.cs.wisc.edu/htcondor/dagman/dagman.html Pegasus https://pegasus.ae/ Makeflow http://ccl.cse.nd.edu/software/makeflow/ WampServer http://www.wampserver.com/en/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management	(ML) (a sub-set of Artificial Intelligence (AI)) is a form of *predictive data analysis*. Computer systems use ML algorithms and models to perform specific tasks without any explicit instructions, and to categorise and correlate data. The system rely on patterns and inference instead. 'Mining' (*exploratory data analysis*, *predictive data analysis*, and a field of study within ML) is the process of discovering patterns in large datasets to transform the new information into a comprehensible structure for further use. 'Clustering' is the grouping of data into a set of meaningful sub-classes (clusters). 'Segmentation' is the process of dividing data up and grouping similar data together based on the chosen parameters so that you can use it more efficiently within marketing and operations. 'Synthesis'			DAGMan https://research.cs.wisc.edu/htcondor/dagman/dagman.html TensorFlow https://www.tensorflow.org/ Nextflow https://www.nextflow.io/ **Integrating** Scriptella http://scriptella.org/ Talend https://sourceforge.net/projects/talend-studio/ Jaspersoft ETL https://community.jaspersoft.com/project/jaspersoft-etl Lab Notebooks (ELNs) https://campusguides.lib.utah.edu/c.php?g=160435&p=1051495 **Cleaning** OpenRefine http://openrefine.org/ Pandas (Python) https://pandas.pydata.org/ Dplyr (R) https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html Optimus (Apache Spark) https://hioptimus.com/ **Analysing/Programming/Modeling** R https://www.r-project.org/ RStudio https://www.rstudio.com/ Apache Hadoop https://hadoop.apache.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management	pools similar data (e.g. from clinical trials) together to obtain an estimate of the overall effect of a particular intervention or variable on a defined outcome. **Data quality monitoring and evaluation (M&E)** - The process that monitors and ensures data quality, in compliance with data quality standards for a specific environment. It is usually performed through automated data quality, data management or monitoring systems. Each environment can set its own data quality metrics and key performance indicators (KPIs). The data quality monitoring process then measures the metrics and KPIs against set criteria, and generate data quality reports. **Modeling** - The process of creating a data model (e.g. using the			Apache Spark https://spark.apache.org/ Apache Storm https://storm.apache.org/ Apache Cassandra http://cassandra.apache.org/ Scala https://www.scala-lang.org/ Python https://www.python.org/ SAS https://www.sas.com/en_za/home.html Java https://www.java.com/en/download/ Alteryx https://www.alteryx.com/ SPSS https://www.ibm.com/products/spss-statistics MapReduce (Apache Hadoop) https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm Hive (Apache Hadoop) https://hive.apache.org/ Pig (Apache Hadoop) https://pig.apache.org/ Qlik https://www.qlik.com/us MATLAB https://www.mathworks.com/products/matlab.html C++ https://isocpp.org/ C http://www.open-std.org/jtc1/sc22/wg14/ MicroStrategy https://www.microstrategy.com/us Cognos https://www.ibm.com/za-en/products/cognos-analytics

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management	Entity Relationship (E-R) Model or UML (Unified Modeling Language)) for the data to be stored in a database. It is a conceptual representation of data objects, associations between different data objects, and the rules. It helps in the visual representation of data and enforces business rules, regulatory compliances, and government policies on the data; it ensures consistency in naming conventions, default values, semantics, security while ensuring quality of the data, and it emphasizes on *what data* is needed and *how it should be organized* instead of what operations need to be performed on the data. It is like an architect's building plan which helps to build a conceptual model and set the relationship between data items.			Teradata https://www.teradata.com/ BigQuery https://cloud.google.com/bigquery/ Kotlin https://kotlinlang.org/ CARTA (Astronomy/SKA) https://cartavis.github.io/ Galaxy (Bioinformatics) https://galaxyproject.org/admin/get-galaxy/ Pegasus https://pegasus.ae/ **Modeling** erwin https://erwin.com/products/erwin-data-modeler/ Sparx Enterprise Architect https://sparxsystems.com/products/ea/index.html MS Visio https://products.office.com/en-za/visio/flowchart-software **Mining** Weka https://www.cs.waikato.ac.nz/ml/weka/ RapidMiner https://rapidminer.com/ Torch https://github.com/torch/torch7 **Machine Learning** MLlib (Apache Spark) https://spark.apache.org/mllib/ Aforge.net http://aforge.net/ **Scikit-learn** https://scikit-learn.org/stable/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management	**Visualise** – Present findings from the analysed data in a graphical and visual way. This is done using graphs, plots, and other tools. It visually presents the key interpretations and key insights extracted from the data. 'Data storytelling' is more than just a graphical representation of the data. It aims at communicating the message from the findings to a specific audience, and is accompanied by a compelling *narrative*. It combines data science, visuals and the (brief) narrative to giving your data a voice, communicating new insights effectively. It does not only tell *what* is happening, but also *why* it is happening.			**Databases** SQL (database management software) https://www.w3schools.com/sql/ Cassandra (Apache) http://cassandra.apache.org/ Bigtable https://cloud.google.com/bigtable/ HBase (Apace) https://hbase.apache.org/ MySQL https://www.mysql.com/ PostgreSQL https://www.postgresql.org/ NoSQL (database) https://www.mongodb.com/nosql-inline Spark SQL (Apache) https://spark.apache.org/sql/ JDBC https://docs.oracle.com/javase/8/docs/technotes/guides/jdbc/ Oracle (application server software) https://www.oracle.com/index.html DataGrip https://www.jetbrains.com/datagrip/ Vertica https://www.vertica.com/ Redshift https://www.redshift3d.com/ SAP ACE (Sybase SQL Server) https://www.sap.com/africa/products/sybase-ase.html OAR Database Schema http://oar.imag.fr/docs/latest/admin/database-scheme.html **Visualisation** R Markdown https://rmarkdown.rstudio.com/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management

Research Data Management

			ggplot2 (R) https://github.com/tidyverse/ggplot2 Tableau https://www.tableau.com Power BI (Microsoft) https://powerbi.microsoft.com/en-us/

	Description	Example Competencies	Example Activities	Example Applications/Sources
Share	Data sharing refers to the process of making newly generated data available for use by others. The access can be managed: Open, Mediated/ Controlled or Restricted. In some instances, data has to be digitised.	1. Appraise/valuate the data to determine whether it is scientifically, socially or culturally significant. If not, dispose of it. 2. Share data according to the funder and publisher policies. 3. Submit sensitive, restricted-access data to a secure research data centre, not online accessible. 4. Adhere to the copyright policy of the funder. 5. Assign a Creative Commons License to the data, according to the funder or publisher policy. 6. Submit the data to a trusted institutional or discipline-specific data repository. 7. When submitting the data, adhere to the metadata standards for the discipline or the repository. Use controlled vocabularies, assign persistent identifiers (DOIs, ORCiDs), use best practice file names, file formats and file sizes. 8. Monitor the impact of/ citations to the data. Also on social media. 9. Reference your own data, and create a link between the published research paper and the data itself. 10. Manage different versions of the dataset according to best practice. 11. Upload any supporting documentation/instruments/software/code/	• Appraise the data. • Verify the conditions of both the funder and the publisher, and share data accordingly. • Acknowledge the funder as part of the metadata. • Use the CCL online tool to assign a license, to specify the conditions of use. • Determine which repository to submit the data to. • When submitting the data, add metadata: descriptive, provenance and administrative metadata.	**Persistent identifiers (PIDs)** ORCiD https://orcid.org/ DOIs https://www.crossref.org/ DataCite https://datacite.org/ **Reference management** Zotero https://www.zotero.org/ EndNote https://endnote.com/ BibTeX http://www.bibtex.org/ WordPres/ZotPress https://wordpress.org/plugins/zotpress/ **Authorship attribution** CRediT https://www.casrai.org/credit.html OpenVIVO https://github.com/OpenVIVO **Web publishing** Jekyll https://jekyllrb.com/ GitHub https://github.com/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management Research Data Management	simulations/models/protocols/workflows developed along with the dataset.		**Version control** Git https://git-scm.com/ Apache Subversion (SVN) https://subversion.apache.org/ Mercurial https://www.mercurial-scm.org/ **Licensing** Creative Commons Licenses https://creativecommons.org/choose/ CC Zero (0) https://creativecommons.org/share-your-work/public-domain/cc0/ GNU General Public License (GNU GPL) https://www.gnu.org/licenses/gpl-3.0.en.html **Impact** Impactstory https://impactstory.org/ PlumX Metrics https://plumanalytics.com/learn/about-metrics/ Altmetrics https://www.altmetric.com/ HuMetrics https://humetricshss.org/blog/humetrics-values/, FAIR Metrics http://fairmetrics.org/ Make Data Count https://makedatacount.org/ **Data repositories** GitHub https://github.com/ DSpace https://duraspace.org/dspace/ Invenio https://invenio-software.org/ Dataverse

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management

Research Data Management

	Description	Example Competencies	Example Activities	Example Applications/Sources
				https://dataverse.org/ World Bank Toolkit http://opendatatoolkit.worldbank.org/en/technology.html Zenodo https://zenodo.org/ SCHOLIX http://www.scholix.org/ GBIF Integrated Publishing Toolkit https://github.com/gbif/ipt
	Description	**Example Competencies**	**Example Activities**	**Example Applications/Sources**
Store	**Data preservation** refers to conserving and maintaining both the safety and integrity of data. Preservation activities are formal and should adhere to institutional data policy requirements. This towards protecting and prolonging the existence and authenticity of data and its metadata. **Data archiving** – Refers to the long-term storage of data. Data is retrieved in the case of a disaster/ trigger event. The data in archive is also not used every day. **Data retention** – Delete data when the storage	1. Convert the data to preservation formats. 2. Decide whether the data should be preserved for the short-term or the long-term (archived). 3. Decide whether the data should be immediately accessible ('hot' storage) or not ('cold' storage). 4. Distinguish between storing active (working), master, raw, sensitive, Big Data, etc. 5. Upload the data to the archival system. Organise the data so that it can be retrieved if needed. 6. Delete obsolete data.	• Convert the data to preservation formats. • Decide whether the data should be preserved for the short-term or the long-term (archived). • Distinguish between storing active (working), master, raw, sensitive, Big Data, etc. • Decide whether the data should be immediately accessible ('hot' storage) or not ('cold' storage). • Upload the data to the archival system. Organise the data so that it can be retrieved if needed. • Delete obsolete data.	CLOCKSS https://clockss.org/ LOCKSS https://www.lockss.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management		capacity has reached its full capacity, and the data no longer have social, cultural or scientific value.			
	Re-use Also see *Acquire*	Use research data for a research activity or purpose other than that for which it was originally intended.	1. Search for existing datasets. Search in trusted data repositories. 2. Utilise web scraping tools to collect data from the web. 3. Utilise web APIs to discover, extract, and enrich existing data, or in transforming legacy data. 4. Interpret and respect the legal framework within which data can be re-used. 5. Cite the original data correctly, and include the DOI reference as part of the citation.	• Search for existing datasets. Search in trusted data repositories. • Utilise web scraping tools to collect data from the web. • Utilise web APIs to discover, extract, and enrich existing data, or in transforming legacy data. • Interpret and respect the legal framework within which data can be re-used. • Cite the original data correctly, and include the DOI reference as part of the citation.	**Data repositories** Registry of Research Data Repositories https://www.re3data.org/ OmicsDI https://www.omicsdi.org/#/home NIH BD2K DataMed https://datamed.org/index.php ELEXIR https://elixir-europe.org/platforms/data/core-data-resources FAIRsharing.org https://fairsharing.org/databases/ figshare https://figshare.com/ OpenDOAR http://v2.sherpa.ac.uk/view/repository_by_country/002.html Also refer to the *AOSP Landscape Study Appendix 1* for more. CoreTrustSeal https://www.coretrustseal.org/ **Licensing** Creative Commons Licenses https://creativecommons.org/choose/ CC Zero (0) https://creativecommons.org/share-your-work/public-domain/cc0/ GNU General Public License (GNU GPL) https://www.gnu.org/licenses/gpl-3.0.en.html

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Data Management			**Web scraping** Web Scraper https://webscraper.io/ Scrapy https://scrapy.org/ **Persistent identifiers (PIDs)** ORCiD https://orcid.org/ DOIs https://www.crossref.org/ DataCite https://datacite.org/ **Reference management** Zotero https://www.zotero.org/ EndNote https://endnote.com/ BibTeX http://www.bibtex.org/ WordPres/ZotPress https://wordpress.org/plugins/zotpress/ **Authorship attribution** CRediT https://www.casrai.org/credit.html OpenVIVO https://github.com/OpenVIVO

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Open Data Science Competency Framework (Non-technical Skills)

Unique Skills

From an analysis of jobs advertised related to data science and data stewardship, the following qualities are highly sought after:

- Growth mind set (demonstrate curiosity, welcomes failure as a learning opportunity);
- Take risks (make bold decisions and recommendations);
- Focus on and integrate external trends;
- Has high performance standards (performance driven and accountable);
- Fast/agile (removes barriers to move faster, experiments and adapts, thrives under pressure and fast pace); and
- Empowered (bring solutions instead of problems, challenges the status quo, has the courage to take an unpopular stance).

Generic Skills

Generic qualities applicable to data scientists and data stewards include:

- Data literacy (incl. digital citizenship, digital literacy, data ethics, rights & responsibilities)
- Language competency
- Communication (interpersonal, presentation, networking)
- Leadership
- Advocacy
- Organisational
- Meticulous attention to detail
- Work under pressure
- Team-player
- Self-motivated, self-driven
- Independent worker
- Pro-active
- Ability to learn quickly
- Versatile
- Multi-task
- Work amidst interruptions
- Switch between tasks
- Prioritise
- Good time management
- Analytical and critical thinker
- Innovative, creative problem solver, troubleshooting
- Conceptual thinking

Research Specific Skills

Data scientists and data stewards should have knowledge and understanding of the following:

- Research lifecycle and research data lifecycle
- Research design and management; research methodology (design experiment, collect data, analyse data, develop algorithms, identify patterns, hypothesise, explain, test hypothesis, M&E, data modeling, visualisation)
- Scientific and report writing
- Informed consent (human data) ethics
- Legislative compliance: personal data protection act/policies (incl. IP, copyright); machine-readable CC-licensing
- Data citation and reference management; version control
- Interoperable and linked data; Blockchain
- Basic & advanced programming skills in OS tools, ML, DL, AI, neural networks
- Understand concepts, tools, platforms, media to promote creation and dissemination of data
- Extract robust, evidence-based knowledge from data

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

Approaches to Capacity Building

There is a dire need for training in Open Data Science on the continent, across all disciplines. Data training of scale can be achieved by building on existing initiatives and utilising educational content available as Open Access.

Where disciplines are grouped together - also as the result of funded projects - context specific capacity building/training in Open Data Science is usually conducted.

Although the ideal is for Open Data Science to be integrated as part of all curricula across the educational ecosystem, the following approaches can also help strengthen capacity building on the continent:

Primary and Secondary Schools

Learners can be taught how to code, e.g. using *Scratch*[12] – a free programming language teaching coding principles to learners from all ages – from as young as 5th grade, or younger. Scratch helps young people learn to think creatively, reason systematically, and work collaboratively – essential skills in the 21st century.

The *Hour of Code*[13] further introduces learners from as young as 4 years to computer science and coding. Uptake in Africa is not clear, but both *Scratch* and *Hour of Code* are freely available for all to upskill themselves in their own time.

Data handling should form part of school curricula – across disciplines, and already from pre-school level. It can be taught from Gr R (pre-school) up to Grade 12, specifically as part of Mathematics, but also embedded as part of Life Skills, Physical Sciences, Computer Technology, Languages, Business Studies, Economics, Geography, Social Sciences and Tourism. An example programme is *Kaggle*[14], allowing learners from all ages to do data science projects.

Universities

Research Data Management is slowly being implemented following funder requirements, and often forms part of Information Literacy Training, conducted by libraries.

Known examples of institutions at which data science is taught include the following:

- Bachelor of Science in Data Science, Sol Plaatje University (South Africa) - https://www.spu.ac.za/index.php/programs-3/
- Masters' Program in Biodiversity Informatics, University of Abomey-Calavi (Benin)
- ICT Centre of Excellence & Open Data – iCEOD, Jomo Kenyatta University of Agriculture and Technology (Kenya)
- BSc Honours in Big Data Analytics, University of the Witwatersrand (South Africa) - https://www.wits.ac.za/csam/academic-programmes/postgraduate-programmes/big-data-analytics-honours/

Short Courses

Example short courses presented by institutions, are the following:

- UCT e-Research (South Africa) - http://www.eresearch.uct.ac.za/eresearch-training
- dLAB - University of Dar es Salaam (Tanzania) - https://dlab.or.tz/what-we-do/capacity-development/
- Explore Data Science Academy (South Africa) - https://www.explore-datascience.net/
- WeThinkCode (South Africa) - https://www.wethinkcode.co.za/

[12] https://scratch.mit.edu/
[13] https://hourofcode.com/za
[14] https://www.kaggle.com/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

- CODATA-RDA Schools of Research Data Science - http://www.codata.org/working-groups/research-data-science-summer-schools

Online Courses
Online courses available can also be utilised for further training, and are open for anyone to enrol/host:

- IBM Digital – Nation Africa https://developer.ibm.com/africa
- Open Science MOOC – https://opensciencemooc.eu/
- Coursera Data Science – https://www.coursera.org/browse/data-science
- Coursera Research Data Management and Sharing – https://www.coursera.org/learn/data-management
- FOSTER Open Science Courses – https://www.fosteropenscience.eu/
- FOSTER Open Science Toolkit - https://www.fosteropenscience.eu/toolkit
- MANTRA for Researchers – http://mantra.edina.ac.uk/
- MANTRA for Librarians – http://mantra.edina.ac.uk/libtraining.html
- Agricultural Information Management Standards (AIMS) – http://aims.fao.org/online-courses
- Online Master's Guides - https://www.discoverdatascience.org/online/
- Google Digital Skills for Africa online courses
  Understand the basics of code
  Understand the basics of machine learning
  Improve your online business security
  Google Cloud Platform Fundamentals: Core Infrastructure
  Google Cloud Platform Big Data and Machine Learning Fundamentals
  Machine Learning Crash Course
  Elements of Artificial Intelligence
  Computational Thinking for Problem Solving
  Programming for Everybody (starting with Python)
  Model thinking
  Introduction to Cybersecurity for Business
  Python Basics
  Exploring and preparing your data with BigQuery
  Enterprise System Management and Security
  Software Development Processes and Methodologies
  Cloud Computing Concepts Part I
  Kotlin for Java Developers
  SQL for Data Science
  Fundamentals of Network Communication
  Database Management Essentials
  What is Data Science?

Discipline-specific
- eBioKit (Bioinformatics) - https://github.com/eBioKit
- H3ABioNet - https://www.h3abionet.org/training

Open Educational Resources
Open licensed educational material available for teaching can be found online, and can be customised according to needs experienced:

- Author Carpentry – https://authorcarpentry.github.io/
- Data Carpentry – http://www.datacarpentry.org/
- Library Carpentry – https://librarycarpentry.github.io/
- World Data System (WDS) Training Resources – https://www.icsu-wds.org/services/training-resources-guide

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0049
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bibliography

AGU100: Advancing Earth and Space Science Data Management Assessment Programme (online). Available at: https://dataservices.agu.org/dmm/(Accessed 9 May 2019)

AOSP (African Open Science Platform). (2018). *The Future of Science and Science of the Future: Vision and Strategy for the African Open Science Platform (v02) (online)*. Available at: http://doi.org/10.5281/zenodo.2222418 (Accessed 9 May 2019)

Australian National Data Service (ANDS) Data Capture (online). Available at: https://www.ands.org.au/working-with-data/data-management/data-capture (Accessed 9 May 2019)

Christiansen, J. (2016). *EMBL Australia Bioinformatics Resource Data and Research (online)*. Available at: https://www.embl-abr.org.au/data/(Accessed 9 May 2019)

Hugo, W. (2017). *Competency framework: engineers, statisticians, data scientists, librarians, data curators & researchers (online)*. Available at: https://www.slideshare.net/AfricanOpenSciencePl/competency-framework-engineers-statisticians-data-scientists-librarians-data-curators-researcherswim-hugo(Accessed 9 May 2019)

Martin, E.R. (2016). The Role of Librarians in Data Science: A Call to Action. *Journal of eScience Librarianship*, 4(2):1-3 (online). Available at: http://dx.doi.org/10.7191/jeslib.2015.1092 (Accessed 9 May 2019)

Mwelwa, J. (2019). *Capacity Building Framework (online)*. Available at: https://drive.google.com/file/d/16gwm3Fh4KFZf2qW3Eg3x0hxY-UnBEnb6/view?usp=sharing(Accessed 9 May 2019)

Mwelwa, J., Smith, I. and Molutsi, P. (2018). *Advancing African Open Science through a Network for Education and Skills (online)*. Available at: http://www.jointmindsconsult.com/wp-content/uploads/2017/01/MWELWA-AND-SMITH-2008.pdf (Accessed 9 May 2019)

Unsworth, K. (2017). *Research data lifecycle: data skills for librarians (online)*. Available at: https://www.slideshare.net/SusanMRob/rscd-2017-bo-f-data-lifecycle-data-skills-for-libs(Accessed 9 May 2019)

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Capacity Building Framework: fostering a culture of Open Data within African National Systems of Innovation*. doi: http://dx.doi.org/10.17159/assaf.2019/0049

This work is licensed under a Creative Commons Attribution 4.0 International License.

African Open Science Platform (AOSP)
e-Infrastructure Framework and Roadmap

Fostering a culture of Open Data within African National Systems of Innovation

Developed by Rob Simmonds[1], Ina Smith[2] and Susan Veldsman[2]

[1]Inter-University Institute for Data Intensive Astronomy and the University of Cape Town

[2]Academy of Science of South Africa

Rationale for a Framework on e-Infrastructure in support of Open Data Science

Concurrent advances towards the 4th Industrial Revolution such as Artificial Intelligence (AI), the Internet of Things (IoT) and robotics will enable tremendous innovations and fundamentally transform science, business, government and society (ITU, 2018). The 4th Industrial Revolution (4IR) is not only "about" data, but "equals" data – it "is" data. According to Alan Marcus (Senior Director, Head of Information Technology and Telecommunications Industries, World Economic Forum) (2015),

> *"[A]at its core, data represents a post-industrial opportunity. Its uses have unprecedented complexity, velocity and global reach. As digital communications become ubiquitous, data will rule in a world where nearly everyone and everything is connected in real time. That will require a highly reliable, secure and available infrastructure at its core, and innovation at the edge."*

To harness the enormous opportunities, and to address the challenges and implications brought on by the 'digital revolution'[1], resulting in an explosion of Big and Broad Data (part of the 4IR or 'data revolution') with potential profound implications on societal, economic and scientific areas, countries will need to create conditions that support the deployment of next-generation networks and service infrastructures (ITU, 2018). Those networks and infrastructures should be supported and guided by policy, funding, capacity building (incl. upskilling of existing ICT workers), and continuous Monitoring and Evaluation (M&E), in order to track the growth and impact of these emerging trends.

Although AOSP phase 1 (2018) proposes a large-scale federated hardware, communications and software research e-infrastructure, including policies and enabling practises to support Open Science in the digital and data revolution, it is very ambitious for Africa. Similar to the European Roadmap (European Strategy Forum on Research Infrastructures (ESFRI)), the future AOSP phase 1 aims to support collaboration and the sharing of a wide spectrum of resources including hardware systems, providing Cloud systems connected by large bandwidth Wide Area Networks (WANs), that will host software systems enabling data analysis and providing access to massive data collections. It will also provide WAN data transfer services, and identity management and access control services to simplify access to the provided systems. In addition human resources will be provided to assist and train users to make best use of these resources. Funding from both African governments and funders will be required to realise an AOSP phase 1.

AOSP follows on existing initiatives and strategies elsewhere in the world, such as the e-infrastructure work by the Joint Information Systems Committee (JISC), US National Science Foundation (NSF), the European Open Science Cloud (EOSC), Compute Canada, the Australian Research Data Commons (ARDC) and the Southern African Development Community (SADC).

[1] Fourth Industrial Revolution (4IR) (the 'data revolution') whilst in the midst of the Third Industrial Revolution (the 'digital revolution'),

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation*. doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

The *AOSP e-Infrastructure Framework and Roadmap* should be viewed and interpreted alongside the *AOSP Open Data Policy Framework*[2] (providing direction as to how research data should be dealt with), the *AOSP Incentives Framework*[3] (creating an enabling environment for data scientists and data stewards to engage scientific data processing and sharing), the *AOSP Research Data Management Framework*[4] (planning how research data will be managed throughout the research data lifecycle), and the *AOSP Capacity Building Framework*[5] (equipping data scientists and data stewards with the required ICT skills, and attracting new people to study data and related fields of engineering).

Focus and Scope

This *AOSP e-Infrastructure Framework and Roadmap* focuses on guiding and informing governments and decision/policy-makers towards crucial aspects to be addressed as part of establishing a shared African e-infrastructure.

Stakeholders
Human resources in the form of system architects, system support staff, user support staff, data engineers, data architects, data stewards and data scientists will be required to design, implement and maintain such an e-infrastructure. Private industry such as Microsoft, Amazon and Google can potentially partner with AOSP in providing required e-infrastructure services, though care would be needed to make sure the program did not become dependent on any particular service provider whose business model may diverge from what is needed. This is particularly relevant with anti-trust discussions going on currently in Europe and North America that could directly impact these large tech companies.

Science Engagement
Science engagement is critical in creating interaction between the various AOSP stakeholders[6], incl. an understanding of needs experienced on computational resource provision, data science level, data support services, future requirements, and ownership. AOSP can potentially offer an overarching brokerage service, bringing the needs experienced by researchers and services available together, negotiating agreements towards collaboration, changing the mind sets of people, and acting as an intermediary/negotiator/agent/middleperson.

This framework and roadmap acknowledges that business models will depend on resources (human, financial, infrastructure, etc.) available at institutional, national and/or continental level. Where institutions lack the necessary resources and expertise, collaboration between National Research and Education Networks (NRENs), advanced computing service providers and research institutions should be considered. AOSP will play an important role in connecting NRENs and providers of computational and storage resources.

e-Infrastructures for Research Data Storage, Processing and Transfer
e-Infrastructure (also 'ICT infrastructure' or 'cyberinfrastructure') in this context encompasses all the devices, networks and middleware that are employed to support distributed research computing. It refers to the building of a user environment to run systems, which then provide users with the required functionalities.

Establishing an AOSP e-infrastructure will require commitment and support from governments for the foreseeable future. Such an e-infrastructure will support policies (for data sovereignty, domain specific data protocols; also see the *AOSP Open Data Policy Framework*), connectivity, training (also see the

[2] http://dx.doi.org/10.17159/assaf.2019/0052
[3] http://dx.doi.org/10.17159/assaf.2019/0051
[4] http://dx.doi.org/10.17159/assaf.2019/0050
[5] http://dx.doi.org/10.17159/assaf.2019/0049
[6] AOSP stakeholders include computational resource and network providers, system architects, system support staff, user support staff, data engineers, data architects, data stewards and data scientist and policy/decision-makers

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048
This work is licensed under a Creative Commons Attribution 4.0 International License.

AOSP Capacity Building Framework), and science engagement activities. The establishment of an African Open Science Research Cloud, including the design, development, testing and deployment of a federated network of computational facilities and services (incl. software/middleware), should be prioritised and completed by end 2021.

Connectivity

As far as connectivity is concerned, the focus of this framework and roadmap is on sustainable and stable connectivity and sufficient bandwidth for research institutions to connect to each other and with partners around the world. This framework and roadmap focuses on those research institutions, without disregarding the importance of services provided by NRENs to schools and Technical and Vocational Education and Training (TVET) institutions, e.g. SABEN[7] (South African Broadband Education Networks). SABEN, as part of the SANReN, is funded through the SA Dept. of Higher Education and Training to connect TVETs to SANReN, using the standard architectural principles of SANReN and TENET. School networks can also request to be connected via SABEN, but need to provide funding for this. Another partnership between an NREN and schools is the KENET Schools Connectivity Initiative (SCI)[8] to coordinate the various commercial, educational and government organizations interested and willing to provide Internet access and promote the use of ICT in Kenyan schools. The SCI is a platform through which public and private sectors partner in an effort to provide scalable and sustainable ICT and Internet access to schools. The SCI model is based on a holistic approach that integrates Internet connectivity, Internet access, relevant educational content and capacity building for teachers.

5G networks – although not in the scope of this framework - are the next generation of mobile Internet connectivity, offering faster speeds and more reliable connections on smartphones and other mobile devices than ever before. AOSP should as part of the next phases, demonstrate the power of uploading data from mobile phones to Cloud resources and re-sharing that data and derived products back to the community. Combining cutting-edge network technology and the very latest research, 5G should offer connections that are multitudes faster than current connections, with average [connectivity] (both download and upload speeds, e.g. for collaborative services such as videoconferencing (VC)) speeds of around 1Gbps for wireless 5G connections. The networks will help power a huge rise in Internet of Things technology, providing the infrastructure needed to carry huge amounts of data, allowing for a smarter and more connected world. New Low Earth Orbits (LEOs) satellite systems are expected to further impact on the cost and availability of network bandwidth.

NRENs and their role as part of Science Communities

In Africa, higher education and research institutions rely on National Research and Education Networks (NRENs) to provide connectivity and specialised services. NRENs are "organisations that are specialised Internet Service Providers dedicated to supporting the needs of the research and education communities within their own country" (GÉANT[9]).

The *AOSP e-Infrastructure Framework and Roadmap* recognises NRENs as key stakeholders in support of research data sharing. It also acknowledges that different NRENs are on different levels of maturity, and that Level 6 NRENs are well positioned to offer elaborated services. According to Greaves (2016)[10], Level 6 NRENs offer numerous value-added services such as videoconferencing, federated identity management and wireless roaming services. Many institutions will purchase commodity bandwidth from a commercial provider in addition to NREN-specific bandwidth, though it should be noted that this is usually done to provide high speed connections to the commercial Internet which is not a primary focus for the NRENs. A culture of collaboration [among African NRENs] is deeply established, well aligned with the collaborative approach called for as part of the African Open Science Platform.

The current status of NRENs are reflected in the table below:

[7] http://www.saben.ac.za

[8] https://schools.kenet.or.ke/

[9] https://www.geant.org/

[10] https://repository.ubuntunet.net/bitstream/handle/10.20374/69/NREN_Capability_Maturity_Model.pdf?sequence=1&isAllowed=y

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

Level 0	Central African Republic, Djibouti, Republic of the Congo, Lesotho, Libya
Level 1	Angola, Comoros, Eritrea, Seychelles, South Sudan, Equatorial Guinea, Guinea Bissau, São Tomé and Príncipe
Level 2	Botswana, Democratic Republic of the Congo (2.5), Malawi (2.5), Mauritius, Rwanda, Somalia, Swaziland, Zimbabwe, Cape Verde, Chad, Gambia, Guinea, Liberia, Sierra Leone, Mauritania
Level 3	Benin, Burkina Faso, Cameroon, Gabon, Ghana (3.5), Mali, Niger, Togo
Level 4	Burundi, Ethiopia (4.5), Madagascar, Mozambique (4.5), Namibia, Sudan (4.5), Tanzania (4.5), Côte d'Ivoire, Nigeria, Senegal, Morocco (4.5), Tunisia (4.5)
Level 5	Uganda, Zambia
Level 6	Algeria, Egypt, Kenya, South Africa, [Zambia – 2019]

Table 1. NREN status of African countries (Foley, 2016)

Similar to the sharing of e-infrastructure proposed as part of AOSP, NRENs promote collaboration among member academic institutions and the sharing of infrastructure, content and high-end ICT talent. The implementation of an AOSP e-infrastructure will be highly dependent on affordable connectivity/bandwidth and specialised services provided by NRENs. Governmental support, funding, enlightened policies, and mandates for NRENs to be the preferred service providers for academic institutions, are required for AOSP to be sustainable.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

Case for Cloud Service Providers to Support Science Communities

Infrastructure as a Service (IaaS) Cloud providers linked to NRENs can support a wide range of scientific activity. Systems such as OpenStack[11] provide a rich set of services that support the provisioning of diverse computing and data access approaches. In general these Cloud systems are differentiated from High Performance Computing (HPC) systems by providing highly flexible environments at the expense of some performance loss when executing massively parallel simulation programs. Even this is changing however, since OpenStack can support the provisioning of HPC environments onto the bare metal servers making up the underling clusters; the approach of using OpenStack to do this is part of the baseline design for the soon to be built SKA radio telescope processing facility. In addition new standards such as AARC2 BluePrint architecture[12] are enabling role based authorized access to distributed Cloud systems providing secure access to data distributed among the federated resources.

Examples of organisations providing IaaS Cloud systems to support research include NeCTAR[13] in Australia and Compute Canada (CC). A slightly different example is EGI that provides a framework and toolset to federate the IaaS Cloud systems hosted at data centres across Europe.

Simply providing IaaS does not provide tools at the level that is required by many researchers. NeCTAR addresses this by providing domain specific support at particular centres so that researchers in those domains can get support from the appropriate centre. In Compute Canada the nationally distributed support team helps researchers to make best use of the CC systems. In a different model CC also provides the IaaS computing and storage resources which are then used by the Canadian Astronomy Data Centre (CADC) to host an environment that supports the astronomy research for the Canadian academic community. In addition to supporting the federating of resources, EGI[14] also supports virtual machine and container libraries that hold scientific applications and workflow tools, that users can provision onto the federated Cloud systems to support their work.

Note that NRENs may be able to provide these computing and storage services in some cases. However, providing these services is not their core business, which is providing large bandwidth network connections between research institutions, providers of HPC systems and research Cloud resources. In most cases the responsivity of providing the advanced research computing and storage provided by IaaS Clouds is not performed by NRENs, but rather by organisations that specialize in the provision of particular types of services.

Challenges in developing an e-Infrastructure

Factors hindering the development of an e-Infrastructure are for example:

- Lack of skills – and a demand for more system architects, system support staff, user support staff, data engineers, data architects, data stewards and data scientists.
- Commercial ISPs offer Internet that are too expensive for Africa. Regulatory bodies to bring Internet costs down and make it more affordable – to engage citizens, primary and secondary school learners, as part of citizen science.
- Cloud services require high-speed network access, which are expensive. Cloud services should be connected to NRENs. Funding for the hardware in support of data sharing is however lacking. One example of a funded project like this is the DIRISA Tier 2 Ilifu project[15], but this is so far the only example in Africa and supports a subset of the South African research community.

[11] https://www.openstack.org/
[12] https://aarc-project.eu/architecture/
[13] https://nectar.org.au/
[14] https://www.egi.eu/
[15] http://www.ilifu.ac.za/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048
This work is licensed under a Creative Commons Attribution 4.0 International License.

- Big Data requires sufficient bandwidth, and stable and reliable Wide Area Network (WAN) connections. Selected areas in Africa struggle with ageing and unreliable power infrastructure and frequent power outages, interrupting Internet service delivery.
- A typical compute cluster that provides Cloud, High Throughput, and High Performance Computing services consumes large amounts of electrical power, much of which is converted into heat, requiring cooling in purpose built data centres.
- NRENs receive limited infrastructure grants or budget support for operational expenditure (OPEX) from their respective governments. They are further poorly understood by the telecom and Internet community, and are regarded as merely specialised Internet service providers that have to compete with very large telephone companies (telcos).
- Universities and research institutions on the continent have very low WAN and Internet access budgets.
- Data repositories – for global participation – should support FAIR data principles and should align with the principles of the CoreTrustSeal[16] for trusted data repositories. At this stage, only one data repository on the continent has been assigned the CoreTrustSeal.
- Projects such as H3ABioNet indicated that a challenge they experience is continent-wide obsolete computer infrastructure that varies between medium-scale server infrastructures to a small number of workstations, with multiple operating systems, and a lack of shared, secure data storage.
- Researchers are often unaware of the availability of Open Source Software tools/applications to collaborate and to share data as part of Open Science.
- The security of data is a huge concern for researchers, and providing a secure infrastructure that can be trusted by researchers, a challenge. Despite this many researchers store data on their own laptops and workstations running operating systems that are highly vulnerable to viruses that aim to expose their data. This points to the need for professionally managed data repository sites that can be trusted.
- Due to a lack of trusted data repositories, data often leaves the continent and are then published on platforms published by the Global North, changing ownership.
- English being a second language for the majority of Africans, it impacts on effective communication.

Suggestions to overcome the mentioned challenges include for example:

- Utilise high quality online capacity building/training opportunities.
- Governments to recognise the important role NRENs play through policy, and to commit funding towards NRENs and the services offered to the research and education communities.
- Train more data stewards, with knowledge on how to manage trusted data repositories.
- Create awareness among researchers about the many opportunities offered through utilising Open Source.
- Manage data within a trusted environment, and where data needs to be protected, make provision for data being stored in a secure environment (offline).
- Make provision for sufficient storage space for all data, emanating from Africa.
- Data engineers, data architects, data scientists and data stewards to engage in dialogue to better understand what the needs are, and what services can be offered in response to those needs.
- Train system and user support staff to run Cloud computing instances.
- Develop and build science engagement capacity.

Technical Skills

System support staff, user support staff, network engineers, data engineers and data architects are to be trained to provide the identified services, and funding should be made available for continuous upskilling. The skills should be aligned with the services and activities needed during and between

[16] https://www.coretrustseal.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation*. doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

every stage of the research data lifecycle (see *Figure 1*). Experts will be required to roll-out services and infrastructure, addressing needs experienced by priority disciplines. Policy/decision-makers (can include data auditors, funders, institutional management, etc.) will further be required to develop policies and make available funding to implement Open Data activities, since data stewardship (maintenance, cleaning, annotating, storing, etc.), e-infrastructure, resources (incl. maintenance, back-ups & upgrades, hardware, software, middleware) are costly, and requires financial, infrastructural and human (time, expertise) resources. When data collections are regarded as assets, it justifies the commitment of resources as a means of accruing reputational and collaborative benefits that will raise the institutional profile. Governments, institutions and researchers should plan, coordinate, share and align the use of resources, for optimal benefit and return on investment.

The following approaches should be considered as part of training and upskilling:

- Generic open online courses e.g. IBM Digital – Nation Africa https://developer.ibm.com/africa
- Discipline-specific online courses, funded through projects (also online) e.g. H3ABioNet https://www.h3abionet.org/training
- Discover Data Science https://www.discoverdatascience.org provides valuable descriptions and links to online courses
- Courses offered by universities and other training institutions

A summary of the skillset[17] (not comprehensive) of a **data engineer** is provided below:

- Running computer clusters to host software services
- Building and designing large-scale applications
- Database architecture and data warehousing
- Data modeling and mining
- Statistical modeling and regression analysis
- Distributed computing and splitting algorithms to yield predictive accuracy
- Proficiency in programming languages, especially R, SAS, Python, C/C++, Ruby Perl, Java, and MatLab
- Database solution languages, such as SQL, as well as Cassandra, and Bigtable
- Hadoop-based analytics, such as HBase, Hive, Pig, and MapReduce
- Operating systems, especially Linux
- Machine learning

A summary of the skillset[18] (not comprehensive) of a **data architect** is provided bellow:

- Applied math and statistics
- Data visualization and data migration
- RDMSs (relational database management systems) or foundational database skills
- Database management system software
- Relational (SQL) and non-relational (NoSQL) databases
- Data flow technologies, like Hadoop, MapReduce, Hive, and Pig
- Information management and data processing on Linux
- Machine learning
- Data mining and modelling tools, especially ERWin, Enterprise Architect, and draw.io
- Programming languages, such as Python and Java, as well as C/C++ and Perl
- Operating systems (Linux)
- Application server software
- Backup/archival software

The skills system administrators, network engineers, security specialists and user support workers require are crucial, especially where Cloud services are concerned. Other positions also to be

[17] https://www.discoverdatascience.org/career-information/data-engineer/
[18] https://www.discoverdatascience.org/career-information/data-architect/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048
This work is licensed under a Creative Commons Attribution 4.0 International License.

included as part of a data science team include: blockchain engineer, data mining specialist, business intelligence analyst, etc.

Towards an AOSP Phase 1 e-Infrastructure

The following section covers:

Figure 1: e-Infrastructure needs a systems architect and experienced system support staff, data scientists and data stewards experience during various stages of the research data lifecycle

Figure 2: African Open Science Platform Ecosystem

Figure 3: Priority Domain-specific Programme (e.g. Bioinformatics/Infectious Diseases (H3ABioNet)) within the AOSP

Figure 4: Alignment of existing e-Infrastructure and Ecosystem with AOSP Phase 1 Strands

Figure 5: AOSP Phase 1 Roadmap

Figure 1: e-Infrastructure needs experienced by data scientists and data stewards during various stages of the research data lifecycle

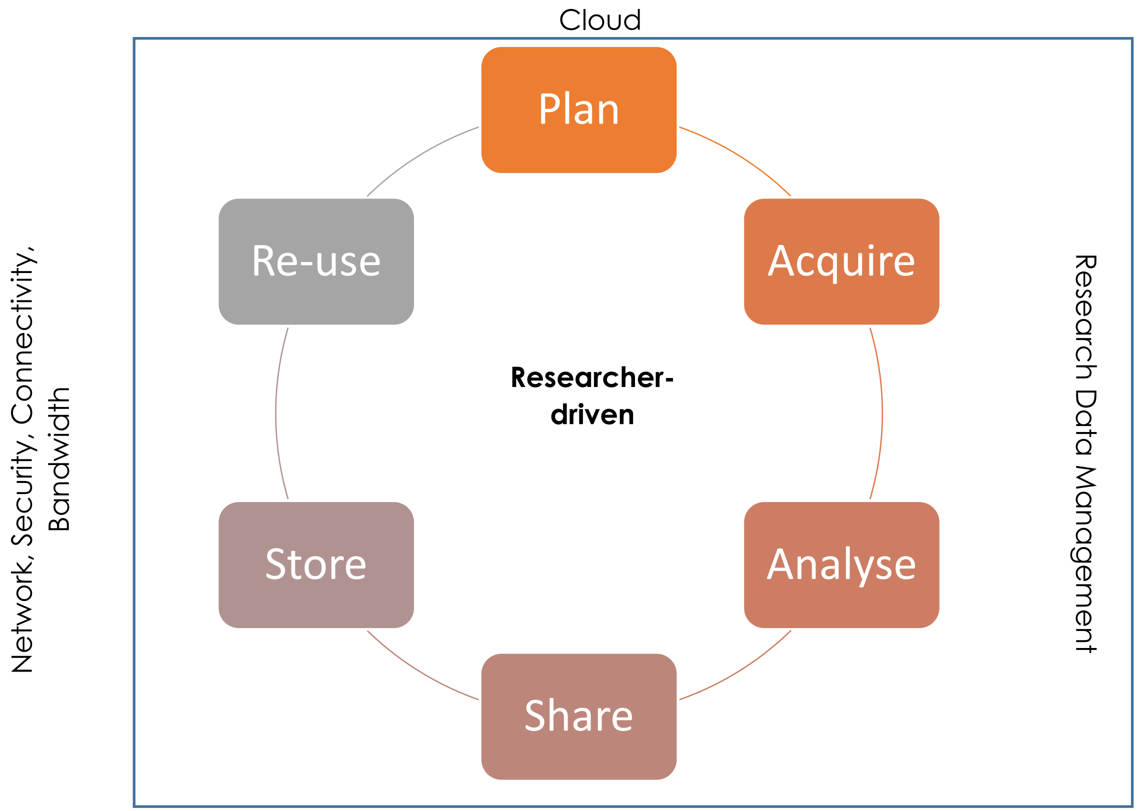
Providing e-infrastructure towards supporting and addressing needs experienced during various stages of the research data lifecycle is essential (see *Figure 1*). Primary to this lifecycle are the services to the left, which include connectivity, bandwidth, security, data transfer, and more. These services are provided by National Education and Research Networks (NRENs), maintaining and regularly upgrading their networks, also offering specialised services which are not offered by commercial ISPs.

For details about the activities and tools to accomplish the activities, refer to the *AOSP Capacity Building Framework*[19].

[19] http://dx.doi.org/10.17159/assaf.2019/0049

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048
This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 1: e-Infrastructure needs experienced by data scientists and data stewards during various stages of the research data lifecycle



© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 2: African Open Science Platform Ecosystem

Five African countries have been identified to participate in AOSP Phase 1. This is based on the performance/readiness of the respective countries based on R&D expenditure, policy development, HDI, Internet connectivity, Academies of Science, Level 6 NRENs', existing H3ABioNet nodes (bioinformatics a priority discipline), Cloud computing infrastructure and more. Each of the countries is from a different AU region: Egypt (North Africa), South Africa (Southern Africa), Ghana (West Africa), Kenya (East Africa), and Uganda (East Africa, but closest to Central Africa due to a lack of a Central African-ready country).

Stakeholders that should form part of discussions towards establishing the AOSP Phase 1 are listed, and include Academies of Science, Centres of Excellence, IP Offices, relevant Ministries, NRENs, Research Councils, Research Institutes, Research Projects, Science Granting Councils, Telecommunication Regulatory Bodies, and Universities. Also refer to the AOSP stakeholder mapping[20] and AOSP Landscape Study[21] for regional bodies that should be involved.

Collaboration on multiple levels will be required for AOSP Phase 1 to succeed, incl.: Institutional, National, Regional, Continental and Global levels.

AOSP Phase 1 proposes 6 Strands to be implemented. The strands and pilot AOSP focus areas are aligned to the left of Figure 1, and key strategies/organisations and activities have been highlighted, to indicate how it fits in as part of the AOSP ecosystem.

Figure 3: Priority Domain-specific Programme (e.g. Infectious Diseases (H3ABioNet)) within the AOSP

A number of priority data-intensive domains have been identified as part of the AOSP landscape study. Given the expertise and network developed by H3ABioNet, it is proposed that the first AOSP demonstrator focuses on bioinformatics/infectious diseases/health. H3ABioNet has for example well established nodes in the Africa countries identified. Different stakeholders will play different roles on different levels. It is expected that the overarching AOSP will play an important role in science engagement and science communication, and that it will be one of its main priorities through playing a coordinating role.

Figure 4: Alignment of existing e-Infrastructure and Ecosystem with AOSP Phase 1 Strands

The AOSPs mission is to put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for a modern society. Its building blocks are:

► a secure federated hardware, communications and software infrastructure, implementing policies that enable best practices to support open science in the digital era;

► a network of excellence in open science that supports scientists and other societal actors in accumulating and using modern data resources to maximise scientific, social and economic benefit.

These objectives will be realised through six related strands of activity:

Strand 1: A federated network of computational facilities and services.
Strand 2: Software tools and advice on policies and practices of research data management.
Strand 3: A Data Science and AI Institute at the cutting edge of data analytics.
Strand 4: Priority application programmes: e.g. cities, disease, biosphere, agriculture.
Strand 5: A Network for Education and Skills in data and information.
Strand 6: A Network for Open Science Access and Dialogue.

[20] https://atlas.mindmup.com/2019/04/51a0845065de11e9b92b8d9fa90c559e/free_mind_map/index.html
[21] http://dx.doi.org/10.17159/assaf.2019/0047

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 4 aligns the 6 objectives with the existing e-Infrastructure and ecosystem, and suggest activities to achieve the various objectives as part of AOSP Phase 1. It is further proposed that the African Research Cloud[22] or an expansion of the existing Ilifu Cloud (towards an African Open Science Cloud) be utilised for the purposes of AOSP.

Figure 5: AOSP Phase 1 Roadmap

The roadmap proposes the steps required to implement AOSP Phase 1, following the AOSP meeting in Egypt (2019).

[22] https://www.arc.ac.za/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 2: African Open Science Platform Ecosystem

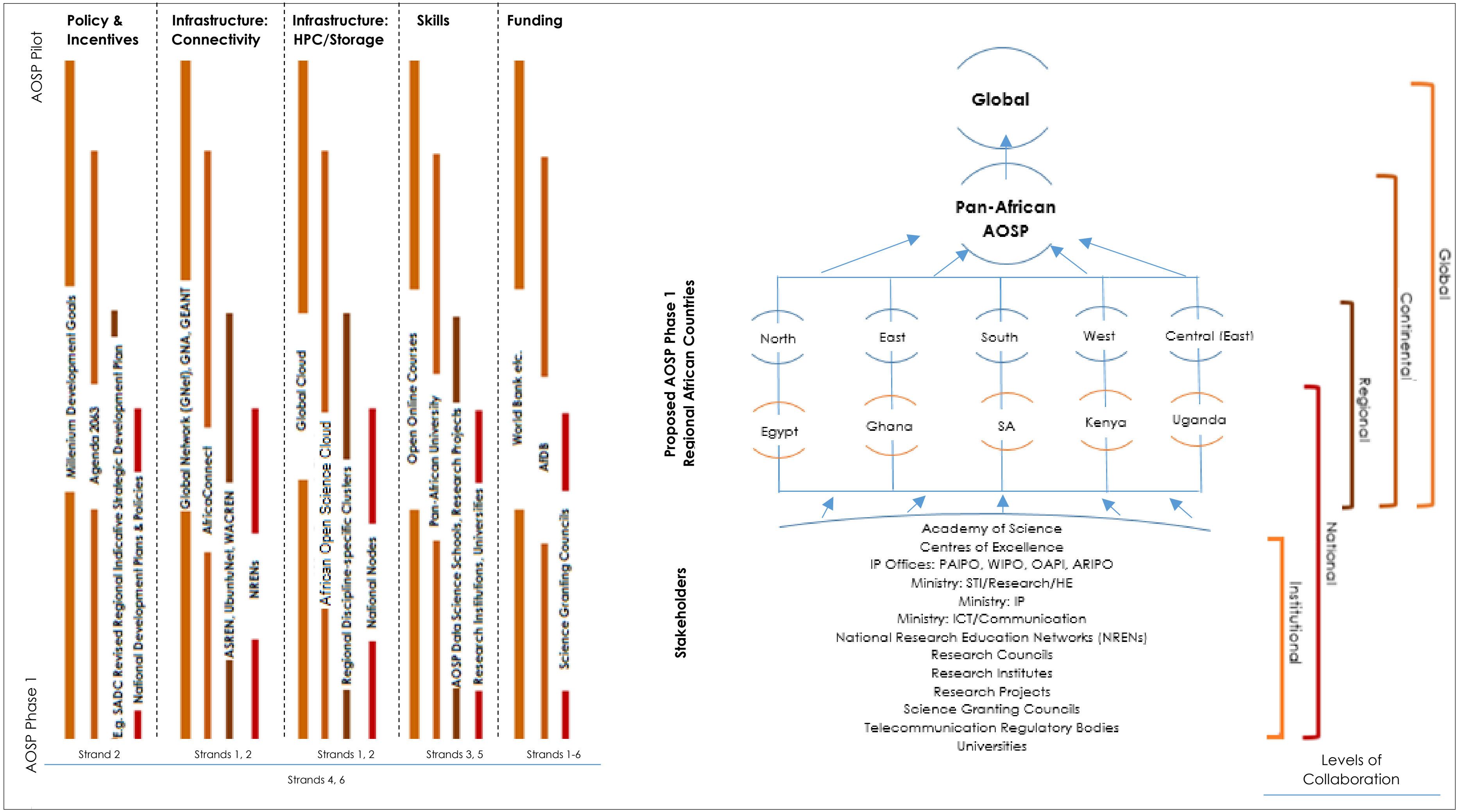
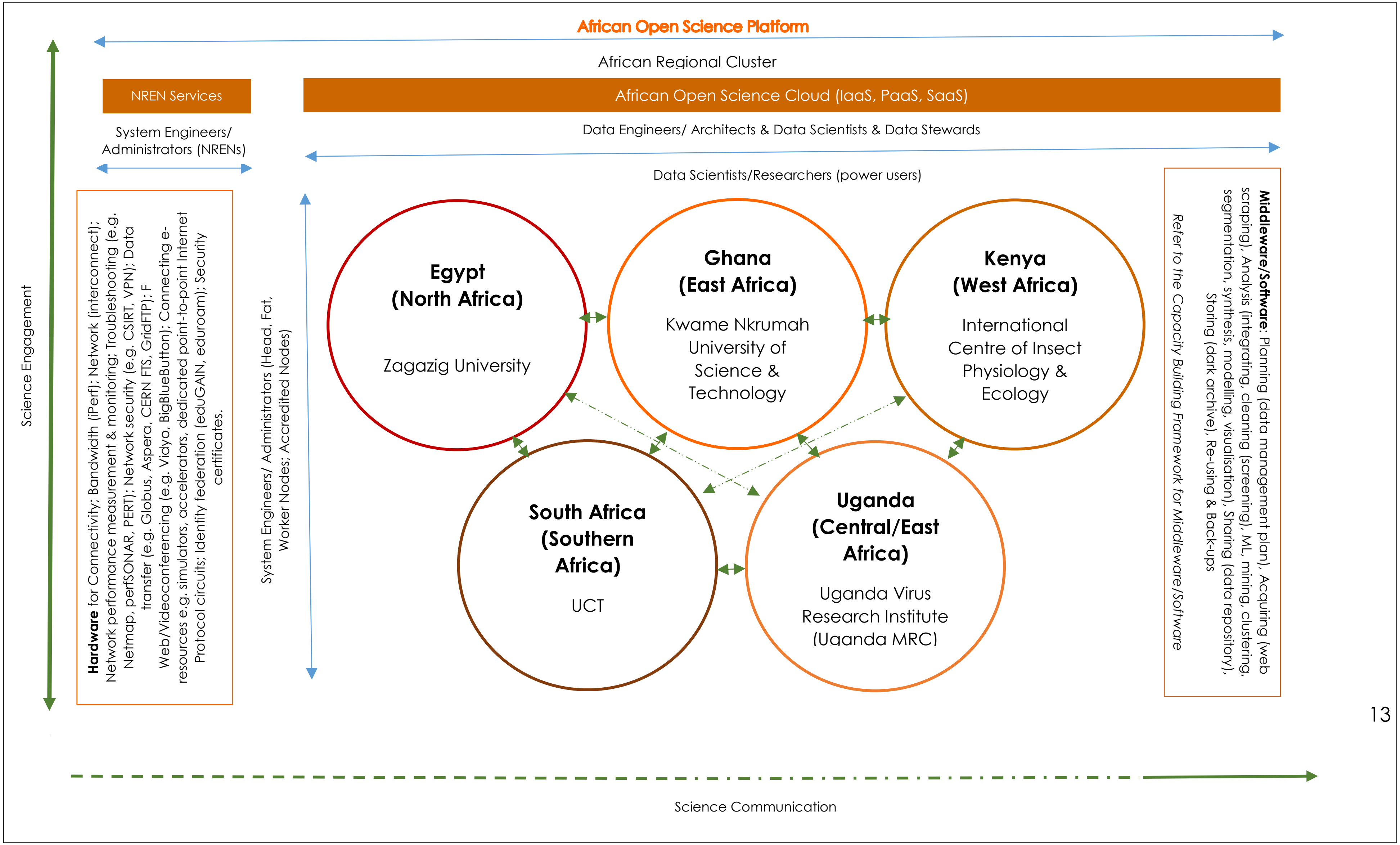

AOSP Pilot
AOSP Phase 1
Policy & Incentives
Infrastructure: Connectivity
Infrastructure: HPC/Storage
Skills
Funding
Millenium Development Goals
Agenda 2063
E.g. SADC Revised Regional Indicative Strategic Development Plan
National Development Plans & Policies
Global Network (GNet), GNA, GEANT
AfricaConnect
ASREN, UbuntuNet, WACREN
NRENs
Global Cloud
African Open Science Cloud
Regional Discipline-specific Clusters
National Nodes
Open Online Courses
Pan-African University
AOSP Data Science Schools, Research Projects
Research Institutions, Universities
World Bank etc.
AfDB
Science Granting Councils
Strand 2
Strands 1, 2
Strands 1, 2
Strands 3, 5
Strands 1-6
Strands 4, 6
Global
Pan-African AOSP
Proposed AOSP Phase 1 Regional African Countries
North
East
South
West
Central (East)
Egypt
Ghana
SA
Kenya
Uganda
Stakeholders
Academy of Science
Centres of Excellence
IP Offices: PAIPO, WIPO, OAPI, ARIPO
Ministry: STI/Research/HE
Ministry: IP
Ministry: ICT/Communication
National Research Education Networks (NRENs)
Research Councils
Research Institutes
Research Projects
Science Granting Councils
Telecommunication Regulatory Bodies
Universities
Global
Continental
Regional
National
Institutional
Levels of Collaboration

Figure 3: Priority Domain-specific Programme e.g. Bioinformatics/Infectious Diseases (H3ABioNet) within the AOSP



African Open Science Platform

African Regional Cluster

NREN Services

African Open Science Cloud (IaaS, PaaS, SaaS)

System Engineers/ Administrators (NRENs)

Data Engineers/ Architects & Data Scientists & Data Stewards

Data Scientists/Researchers (power users)

Science Engagement

Hardware for Connectivity; Bandwidth (iPerf); Network (interconnect); Network performance measurement & monitoring; Troubleshooting (e.g. Netmap, perfSONAR, PERT); Network security (e.g. CSIRT, VPN); Data transfer (e.g. Globus, Aspera, CERN FTS, GridFTP); F Web/Videoconferencing (e.g. Vidyo, BigBlueButton); Connecting e-resources e.g. simulators, accelerators, dedicated point-to-point Internet Protocol circuits; Identity federation (eduGAIN, eduroam); Security certificates.

System Engineers/ Administrators (Head, Fat, Worker Nodes; Accredited Nodes)

Egypt (North Africa)
Zagazig University

Ghana (East Africa)
Kwame Nkrumah University of Science & Technology

Kenya (West Africa)
International Centre of Insect Physiology & Ecology

South Africa (Southern Africa)
UCT

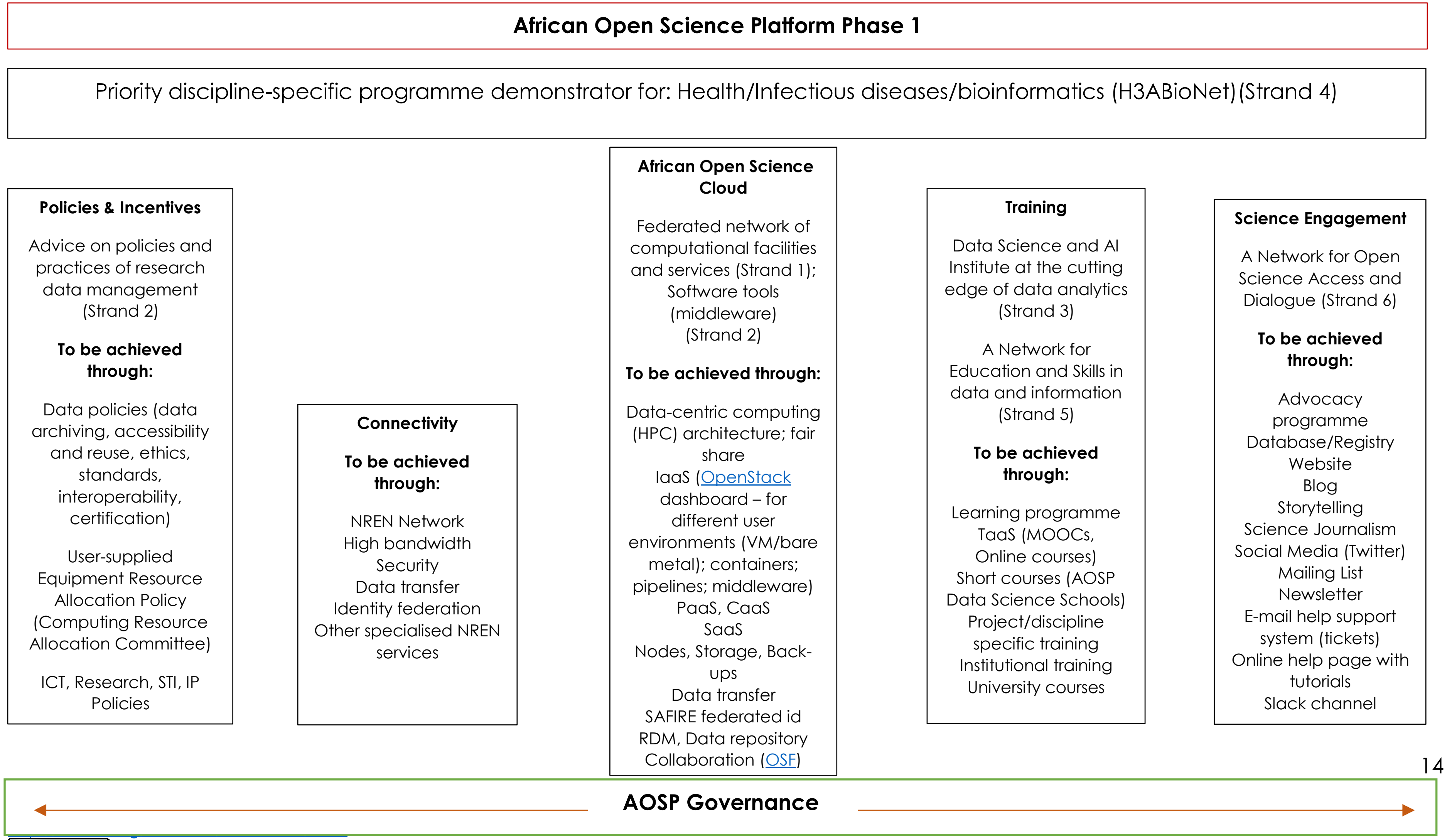
Uganda (Central/East Africa)
Uganda Virus Research Institute (Uganda MRC)

Middleware/Software: Planning (data management plan), Acquiring (web scraping), Analysis (integrating, cleaning (screening), ML, mining, clustering, segmentation, synthesis, modelling, visualisation), Sharing (data repository), Storing (dark archive), Re-using & Back-ups

Refer to the Capacity Building Framework for Middleware/Software

Science Communication

Figure 4: Alignment of existing e-Infrastructure and Ecosystem with AOSP Phase 1 Strands



African Open Science Platform Phase 1

Priority discipline-specific programme demonstrator for: Health/Infectious diseases/bioinformatics (H3ABioNet) (Strand 4)

Policies & Incentives

Advice on policies and practices of research data management (Strand 2)

To be achieved through:

Data policies (data archiving, accessibility and reuse, ethics, standards, interoperability, certification)

User-supplied Equipment Resource Allocation Policy (Computing Resource Allocation Committee)

ICT, Research, STI, IP Policies

Connectivity

To be achieved through:

NREN Network
High bandwidth
Security
Data transfer
Identity federation
Other specialised NREN services

African Open Science Cloud

Federated network of computational facilities and services (Strand 1); Software tools (middleware) (Strand 2)

To be achieved through:

Data-centric computing (HPC) architecture; fair share
IaaS (OpenStack dashboard – for different user environments (VM/bare metal); containers; pipelines; middleware)
PaaS, CaaS
SaaS
Nodes, Storage, Back-ups
Data transfer
SAFIRE federated id
RDM, Data repository
Collaboration (OSF)

Training

Data Science and AI Institute at the cutting edge of data analytics (Strand 3)

A Network for Education and Skills in data and information (Strand 5)

To be achieved through:

Learning programme
TaaS (MOOCs, Online courses)
Short courses (AOSP Data Science Schools)
Project/discipline specific training
Institutional training
University courses

Science Engagement

A Network for Open Science Access and Dialogue (Strand 6)

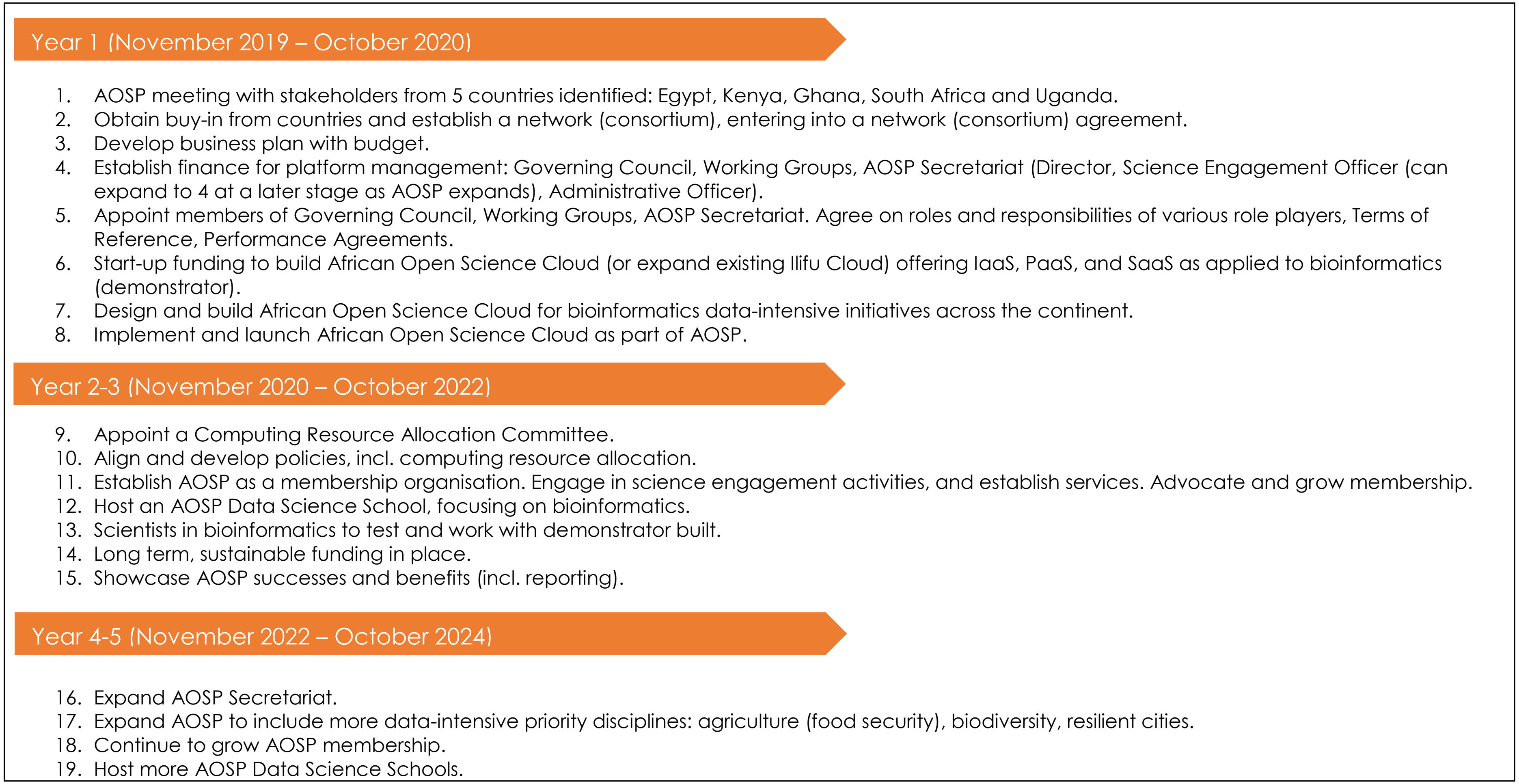
To be achieved through:

Advocacy programme
Database/Registry
Website
Blog
Storytelling
Science Journalism
Social Media (Twitter)
Mailing List
Newsletter
E-mail help support system (tickets)
Online help page with tutorials
Slack channel

AOSP Governance

This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 5: Roadmap AOSP Phase 1



Year 1 (November 2019 – October 2020)

1. AOSP meeting with stakeholders from 5 countries identified: Egypt, Kenya, Ghana, South Africa and Uganda.
2. Obtain buy-in from countries and establish a network (consortium), entering into a network (consortium) agreement.
3. Develop business plan with budget.
4. Establish finance for platform management: Governing Council, Working Groups, AOSP Secretariat (Director, Science Engagement Officer (can expand to 4 at a later stage as AOSP expands), Administrative Officer).
5. Appoint members of Governing Council, Working Groups, AOSP Secretariat. Agree on roles and responsibilities of various role players, Terms of Reference, Performance Agreements.
6. Start-up funding to build African Open Science Cloud (or expand existing Ilifu Cloud) offering IaaS, PaaS, and SaaS as applied to bioinformatics (demonstrator).
7. Design and build African Open Science Cloud for bioinformatics data-intensive initiatives across the continent.
8. Implement and launch African Open Science Cloud as part of AOSP.

Year 2-3 (November 2020 – October 2022)

9. Appoint a Computing Resource Allocation Committee.
10. Align and develop policies, incl. computing resource allocation.
11. Establish AOSP as a membership organisation. Engage in science engagement activities, and establish services. Advocate and grow membership.
12. Host an AOSP Data Science School, focusing on bioinformatics.
13. Scientists in bioinformatics to test and work with demonstrator built.
14. Long term, sustainable funding in place.
15. Showcase AOSP successes and benefits (incl. reporting).

Year 4-5 (November 2022 – October 2024)

16. Expand AOSP Secretariat.
17. Expand AOSP to include more data-intensive priority disciplines: agriculture (food security), biodiversity, resilient cities.
18. Continue to grow AOSP membership.
19. Host more AOSP Data Science Schools.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

Bibliography

African Open Science Platform (AOSP). (2018). *The Future of Science and Science of the Future: Vision and Strategy for the African Open Science Platform (v02) (online).* Available at: http://doi.org/10.5281/zenodo.2222418 *(Accessed on 16 May 2019)*

Foley, M. (2016). *The Role and Status of National Research and Education Networks (NRENs) in Africa (online)*. Available at: http://documents.worldbank.org/curated/en/233231488314835003/pdf/113114-NRENSinAfrica-SABER-ICTno05.pdf *(Accessed on 16 May 2019)*

Greaves, D. (2009). An NREN Capability Maturity Model (online). Available at: https://repository.ubuntunet.net/handle/10.20374/69 *(Accessed 0n 16 May 2019)*

International Telecommunication Union (ITU). (2018). *Measuring the Information Society Report Volume 1 (online).* Available at: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2018/MISR-2018-Vol-1-E.pdf *(Accessed on 16 May 2019)*

Marcus, A. (2015). *Data and the fourth industrial revolution (online).* Available at: https://www.weforum.org/agenda/2015/12/data-and-the-fourth-industrial-revolution/ *(Accessed on 16 May 2019)*

Southern African Development Community (SADC). (2016). *Cyberinfrastructure Framework (online).* Available at: https://drive.google.com/drive/u/0/folders/1RhzWdy3HPGtVSMakk3bhczlC6PFSUUjy. *(Accessed on 16 May 2019)*

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) e-Infrastructure Framework and Roadmap: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0048

This work is licensed under a Creative Commons Attribution 4.0 International License.

African Open Science Platform (AOSP) Incentives Framework

Fostering a culture of Open Data within African National Systems of Innovation

Developed by Louise Bezuidenhout[1], Ina Smith[2], and Susan Veldsman[2].

[1]Institute for Science, Innovation and Society, University of Oxford, United Kingdom & Steve Biko Centre for Bioethics, University of the Witwatersrand, South Africa
[2]Academy of Science of South Africa

Rationale for a Framework on Incentives

During recent years, there's been a shift from "publishing as fast possible" to "sharing knowledge as early possibly", and without any barriers to access. Science activities are increasingly becoming more collaborative, transparent, reproducible, and publicly available, through sharing digital technologies & utilising collaborative tools. Open Science (OS) rests on three key pillars: Open Access (OA) to scientific literature, Open Data (OD), and open engagement with society. OS is both an ideological commitment to the just distribution of resources, and a practical commitment to use evolving digital research tools. It represents a new vision for African research, whereby resources can be effectively shared so as to foster a globally competitive, yet distinctly African, style of research. Research stakeholders are offered a unique opportunity to develop a coordinated and comprehensive strategy for embedding openness in research in a manner that represents the priorities and preferences of the African continent.

The *AOSP Incentives Framework* provides guidelines to national science agencies, incl. policy and decision-makers from countries in Africa, towards identifying openness as a key component of responsible research practices, making a compelling case for how Open Data activities can enhance research quality, increase research and researcher visibility, and have higher impact. For this to happen, data needs to be open in an intelligible way. It is imperative that national governments recognise the value to be gained from intelligible Open Data, for national science agencies to adopt a coordinating role, for science policy and decision-makers to set incentives for openness from universities and research institutions, for these institutions to support Open Data processes by their researchers, and for learned societies that articulate the priorities and practices of their disciplines, to advocate and facilitate Open Data processes as important priorities.

Focus and Scope

The *AOSP Incentives Framework* provides guidelines to continental, regional, national, and institutional stakeholders, national science agencies and policy and decision-makers, to create a research environment conducive for the sharing of research data (Open Data), to develop comprehensive regulatory frameworks to govern research data (Open Data), at the same time implementing Open Science practices.

This framework speaks to multiple stakeholders involved in African science, including researchers, librarians, and academic institutions, funding bodies, publishers, National Research Education Networks (NRENs) and national governments.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

Challenges

- Physical, regulatory, social barriers
- Universities not adapting curricula and performance appraisal systems fast enough
- Authorship – no acknowledgment for work by data scientists and data stewards
- Shortage of resources
- Lack of awareness, skill shortages and shortage of training
- Diversity of practices and policies
- Intellectual property concerns
- Ethical concerns
- Enhanced infrastructural challenges
- Lack of hardware and software
- Traditional research cultures

Opportunities

- Tracking the African digital research footprint, and the impact of African research
- Rise in measurable Open Data events within African academia
- Changing metrics of evaluation used in Africa
- Increase in resources dedicated to Open Data activities
- Increased visibility of African research
- Online and re-use of African research outputs
- Growth of the active African Open Data community
- Changing the way African science is practiced
- Development of robust Open Data infrastructures
- Use OD and OA research output for higher social returns, stimulating innovation, growing the economy

Incentivising Open Data – from Theory to Action

General Actions

- Raise awareness of the benefits of Open Data amongst African research stakeholders
- Contextualise Open Data discussions and engage all potential data users
- Promote understandings of data collections as assets
- Collect and curate positive examples of "Open Data Champions"
- Encourage a range of Open Science/Open Data activities
- Develop informal cultures of Open Data in research environments
- Ensure that all research stakeholders are involved in data policy development
- Integrate data skills teaching in all levels of tertiary academic teaching
- Engage the public in Open Data activities
- Develop continental, regional, national and institutional cultures of mutual support
- Address challenges:
    - Systems of credit attribution that reflect Open Data contributions
    - Shortage of resources
    - Diversity of practices and policies
    - Intellectual property
    - Authorship and data curation (stewardship)
    - Ethical issues
    - Infrastructure, hardware and software

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

Specific Actions

Action	Discussion/Guidelines
Create an enabling environment for Open Science, through restructuring National Systems of Innovation (NSIs)	• Explore global, continental, regional, national, and institutional Open Science initiatives, e.g. Plan S, AmeliCA. • Where in agreement, national governments can benefit from membership of a continental-wide coordination initiative such as AOSP. • Build on existing open government data initiatives e.g. SHaSA (Strategy for the Harmonisation of Statistics in Africa) (an initiative by the UN Economic Commission for Africa (UNECA)), the African Development Bank (AfDB), and the African Union (AU)), AFRISTAT, and the Pan-African Institute for Statistics (STATAFRIC); Open Government Partnership. • Re-visit existing national metrics for assessing and awarding research output and research impact by institutions and individual researchers. • Promote a collaborative culture of openness vs a traditional/conservative/competitive academic culture. • Address the position of indigenous language/s vs English as scientific language. • Address conflicting or incomplete national, institutional and funder policies. • Engage all stakeholders in national policy formulation – foster a sense of “ownership” among all; engage neutral parties to oversee agenda-setting of these meetings; precede meetings on national level with institutional/discipline-specific workshops. • Engage academia in rolling out new systems (testing and development). • Target institutional management. • Make provision for social science input into natural science. • Make use of existing resources to develop policy, design infrastructure, and conduct training. • Regard data collections as assets – to attract funding and collaboration, and increase global visibility. • Conduct national scoping exercises, collecting examples/case studies on how data have added value, benefitted all. • Establish a network of champions, including all stakeholders. • Encourage the development of trusted data science communities, across disciplines, sharing experiences, finding solutions – also informal. Across stakeholder groups. Foster learning networks between institutions to circulate learning experiences and best practices. • Host national competitions (awards), hackathons, and involve private industry. • Link Open Data to networking and public engagement activities. • Embed Open Data in institutional definitions of “responsible research”. • Make openness an aspirational character trait for institutional staff and students. • Establish support systems within governmental and non-governmental bodies. • Set an example on national level. Then encourage data sharing on institutional level.
Create an enabling environment for Open Science, through developing, aligning and implementing related	• National & Institutional Policies – alignment required between: o Intellectual Property Rights (IPR) Policy

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

policies, towards a legal/regulatory framework conducive for data sharing	o Information and Communications Technology (ICT) Policy (incl. Internet connectivity/bandwidth) o Higher Education (HE) Policy o Science, Technology and Innovation Policy (incl. regulation of data; ethical guidelines; research data management (RDM)) o Research Output Policy o Research Ethics Policy • Provide guidance towards strategic planning on institutional level and implementation of national policies. • Make available required infrastructure, resources to enable science community to implement policy. • Policy to determine which data to curate, since curation is expensive (incl. Big Data vs "long tail" small data; raw data vs processed data, etc.). • Align expectations in terms of: o Data availability and accessibility: commit users to release data with as few restrictions as possible in a timely and responsible manner o Data management: identify relevant standards and community best practice o Data re-use and discoverability: highlight expectations regarding metadata, IP as applied to metadata etc. o Legal, ethical and commercial issues: take into consideration, explain and endorse any constraints on the release of research data o Data citation: commit users to properly credit intellectual contributors, data sources, terms etc. o Efficiency and cost-effectiveness: highlight the importance of maximising the research benefit which can be gained from budgets; the mechanisms for the research activities should be efficient and cost-effective in the use of public funds. • Policies should reflect values such as openness, collegiality, quality and community. It must also support associated RDM values, such as flexibility, transparency, legal conformity and accountability, together with the FAIR principles of interoperability and reproducibility. • Differentiate between policies in terms of personal funds used, and public funds used. Consider when data can be commercialised to generate an income. • Address conflicts between national, institutional, publisher and funder policies and requirements (incl. IP and licensing) • Involve IP stakeholders on continent e.g. WIPO, OAPI, ARIPO PAIPO. • Acknowledge and address ethical concerns about privacy, consent, security, or the possible causing of harm when data is shared, as part of policy. • Prepare – through policy and regulation for the new challenges faced as part of the 4IR: unanticipated data re-use, impact of linking disparate data sets, Big Data mining, use of surveillance technologies, and Artificial Intelligence (AI) raising ethical concerns.
Make provision for funding	• Funders on all levels: Global, Continental, National, Institutional, Private level. • Funders to make provision for costs regarding data curation (maintenance, cleaning, annotating, storing, etc.), ICT infrastructure, resources (incl. maintenance & upgrades, hardware, software).

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

	- Make funding available for researcher, technical staff, librarian (data stewardship) training (capacity building/skills). - Citizen science outreach initiatives require funding. - Funders to include FAIR sharing of data as a set of evaluation criteria, making it a strong requirement in all proposals for funding. - If researchers do not comply, they become temporarily ineligible for further funding.
Create awareness	- Workshops on national & institutional levels, for all stakeholders; precede national workshops with institutional/discipline-specific workshops. - Stakeholders in turn to also create awareness and promote OS practices – practice what is preached. - Host disciplinary forums. - Conduct formal & informal training. - Make anonymous help-lines/chatbots available. - Collect and share positive examples of the impact of data sharing – consider storytelling. - Share benefits of open data/data sharing. - Accommodate restricted/controlled/mediated access to data e.g. in health. - Create awareness of free and open source tools available. - Utilise webinars & recordings thereof to create awareness. - Address fears re scooping, unethical data re-use, lack of attributing/crediting.
Build capacity and increase data skills	- Integrate data skills as part of curricula, across all university courses. Build on existing OA training material/courses, and contextualise for local use. - Build data skills in as part of Continuous Professional Development (CPD). - Online training: utilise Learning Management Systems (LMSs) and MOOCs (Massive Open Online Courses). - Curriculum should be flexible, and make provision for practical training/learning. - Values and norms of Open Science (incl. collegiality, responsibility, justice, and equity) should be taught. - Address fears re scooping, unethical data re-use, lack of attributing/crediting, privacy, consent, security, or the possible causing of harm. - Teach ethics of data usage & authorship (citation, credit, attribution). - Teach Research Data Management (RDM) – see *the AOSP RDM Framework*[1]. - Address FAIR (findable, accessible, interoperable, re-usable) data principles as part of all curricula. - Introduce and create awareness on the use of data tools (incl. collection, cleaning, analysis, visualisation/modelling, sharing/publishing) – also free and open source (FOSS) tools. Reduce reliance on expensive proprietary software. - Equip Research Ethics Committees (RECs) to deal with complexities of data sharing, incl. Big Data and "long tail" small data sets. - Partner more traditional researchers with researchers practicing OS, sensitising and supporting the transition. - Train on self-archiving, standards for the identification, formatting and curation of data and metadata to make

[1] http://dx.doi.org/10.17159/assaf.2019/0050

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051
This work is licensed under a Creative Commons Attribution 4.0 International License.

	data reusable, publishing venue and related licensing, metrics and acknowledging and crediting the reuse. • Standards: agree on adhering to the "acceptable minimum" standards, and not necessarily "best practice". • Create awareness on open journals and open standards. • Implement and provide training on persistent identifiers (DOIs or digital object identifiers) and ORCiDs. • Create awareness on where to share data, and how to identify a trusted and sustainable data repository: see DATAD-R[2], FAIRsharing[3], re3data[4] , CoreTrustSeal[5].
Recognise and award data sharing	• Identify ways in which Open Data practices can be incentivised within African science. • Incentivise intellectual property (IP) creators at public research institutions to disclose IP to their institution's office of technology transfer (OTT) at the point of publication, as early as possible, once the researcher has exploited the data for his/her own research and objectives have been met. • Add data sharing as a criteria to key performance indicators in institutional evaluations. • Re-visit national existing metrics for assessing research impact. Currently metrics are very much paper-focused, it overlooks data sharing and undermines OS. Adapt metrics to measure re-use of Open Data, as well as the sharing of Open Data. • Citation-based metrics can be used for data published in data journals (measuring use of persistent identifiers/DOIs). • Alternative data-level metrics are needed to measure attention and uptake of a dataset (e.g. HuMetrics[6], FAIR Metrics[7], Make Data Count[8] initiatives). • Utilise scientometrics to track authorship, data re-use and other indicators of visibility, collaboration and re-use, open access to publications, open data, open peer review, research integrity, citizen science and stakeholder engagement. • Avoid the exclusive use of quantitative assessment mechanisms such as impact factors in assessment processes. • Avoid exploitation of the national data sharing policy through data dumping or excessive publication. • Put national systems in place which makes it easy to report for institutions and researchers. • Adapt the Open Science Career Assessment Matrix (OS-CAM) (o'Carroll, Rentier, et al., 2017) as a template to address African research recognition (see *AOSP Incentives Framework*[9] p. 49). Include all stakeholders during the adoption process.
Provide reliable ICT infrastructure and resources	• Data sharing requires financial, infrastructural and human (time, expertise) resources. • When data collections are regarded as assets, it justifies the commitment of resources as a means of accruing reputational and collaborative benefits that will raise the institutional profile.

[2] Database of Theses and Dissertations and Research http://datad.aau.org/

[3] https://fairsharing.org/

[4] https://www.re3data.org/

[5] https://www.coretrustseal.org/

[6] https://humetricshss.org/blog/humetrics-values/

[7] http://fairmetrics.org/

[8] https://makedatacount.org/

[9] https://drive.google.com/open?id=1qV7Bu2MGetCTbGw2d2yPn3jzE3P8-b6P

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

	• Data curation is expensive, and therefore policies should address which data to curate. • Governments, institutions and researchers should plan, coordinate, and align the use of resources, for optimal benefit and return on investment. • Develop metrics for resource investment. • Publishing platform standards: OA journal and OD repository standards, to guide institutions and stakeholders, and align with global standards. • Make provision for Variety, Velocity, Volumes of data, across disciplines. • Big Data requires sufficient bandwidth, and stable & reliable Internet. Selected areas in Africa struggle with ageing and unreliable power infrastructure and frequent power outages. • Refer to the *AOSP e-Infrastructure Framework*[10], for guidelines on establishing an ICT infrastructure conducive for data sharing.
Include citizens as stakeholders through community engagement, science communication efforts	• Create opportunities for citizens to engage with research data and academics, e.g. during national science forums, institutional science forums. Increase trust in science. • Demonstrate how private industry benefit from Open Data, yielding benefits in terms of economy, job creation and innovation. • Involve citizens in data collection efforts. • Demonstrate return on investment in R&D for public, and how they benefit. • Engage with school learners e.g. during national science week, science festivals, outreach activities. Foster data literacy and an appreciation for Open Science.

[10] http://dx.doi.org/10.17159/assaf.2019/0048

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

Open Science Career Assessment Matrix (OS-CAM)

The Open Science Career Assessment Matrix (o'Carroll, Rentier, et al., 2017)

Open Science Activities	Possible Evaluation Criteria
RESEARCH OUTPUT	
Research Activity	Pushing forward the boundaries of OS as a research topic
Publications	Publishing in OA journals registered with DOAJ[11] Self-archiving research output in OA institutional repositories
Datasets and Research Results	Using the FAIR data principles Adopting quality standards in OD management and open datasets Making use of OD by other researchers
Open Source	Using open source software and other open tools Developing new software and tools that are open to other users
Funding	Securing funding for OS activities
RESEARCH PROCESS	
Stakeholder Engagement/ Citizen Science	Actively engaging society and research users in the research process Sharing provisional research results with stakeholders through open platforms Involving stakeholders in peer-review process
Collaboration & Interdisciplinarity	Widening participation in research through open collaborative projects Engaging in team science through diverse cross-disciplinary teams
Research Integrity	Being aware of the ethical and legal issues relating to data sharing, confidentiality, attribution, and environmental impact of OS activities Fully recognising the contribution of others in research projects, including collaborators, co-authors, citizens, OD providers
Risk Management	Taking account of the risks involved in OS
SERVICE & LEADERSHIP	
Leadership	Developing a vision and strategy on how to integrate OS practices in the normal practice of doing research Driving policy and practice in OS Being a role model in practicing OS
Academic Standing	Developing an international or national profile for OS activities Contributing as editor or advisor for OA or OS journals or bodies
RESEARCH IMPACT	
Communication & Dissemination	Participating in public engagement activities Sharing research results through non-academic dissemination channels Translating research into a language suitable for public understanding
IP (Patents, Licenses)	Being knowledgeable on the legal and ethical issues relating to IPR Transferring IP to the wider community
Societal Impact	Evidence of use of research by societal groups Recognition from societal groups or for societal activities
Knowledge Exchange	Engaging in open innovation with partners beyond academia
TEACHING & SUPERVISION	
Teaching	Training other researchers in OS principles and methods Developing curricula and programmes in OS methods, incl. OS management Raising awareness and understanding in OS in undergraduate and masters' programmes
Mentoring	Mentoring and encouraging others in developing their OS capabilities
Supervision	Supporting early stage researchers to adopt an OS approach
PROFESSIONAL EXPERIENCE	
Continuing Professional Development (CPD)	Investing in open professional development to build OS capabilities
Project Management	Successfully delivering OS projects involving diverse research teams
Personal Qualities	Demonstrate the personal qualities to engage society and research users with OS Showing the flexibility and perseverance to respond to the challenges of conducting OS

[11] Directory of Open Access Journals https://doaj.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

References

Bezuidenhout, L. (2018). To share or not to share … incentivizing data sharing in life science communities, *Developing World Bioethics*. John Wiley & Sons, Ltd (10.1111). doi: 10.1111/dewb.12183.

Bezuidenhout, L. (2019). *Incentivising data sharing within African academia*. url: https://drive.google.com/open?id=1qV7Bu2MGetCTbGw2d2yPn3jzE3P8-b6P.

o'Carroll, C., Rentier, B., Cabello Valdes, C., Esposito, F., Kaunismaa, E., Maas, K., Metcalfe, J., McAllister, D. and Vandervele, K. (2017). *Evaluation of Research Careers Fully Acknowledging Open Science Practices*. Brussels. doi: 10.2777/75255.

OECD (2004). *OECD Declaration on Open Access to Publicly Funded Data*.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Incentives Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0051

This work is licensed under a Creative Commons Attribution 4.0 International License.

African Open Science Platform (AOSP) Open Data Policy Framework

Fostering a culture of Open Data within African National Systems of Innovation

Developed by Joseph Wafula[1], Ina Smith[2], Susan Veldsman[2], Paul Uhlir[3] and Audrey Masizana[4]

[1]ICT Centre of Excellence & Open Data (iCEOD), Jomo Kenyatta University of Agriculture and Technology, Kenya
[2]Academy of Science of South Africa
[3]Consultant on Information Policy & Management, USA
[4]University of Botswana

Rationale for a Framework on an Open Data Policy

Open Science is a combination of concepts, tools, platforms and media to promote the creation and dissemination of knowledge in free, open and more inclusive ways, to enable reaping of wider research benefits (Picarra, 2016). The FOSTER Project[1] defines Open Science as "… 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 re-use, redistribution and reproduction of the research and its underlying data and methods." According to OCSDNet[2] "Open Science moves beyond Open Access research articles, towards encompassing other research objects such as **data**, **software codes**, **protocols** and **workflows**. The intention is for people to use, re-use and distribute content without legal, technological or social restrictions. In some cases, Open Science also entails the opening up of the entire research process from agenda-setting to the dissemination of findings."

Open Data – being one of a few Open Science branches[3] – refer to "data that can be freely used, shared and built-on by anyone, anywhere, for any purpose" (Open Knowledge International[4]). This includes both government data and research data.

For society to benefit from research, research output and the underlying research data need to be accessible, so that it can be discussed, challenged, reproduced, tested, scrutinised, and enhanced, but also so that all of society can equally benefit from the resulting scientific discoveries. New research builds on established results from previous research (Schiltz, 2018). Lack of data management and publication paywalls for both the research data and literature have the opposite effect, creating barriers to science and slowing down scientific discovery, and increasing the existing digital divide between North and South even further.

Governments – through policies – have the opportunity to change the negative effect paywalls have on restricting access to research, and to prevent research data to fall into the exact locked down position many scholarly articles in published journals find themselves in. The *AOSP Open Data Policy Framework* provides guidelines towards implementing Open Data policies, for consideration by governments and research institutions. Up to date, there is no real evidence that national-level Open Data policies have been finalised and implemented on the African continent. However, selected African countries such as Botswana, Kenya, Madagascar, Mauritius, South Africa, and Uganda have made some progress towards developing policies and strategies for Open Data. In addition there exists a positive development resulting from the *African Union African Observatory of Science and*

[1] https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition
[2] https://ocsdnet.org/about-ocsdnet/about-ocs/
[3] https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition
[4] https://blog.okfn.org/2013/10/03/defining-open-data/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

Technology Indicators (AOSTI)[5] initiative, which endeavours to assist African countries to build capacity for STI policy activities and initiatives. The development is articulated in a report on the *Assessment of Scientific Production in the African Union, 2005–2010* (2014), recommending "creating open and free access publication outlets for Africa, with improved review committees", and highlighting the challenge of high article fee requirements for publishing in citation-indexed journals and the high subscription prices to commercially available databases. Although this applies to publications specifically, research data can potentially form part of this strategy.

Furthermore, Science Granting Councils on the African continent have confirmed – through a statement released as an outcome of the *Science Granting Councils Initiative in Sub-Saharan Africa Annual Forum, Global Research Council Africa Regional Meeting, 5-8 November 2018*[6]:

1) their commitment to support and advocate for the development and use of Open Science platforms that widen access to knowledge and allow integrated problem solving at a potentially transformative (as opposed to incremental) scale, and
2) to commit funding towards the development of the human capital necessary for leveraging the potential of Big Data, as well as invest in the infrastructure required for implementing Open Science platforms.

Open Data being one branch of "Open Science", as explained in the FOSTER taxonomy[7] of Open Science, it is therefore assumed that this statement also applies to Open Data.

Representatives of the African Open Science Platform, AmeLICA, cOAlition S, OA2020, and SciELO – five of the major worldwide Open Access initiatives – met on 01 May 2019 during the annual meeting of the Global Research Council (GRC) in São Paulo. It was during this meeting that a further statement on Open Access was issued, namely the São Paulo Statement on Open Access[8].

In 2004 science ministers from all nations of the Organisation for Economic Co-operation and Development (OECD)[9] signed a declaration to ensuring opening access to publicly funded archived data. On a global level, the OECD *Principles and Guidelines for Access to Research Data from Public Funding* (OECD 2004)[10] provide broad policy recommendations to the governmental science policy and funding bodies of member countries (currently adherence by three non-member African countries, including Egypt, Morocco and Tunisia), on access to and management of research data from public funding. The ultimate goal of these Principles and Guidelines is to improve the efficiency and effectiveness of the global science system.

National Open Data policies should be positioned within a broader regulatory framework, amidst policies regarding intellectual property rights (IPRs) (through work done by WIPO[11], PAIPO[12], OAPI[13] and ARIPO[14]), research policies (including ethics and research output), policies for STI, funding policies, and ICT policies. Related policies are however not always aligned with one another, lacking regulatory convergence/harmonisation of national policies. Although there are efforts towards aligning IP and STI policies on the continent with one another and with international policies, African countries still have a long way to go.

Lastly – for Open Data policies to succeed, researchers should be incentivised, training should be provided, and an enabling ICT infrastructure has to be established within which high quality, trusted data can be curated, and research can be conducted and published. Unfortunately African

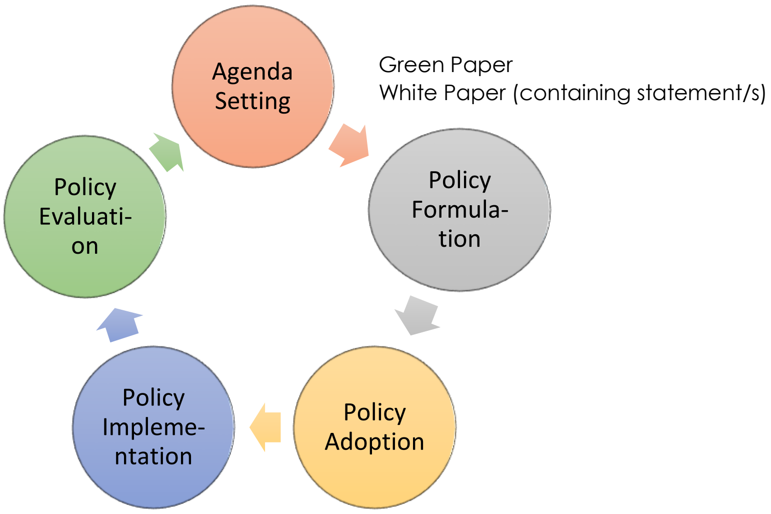
[5] http://aosti.org/
[6] https://www.nrf.ac.za/media-room/news/first-forum-science-granting-councils-initiative-sub-saharan-africa
[7] https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition
[8] https://www.nrf.ac.za/media-room/news/s%C3%A3o-paulo-statement-open-access
[9] http://www.oecd.org/
[10] http://www.oecd.org/sti/inno/38500813.pdf
[11] World Intellectual Property Organization https://www.wipo.int/portal/en/index.html
[12] Pan-African Intellectual Property Organization https://au.int/en/node/32549
[13] Organisation Africaine de la Propriété Intellectuelle http://oapi.int/index.php/fr/
[14] African Regional Intellectual Property Organization http://www.aripo.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052
This work is licensed under a Creative Commons Attribution 4.0 International License.

countries invest far too little in ICT, STI and R&D. As far as R&D concern, only 2 of the 54 African countries contribute 0.8% of its GDP (Gross Domestic Product) to R&D (UNESCO Institute for Statistics[15]), which is the closest to the AU's (African Union's) suggested 1%. This trend is attributed to the absence of political will, a lack of resources and priorities differing from one country to another.

Focus and Scope

The purpose of the *AOSP Open Data Policy Framework* is to guide on development of Open Data policies and practices at continental, regional, national and institutional level. All policy and decision-makers (incl. Policy Advisory Committees) at all levels are encouraged to use this document to direct Open Data policy through the various stages of the policy lifecycle, visualised as follows:



It is important to note that a policy can be preceded by a Green Paper (a tentative government report and consultation document of policy proposals for debate and discussion) and a White Paper (communicating governments' official position and statement of emerging or uncodified government policy).

The following three policy/strategy examples are intended to be instructive:

- Ireland National Open Research Forum (NORF) National Statement on the Transition to an Open Research Environment (draft, 2018)[16]
- Jomo Kenyatta University of Agriculture and Technology Open Data Policy[17]
- Australian Code for the Responsible Conduct of Research 2007[18]

Although it will vary from one country to another, depending on the availability of resources, a roadmap is suggested to include the following:

- Stage 1: Identify the problem, and set the agenda
- Stage 2: Establish a national forum, involving all key stakeholders
- Stage 3: Kick-off meeting
- Stage 4: Develop a Green Paper
- Stage 5: Stakeholder consultation

[15] http://uis.unesco.org/apps/visualisations/research-and-development-spending/
[16] http://norf-ireland.net/wp-content/uploads/2018/11/NORF-National-Statement-on-Transitioning-to-an-Open-Research-Environment_Public.pdf
[17] https://drive.google.com/file/d/0B1DCCKRXRF8ydHR5a2QtOWk3Q3M/view?usp=sharing
[18] https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2007

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052
This work is licensed under a Creative Commons Attribution 4.0 International License.

- Stage 6: Finalise and publish Green Paper
- Stage 7: Establish a Policy Advisory Committee
- Stage 8: Develop a White Paper
- Stage 9: Stakeholder consultation
- Stage 10: Finalise and publish White Paper
- Stage 11: Policy adoption by parliament
- Stage 12: Policy implementation - develop frameworks
- Stage 13: Develop an implementation plan, that includes Monitoring and Evaluation

This framework speaks to multiple stakeholders involved in African science, including researchers, librarians, academic institutions, funding bodies, publishers, National Research Education Networks (NRENs) and national governments. In addition to the many stakeholders, NGOs, country citizens and private industry should also be invited to engage in dialogue towards policy formulation. Various opportunities should be created by governments to create dialogue and inclusiveness.

Challenges Related to Data Sharing

There are many potential barriers to Open Data that have been identified by various observers over the years (Levin et al., 2016; Babini, 2014; Brant & Lohse, 2014), including the following:

- The current science culture can be described as traditional/conservative/competitive. The competitiveness of a given field influences the researcher's ability to collaborate and share research data with peers. There are very few efforts towards developing integrative platforms for accessible research output.
- Funder policies are in conflict with institutional or national policies (including IPRs and licensing).
- Policies lack regulatory convergence and are not aligned with one another.
- Currently metrics to measure research impact - based on publishing in high impact factor (IF) journals, published by the "Big Five" - are very much paper-focused. Existing metrics largely overlook data sharing and undermine Open Science.
- Researchers are not incentivised for making their research Open Access, or for sharing research data.
- Collaboration with industrial partners and the commercialisation of new innovations place constraints on the sharing and dissemination of resources and research findings.
- Patents often delay publications by several months, while granting of patents can lead to restrictions on information to be made available in follow-up publications.
- Copyright law and other relevant intellectual property rights can have a negative effect on the dissemination of scientific results, and open licensing approaches are not yet in common use in many areas, especially in Africa.

Opportunities When Sharing Data

Although there are significant barriers to Open Data, there are many benefits or opportunities that have been identified over the years (Masuzzo & Martens, 2017; Jong & Slavova, 2014; Uhlir & Schröder, 2007), such as:

- Accelerate science and discoveries to benefit society in general
- Increased access to scientific research publications and data can cause spill-over to scientific systems, and boost innovations since unrestricted access to use and re-use of the outputs can produce new products and services
- Reduce duplication of efforts, and costs in collecting, creating, transferring and reusing data and scientific material
- Allow for the verification of findings through shared interpretations and views
- Improve effectiveness and productivity of science systems by increasing return on investment

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

- Research is a public good that anyone can utilise for higher social returns at no additional costs (maximising the economic and social value of data)
- Access to research and data can:
  - boost innovation, resulting in new products and services;
  - align science activities on the continent to significantly contribute to the global scientific knowledgebase;
  - decrease reliance on externally resourced data and information resources; and
  - promote cross boarder knowledge sharing for increased collaboration among countries.

Policy Statements and Enabling Processes

According to the LEARN Toolkit[19], research data is only one part of the knowledge capital (an "asset") of research intensive institutions. "In data-driven science, good data management promotes discovery, efficiency and increases reliability by ensuring consistent quality with a high level of comparability." The following broad Open Data policy statements and enabling processes are proposed, towards the successful implementation of an Open Data policy.

Policy Statement	Enabling Processes
Adopt the FAIR (**Fi**ndable, **A**ccessible, **I**nteroperable, **R**e-usable) data principles[20]	1. Include FAIR data as a requirement, part of institutional policy. Support the use/downloading of FAIR data, and sharing/uploading of FAIR data. 2. Endorse the Panton Principles for Open Data in Science[21]. 3. Endorse the Science International Accord on Open Data in a Big Data World[22]. 4. Participate in the global dialogue on Open Science, and specifically Open Data.
Apply instrumental, procedural and distributive data justice	1. *Instrumental data justice* means FAIR **use** of data; hence focuses on the outcome of use of data. 2. *Procedural data justice* means FAIR **handling** of data. 3. *Distributive data justice* means FAIR **distribution** of data understood from a rights-based perspective (Heeks & Renken, 2018).
Create an enabling environment in support of Open Access to all research output	1. Publish (or make the transition to) high-quality scholarly Open Access journals, and apply for inclusion in the Directory of Open Access Journals (DOAJ)[23]. 2. Publish (or make the transition to) high-quality scholarly Open Access monographs, and apply for inclusion in the Directory of Open Access Books (DOAB)[24]. 3. Publish (or make the transition to) high-quality Open Access conference proceedings. 4. Publish research data in an institutional or discipline data repository, making it openly accessible. Do so prior to submitting the final manuscript. Creating links between the data and

[19] http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf
[20] https://www.go-fair.org/
[21] https://www.pantonprinciples.org/
[22] http://www.science-international.org/
[23] https://doaj.org/
[24] https://www.doabooks.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

	manuscript will increase transparency, reproducibility and validation of the results, also aiding peer-review. If the institution does not offer an institutional repository service, consider uploading the data in a trusted and sustainable generic or discipline-specific data repository. Re3data.org[25] serves as a global directory of available data repositories. 5. Publish copies (pre- or post-prints) of papers appearing in subscription-based journals, in the institutional or discipline repository. 6. Make theses/dissertations as well as the underlying data, available as Open Access, through the institutional repository.
Support the use (application and implementation) of open licenses for both research data and literature (Carroll, 2015)	Licenses limiting commercial re-use or the production of derivative works are discouraged since they hinder the effective re-use of data, and also prevent commercial activities that could lead to innovations. The CC0[26] and CC BY[27] licenses generally are recommended for use with Open Data, and CC BY is already broadly in use for promoting Open Access to the research literature.
Provide a shared and interoperable research infrastructure	Shared and interoperable infrastructure facilitates the dissemination of research results and promotes scientific collaboration so as to avoid duplication of effort (Lasthiotakis et al., 2015). This applies to networking services (bandwidth, connectivity), computational capacity (server space, high performance computing facilities), digitisation, data stewardship and more. National Research and Education Networks (NRENs) are key to providing connectivity and bandwidth, and governments are encouraged to invest in those. Refer to the AOSP e-Infrastructure[28] for more information.
Incentivise data sharing and Open Access publishing	Funders should make the sharing of data mandatory as part of grant agreements. Researchers should be acknowledged for sharing data. The incentivisation can take the form of an enabling and supportive research environment, implementing policy, providing clear guidelines, providing the required infrastructure, and supporting participation in training and capacity building. Refer to the *AOSP Incentives Framework*[29]. Online badges can be used to acknowledge and signal recognition of open research practices. Awards can further be implemented in recognition of open research data sharing.
Encourage the use of Open Source software	Researchers, who write code as a means to obtain results from data, should release the code as well. The

[25] Registry of Research Data Repositories https://www.re3data.org/
[26] https://creativecommons.org/share-your-work/public-domain/cc0/
[27] https://creativecommons.org/licenses/by/2.0/za/
[28] http://dx.doi.org/10.17159/assaf.2019/0048
[29] http://dx.doi.org/10.17159/assaf.2019/0051

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052
This work is licensed under a Creative Commons Attribution 4.0 International License.

	use of a Version Control System (VCS) e.g. Git or Apache Subversion (SVN) is recommended. Also use a GNU General Public License (GNU GPL)[30]. Clear arrangements for storage and preservation of the code should be made while compiling and operating instructions will allow the code to run without issues. The code should be accompanied by a description of its core functionalities alongside the hard- and software requirements for its use (Masuzzo & Martens, 2017).
Consider Open Innovation in limited circumstances	Often research leads to new and innovative products and services. Partnerships between research institutions and private industry remain important, because not all good ideas are developed within a private company, and not all ideas should necessary be developed within the company. There are huge opportunities for research institutions and private industry to collaborate in terms of new innovations, industry attracting new talent, and universities getting the opportunity to expose students to industry, preparing them for the world of work.
Integrate data science and data management into institutional curricula	Multiple free online courses are available to benefit from. Research intensive institutions are encouraged to integrate data management skills across all disciplines – both natural sciences and the humanities. The ICT sector is encouraged to invest in developing more skills related to ICT infrastructure and programming, incl. high performance computing. See the *AOSP Capacity Building Framework*[31] and *AOSP Research Data Management Framework*[32] for more information.
Include citizens as stakeholders through community engagement, science communication efforts	Create opportunities for citizens to engage with research data and academics, e.g. during national science forums, institutional science forums. Increase trust in science. Refer to the *AOSP Incentives Framework* for more information.

Elements of an Open Data Policy

The following – with some adaptions - is based on the LEARN Toolkit[33] containing the *Model Policy for Research Management (RDM) at Research Institutions/Institutes*. When developing a policy, a “soft” approach can be followed (using ‘recommended’, ‘should’), or a “hard” approach (using ‘required’, ‘will’, ‘must’). The example text below follows a “hard” approach to Open Data policy.

Although the content refers to “institutions” and “organisations”, it can also be adapted for national policies on government level.

Policy Element	Description	Example Text
Header info	This includes the: • Document title • Institutional logo	

[30] https://www.gnu.org/licenses/gpl-3.0.en.html
[31] http://dx.doi.org/10.17159/assaf.2019/0049
[32] http://dx.doi.org/10.17159/assaf.2019/0050
[33] http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

Title of policy	Description of the pursued issue	
Subtitle	If necessary, add an extension of the title	
Authorship	It should be clear who defines the policy ("the speaking entity"/sponsoring organisation/body), and why this entity defines the policy. Also – what is the role of this authorship/speaking entity?	
Aim of policy	Indicate what the goal/objective of the policy is, and what does the institution want to achieve through the policy? Does it include recommendations only, or is it compulsory?	
Subject of policy	According to the statutes of the institution and its published guidelines, what is the subject of the policy?	
Definitions	Define – among others - what is meant by the following: • Research • Research data • Researcher • Open Data incl. the scope/relationship with Open Access and Open Science • Research data management (RDM) (for the purpose of this document, it includes research records, methods, software, code, instruments, etc.) • Data steward • Data scientist	See section on "Annexes".
Preamble	The preamble describes the context: • It is an introductory statement or a description of an initial situation. • It defines why there should be a policy and how to contextualise it within the institution. Align it with the prevailing philosophy and mission of the organisation. • The preamble ensures consistency on institutional level, regardless of the huge variety of approaches in various scientific disciplines. • Describe the propositions that serve as foundation for the reasoning within the policy.	The [name of research institution] recognises the fundamental importance of research data [use a footnote to refer to definitions in the Annex] and the management of related administrative records in maintaining quality research and scientific integrity, and is committed to pursuing the highest standards. The [name of research institution] acknowledges that correct and easily retrievable data (FAIR data) [use a footnote to refer to definitions in the Annex] are the foundation of and integral to every research project. They are necessary for the verification and defence of research processes and results. RDM policies are highly valuable to current and future researchers. Research data have a long-term value for research and academia, with the potential for widespread use and re-use in society.
Scope & Coverage	• The scope of the policy must be defined according to space and time. • Is this a national/ institutional/faculty-wide (or other organisational units)/ discipline-wide/project/group-wide (e.g. research staff,	This policy for the management of research data applies to all research data generated through conducting research, by all staff members at the [name of research institution], while

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

	research support staff, IT services, students) policy? • Does the policy apply to all research data? • Does the policy include/exclude a selection of the non-digital results of research processes? • The relationship between the policy and non-research and research institution guidelines and statutes must be clarified in the policy. • Compliance with legal and contractual provisions must be maintained. • New policies should align with related policies, for example IPR, ICT, HE, STI, Research Output, and Research Ethics policies. • The policy should contain a statement showing which policy takes precedence when research is funded by external funders, showing the expectations placed by the institution on external research partners.	being employed at the said institution. The policy was approved by the [dean/commission/authority] on [date]. In cases when research is funded by a third party, any agreements made with that party concerning intellectual property rights, access rights and the storage of research data, take precedence over this policy.
Intellectual Property Rights (IPR)	The policy must take into account all contracts entered into with funders and publishers, as well as contracts between researchers and their institutions, which have precedence. • Who owns the research data? Who holds rights in such data? The copyright owner to the data will determine who has the right to apply a license, such as a CC0 or CCBY license. • What are the terms of use applied to the data? • Which license/s apply to the use of the data? • How will data be protected? • Address: privacy rights, usage rights, exploitation rights, copyrights	Intellectual property rights (IPR) apply as defined in the work contract between the researcher and the [name of research institution]. Further agreements (e.g. grant or consortia agreements) will be taken into account. Where the IPR belongs to the [name of research institution] as defined in the work agreement, the [name of research institution] has the right to choose how to publish and share the data. This will be done in consultation with the researcher, serving the best interest of the [name of research institution]. The [name of research institution] will by default make research data available under an open license, unless legal obligations, third party rights, intellectual property rights and privacy rights prelude this. The license will be selected according to the type of data and in order to label the data and facilitate its utilisation. For source code, a General Public License (GPL)[34] will be considered. For all other data, CC0[35] or CCBY[36] licenses will be considered. Data which are not restricted by copyright will be marked with the Creative Commons Public Domain Mark[37].
Research Data Stewardship	**Data should be FAIR** This section refers to all processes dealing with institutional research data, throughout the research data lifecycle. Institutions supporting data sharing	• Research data should be stored and made available for use in a suitable repository or archiving system, such as [name of

[34] https://www.gnu.org/licenses/gpl-3.0.en.html
[35] https://creativecommons.org/share-your-work/public-domain/cc0/
[36] https://creativecommons.org/licenses/by/2.0/za/
[37] https://creativecommons.org/share-your-work/public-domain/pdm/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

	are encouraged to adopt the **FAIR** data principles: research data should eventually be **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable. Wilkinson et al. (2016) explains it as follows: • **Findable:** Data should be easy to identify and find for both humans and computers, with metadata that facilitate searching for specific datasets. • **Accessible:** Data should be stored for the long term so that they can easily be accessed and/or downloaded with well-defined access conditions, whether at the level of metadata, or at the level of the actual data. • **Interoperable:** Data should be ready to be combined with other datasets by humans or computers, without ambiguities in the meanings of terms and values. • **Re-usable:** Data should be ready to be used for future research and to be further processed using computational methods. This requires adequate information about how the data were obtained and processed (provenance), within the conditions of an appropriate license assigned to the data. **Exceptions applicable to data sharing** Exceptions should be acknowledged, e.g. restricted access to data based on grounds of national and international legislation, intellectual property, personally identified data and sensitive information. This may also concern the "right to be forgotten" (deletion of data). The policy must be clear on which data can be deleted, when, and by whom this decision must be taken. Consider replacing the data entry with a record confirming it has been deleted ("tomb stone"). If needed or foreseen, specify regulations for *open access*, *restricted/mediated/controlled access* and/or *closed data*. **Data preservation** Data should be kept safe from medium failures, natural disasters, as well as software and hardware obsolescence. Keeping data in a dark archive will protect it from any disaster, with no online user access, unless a "trigger event" occurs. This in addition to an online open access data platform. Members of digital preservation infrastructures such as CLOCKSS[38] serve as mirror sites of one another to back-up and repair the disrupted location's archive, should it occur. **Data storage and access** • The policy should address where data will be stored, and how it will be accessed. If possible,	institutional repository/archiving system, if available]. • Data should be provided with persistent identifiers, for citation purposes. • Adherence to citation norms and requirements regarding publication and future research should be assured, sources of subsequently-used data explicitly traceable, and original sources be acknowledged. • The integrity of the research data should be preserved at all times. Research data must be stored in a correct, complete, unadulterated and reliable manner. • Research data must be FAIR (findable, accessible, interoperable, re-usable), and be readily available for subsequent use. • In compliance with intellectual property rights, and if no third-party rights, legal requirements or property laws prohibit it, research data should be assigned a licence for open use. • Research data and records are to be stored and made available according to intellectual property laws or the requirements of third-party funders, within the parameters of applicable legal or contractual requirements, e.g. AU restrictions on where identifiable personal data may be stored. • Research data of future historical interest and the administrative records accompanying research projects should also be archived. The minimum archive duration for research data and records is 10 years after either the assignment of a persistent identifier or publication of a related work following project completion, whichever is later. • In the event that research data and records are to be deleted or destroyed, either after expiration of the required archive duration or for legal or ethical reasons, such action will be carried out only after considering all legal and ethical perspectives. The interests and

[38] https://clockss.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

	there should be a recommendation for the use of institutional research infrastructures. - The minimum recommended retention period of research data is 10 years. - Keep data permanently when it has community/heritage/scientific value, and when gene therapy and seismological data. - Clinical trials might be kept for 15 years. - Data from short-term projects completed by undergraduate students might be kept for 12 months.	contractual stipulations of third-party funders and other stakeholders, employees and partner participants in particular, as well as the aspects of confidentiality and security, must be taken into consideration when decisions about retention and destruction are made. Any action taken must be documented and be accessible for possible future audit.
Responsibilities, Rights, Duties	Regulations concerning the **responsibilities, rights and duties of the following persons and institutions** should be formulated with regard to research data: - Researchers and research data producers (e.g. PhD students) - Funders and funders' regulations (the policy should acknowledge that funders have rights and regulations, and demonstrate that these will be given precedence where appropriate) - Institutions - Research supporting agencies (e.g. libraries, IT services, research support centres, etc.) Define **roles, responsibilities and competencies** in order to assign objectives and define timeframes. Relevant questions to ask include: - Who is in charge of ensuring legal compliance? - Who will provide legal advice? - Who is in charge of the quality of the content? - Who is in charge of defining acceptable formats? - Who is in charge of maintaining the currency of formats over time? - Who will provide technical support? - Who will promote services? - Who will provide training? - What is the role of the data steward? - What is the role of the data scientist?	The responsibility for research data management during and after a research project lies with the [name of research institution] and its researchers, and should be compliant with codes for the responsible conduct of research. **Researchers are responsible for:** - Management of research data and data sets in adherence with principles and requirements expressed in this policy; - Collection, documentation, archiving, access to and storage or proper destruction of research data and research-related records. This also includes the definition of protocols and responsibilities within a joint research project. Such information should be included in a Data Management Plan (DMP)[39], or in protocols that explicitly define the collection, administration, integrity, confidentiality, storage, use and publication of data that will be employed. Researchers will produce a DMP for every research project. - Compliance with the general requirements of the funders and the research institution; special requirements in specific projects should be described in the DMP; - Planning to enable, wherever possible, the continued use of data even after project completion. This includes defining post-project usage rights, with the assignation of appropriate licences, as well as the clarification of data storage and archiving in the case of

[39] Refer to the *AOSP Research Data Management Plan Framework*

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

		discontinued involvement at the [name of research institution]; • Backup and compliance with all organisational, regulatory, institutional and other contractual and legal requirements, both with regard to research data, as well as the administration of research records (for example contextual or provenance information); • To ensure appropriate institutional support, it is required that new research projects are registered at the proposal stage at the [name of research institution]. **The [name of research institution] is responsible for:** • Empowerment of organisational units, providing appropriate means and resources for research support operations, the upkeep of services, organisational units, infrastructures, and employee education; • Support of established scientific practices from the beginning. This is possible through the drafting and provision of DMPs, monitoring, training, education and support, while in compliance with regulations, third-party contracts for research grants, university/ institutional statutes, codes of conduct, and other relevant guidelines; • Developing and providing mechanisms and services for the storage, safekeeping, registration and deposition of research data in support of current and future access to research data during and after the completion of research projects; • Providing access to services and infrastructures for the storage, safekeeping and archiving of research data and records, enabling researchers to exercise their responsibilities (as outlined earlier) and to comply with obligations to third-party funders or other legal entities.
Approval, review, validity and timeline	• This applies to the date of release of the policy and how long the current policy will be valid. This can be done on a regular basis, which may	This policy will be reviewed and updated as required by the head of/the director of the [name of

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

	be externally defined, or based upon needs. The key dates must be included. • The policy should be subjected to periodic review. The changes in each revision must be listed. • Relevant questions to ask include: o How long are the terms of the policy valid? o Who/which body is responsible for reviewing and updating the policy? o What should be done after the end of the defined timeline or period? • Include the following footer info: page numbers, version number, status, etc.	research institution], every [two years].
Annexes	Include Annexes covering the following: • Definition of key terms • Excerpts from/links to relevant funder policies or expectations • List of related institutional policies (with hyperlinks/URLs)	**Annex: Definitions** **Research** is any creative and systematically performed work with the goal of furthering knowledge, including discoveries regarding people, culture and society, in addition to the use of such knowledge for new applications. **Researchers** refers to all research-active members of an institution including employees and doctoral candidates. Persons not directly affiliated with an institution, but who, for purposes of research, make use of or are physically present at the institution, are also included in the term. Visiting researchers or collaborators may also be expected to comply with the policy. **Research data** refers to all information (independent of form or presentation) needed to support or validate the development, results, observations or findings of a research project, including contextual information. Research data include all materials which are created in the course of academic work, including digitisation, records, source research, experiments, measurements, surveys and interviews. This includes software and code. Research data can take on several forms: during the lifespan of a research project, data can exist as gradations of raw data, processed data (including negative and inconclusive results), shared data, published data and Open Access published data, and with varying levels of access, including open data, restricted data and closed data.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

Bibliography

AOSTI (African Observatory of Science, Technology and Innovation). (2014). *Assessment of scientific production in the African Union, 2005–2010 (online)*. Available at: http://aosti.org/images/assessment_of_scientific_production_in_the_african_union_2005-2010.pdf (Accessed on 23 April 2019)

Babini, D. (2014). Open access in Latin America and the Caribbean. In *Open Access in the Americas Research Without Borders: The Changing World of Scholarly Communications* (pp. 1–26) *(online)*. Columbia: Columbia University Scholarly Communication Program and Digital Humanities Center. Available at: https://doi.org/10.13140/2.1.2988.7364 (Accessed on 23 April 2019)

Brant, J., & Lohse, S. (2014). *The Open. Innovation and Intellectual Property (Vol. 3) (online)*. Available at: https://doi.org/10.3860/krit.v3i2.1532 (Accessed on 23 April 2019)

Carroll, M.W. (2015). Sharing Research Data and Intellectual Property Law: A Primer. *PLoS Biol 13(8): e1002235 (online)*. Available at: https://doi.org/10.1371/journal.pbio.1002235 (Accessed on 23 April 2019)

Communia-project.eu. (2019). *Panton Principles for Open Data in Science | COMMUNIA - The European Thematic Network on the Digital Public Domain (online).* Available at: https://communia-project.eu/content/panton-principles-open-data-science.html (Accessed on 23 April 2019)

Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean? *Information Development*, *34(1), 90–102 (online)*. Available at: https://doi.org/10.1177/0266666916678282 (Accessed on 23 April 2019)

Jong, S., & Slavova, K. (2014). When publications lead to products: The open science conundrum in new product development. *Research Policy*, *43(4), 645–654 (online)*. Available at: https://doi.org/10.1016/j.respol.2013.12.009 (Accessed on 23 April 2019)

Lasthiotakis, H., Kretz, A. and Sá, C. (2015). Open science strategies in research policies: A comparative exploration of Canada, the US and the UK. *Policy Futures in Education*, *13(8), pp.968-989 (online)*. Available at: https://doi.org/10.1177/1478210315579983 (Accessed on 23 April 2019)

LEARN. (2017). LEARN Toolkit of Best Practice for Research Data Management. *Leaders Activating Research Networks (LEARN) (online))*. Available at: http://dx.doi.org/10.14324/000.learn.00 (Accessed on 23 April 2019)

Levin, N., Leonelli, S., Weckowska, D., Castle, D. and Dupré, J. (2016). How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice. *Bulletin of Science, Technology & Society*, *36(2), pp.128-141 (online)*. Available at: https://doi.org/10.1177/0270467616668760 (Accessed on 23 April 2019)

Masuzzo P., & Martens L. (2017). Do you speak open science? Resources and tips to Learn the language. *PeerJ Preprints 5:e2689v1 (online)*. Available at: https://doi.org/10.7287/peerj.preprints.2689v1 (Accessed on 23 April 2019)

Picarra, M. (2016). Discussion Paper: Researchers and Open Science. *Pasteur4OA (online). Available at:* http://pasteur4oa.eu/sites/pasteur4oa/files/resource/Discussion%20Paper_Researchers%20and%20Open%20Science.pdf (Accessed on 23 April 2019)

Schiltz, M. (2018). Science Without Publication Paywalls: cOAlition S for the Realisation of Full and Immediate Open Access. *Frontiers in Neuroscience*, *12 (online)*. Available at: https://doi.org/10.3389/fnins.2018.00656 (Accessed on 23 April 2019)

Uhlir, P. F., & Schröder, P. (2007). Open Data for Global Science. *Data Science Journal*, 6(June), *OD36-OD53 (online)*. Available at: https://doi.org/10.2481/dsj.6.OD36 (Accessed on 23 April 2019)

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18 *(online)*. Available at: https://www.nature.com/articles/sdata201618 (Accessed on 23 April 2019)

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Open Data Policy Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0052

This work is licensed under a Creative Commons Attribution 4.0 International License.

African Open Science Platform (AOSP)
Research Data Management Framework

Fostering a culture of Open Data within African National Systems of Innovation

Developed by Thembelihle Hwalima[1], Nancy Kwangwa[2], and Ina Smith[3]

[1]Lupane State University, Zimbabwe
[2]University of Zimbabwe
[3]Academy of Science of South Africa

Rationale for a Framework on Research Data Management

The increasingly data-driven research landscape calls for renewed and continued efforts in the management of research data, referred to as Research Data Management (RDM). According to Whyte & Tedds (2011), RDM concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information.

RDM is seen as a complex issue involving multiple activities carried out by various actors addressing a range of drivers and influenced by a large set of factors (Pinfield et al., 2014). The complexity in the delivery of RDM services calls for a framework that provides guidance in the implementation of RDM services in research and academic institutions. Implementing a RDM service further aligns with the *AOSP Open Data Policy Framework*[1], which calls for the mandatory submission of a data management plan by researchers when applying for funding – on institutional or funder level. More and more funders are requiring researchers to submit data management plans, in order to qualify for funding. A data management plan can furthermore support a research proposal as well as application for ethical clearance. Data management plans are highly valuable to current and future researchers, especially where data is re-used. Such a plan can inform potential re-users of the research data on how the data were collected, which data collection instruments were used, and so forth.

[1] http://dx.doi.org/10.17159/assaf.2019/0052

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

Focus and Scope

This framework speaks to multiple stakeholders involved in African science, including researchers, librarians, academic institutions, funding bodies, publishers, National Research Education Networks (NRENs) and national governments. It is an outcome of the *AOSP Open Data Policy Framework*, providing guidance towards implementing the Open Data policy.

RDM partners at institutional level commonly include the library, information technology services, the office of research, records and archives services, the institutional quality management unit, the institutional ethics committee chair, research chairs, heads of research units and centres, and the centre for postgraduate studies (Chiware & Mathe, 2015). Institutional collaboration and inter-institutional partnerships with different stakeholders are key towards the successful implementation of a successful RDM service.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

Approach towards Implementing a RDM Service

1. Planning for RDM - Strategy
Developing a clear strategy is crucial in the development of sustainable RDM services in African higher education institutions. An effective RDM strategy should highlight major goals, objectives as well as activities for the attainment of set objectives. Developing a clear RDM strategy serves the following:

- It provides direction and informs action plans.
- It prioritises and aligns activities.
- It provides a framework for ongoing decision making.
- It enhances communication and commitment.

2. Feasibility Study and Business Case
RDM spans across the individual stages of the research data lifecycle. Nhendodzashe & Pasipamire (2017) recommend that organisations that seek to introduce RDM services should conduct a feasibility study around the research data lifecycle model. The introduction of RDM services should further be treated as a business case. The business case can be used to define how RDM infrastructure and support services will be resourced, what are the anticipated benefits, and make the case for investment.

3. Timeline
An incremental approach towards rolling out RDM into functional services is recommended. In some instances, it might take time before the new RDM service gain momentum, depending on the culture of the organisation and its response to change. Piloting the potential services is recommended during the early stages. To ensure that the strategy is accepted across the institution, it is important to engage key stakeholders from the conception of the RDM service. The RDM service can also be embedded in existing research strategies such as the library information literacy strategy, the institutional Intellectual Property Rights strategy, and/or the broad institutional research strategy.

4. Policy Development
Refer to the *AOSP Open Data Policy Framework*.

5. RDM Training
The *AOSP Capacity Building Framework*[2] provides a list of generic competencies required during the different stages of the research data lifecycle, as well as a list of online courses on RDM. The following core skills for different categories of data workers are proposed by the Research Data Management Forum (2008):

[2] http://dx.doi.org/10.17159/assaf.2019/0049

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

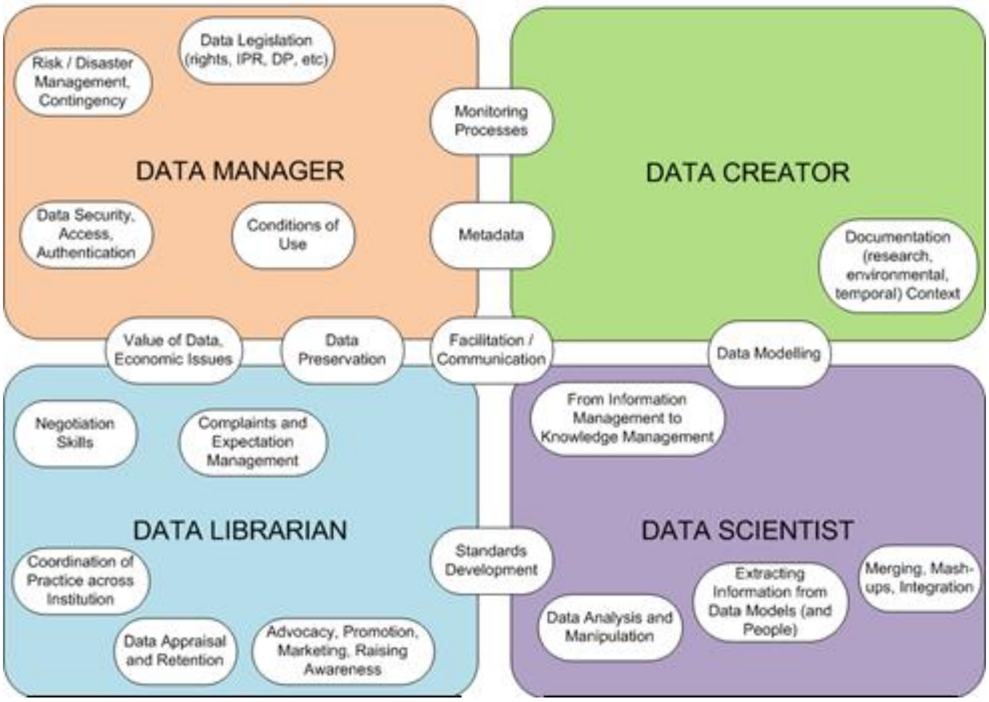

DATA MANAGER
Risk / Disaster Management, Contingency
Data Legislation (rights, IPR, DP, etc)
Monitoring Processes
Data Security, Access, Authentication
Conditions of Use
Metadata
DATA CREATOR
Documentation (research, environmental, temporal) Context
Value of Data, Economic Issues
Data Preservation
Facilitation / Communication
Data Modelling
Negotiation Skills
Complaints and Expectation Management
From Information Management to Knowledge Management
DATA LIBRARIAN
DATA SCIENTIST
Coordination of Practice across Institution
Standards Development
Merging, Mash-ups, Integration
Extracting Information from Data Models (and People)
Data Analysis and Manipulation
Data Appraisal and Retention
Advocacy, Promotion, Marketing, Raising Awareness

Figure 1. Core Skills for Data Management (Research Data Management Forum, 2008)

Training on institutional level will depend on who takes ownership for rolling out the RDM service. Librarians have highly relevant information standard and organisational skills to provide quality support services and training in terms of RDM (Burnett, 2013). To train others, librarians will have to be trained themselves. Training can be by means of short courses, university modules, or online courses, such as:

- MANTRA for Librarians available at http://mantra.edina.ac.uk/libtraining.html
- Coursera Research Data Management and Sharing – https://www.coursera.org/learn/data-management
- FOSTER Integrating Open Science in Information Literacy Education - https://www.fosteropenscience.eu/node/2016
- FOSTER Managing and Sharing Research Data - https://www.fosteropenscience.eu/node/2328

6. Develop RDM Services

When trained or once upskilled, data librarians can be embedded in research projects. RDM services can be broadly categorised as informational/consultative and technical services.

Consultative Services

- During initial stages of RDM service deployment.
- Includes needs assessments, training, crafting policies and guidelines, providing assistance on how to complete a data management plan, assigning metadata, data curation services and Open Science (incl. Open Access) advocacy.
- Advocate for transparency, openness in research (Open Science), and access to research data (Open Data) that is FAIR (**f**indable, **a**ccessible, **i**nteroperable, **r**e-usable).
- RDM services are linked to other services provided by librarians, such as information literacy training (incl. literature searches, referencing, citation, etc.), and can be included as part of those.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

- Refer researchers to trusted data repositories and/or data sets – both for downloading re-usable data sets, and uploading where there is no other local data repository available.
- Create awareness on available data services, applicable to all stages of the research data lifecycle.
- Advise on Intellectual Property Rights (incl. copyright), licensing, citations, persistent identifiers, version control, and more.
- Understand the research lifecycle and research data lifecycles, and which tools/applications/middleware are required during the various stages of the data research lifecycle.

Technical RDM Services

- Involves setting up trusted data repositories, managing the data for the unforeseeable future once the data scientists has exploited its value for a specific research project. In some cases, existing institutional literature repositories are expanded to also include data.
- Manage data within trusted data repositories, according to the CoreTrustSeal[3], and register the data repository with trusted directories, such as the Registry of Research Data Repositories (re3data.org).
- Collaborate and liaise with researchers and ICT, addressing data needs.
- Monitor the impact of research data sets once shared.
- Have knowledge of discipline-specific taxonomies, ontologies, vocabularies, standards and more.
- In some cases, librarians perform quality checks and verify metadata, file formats and documentation (Matusiak & Sposito, 2017).

7. Develop an Institutional Data Repository

Setting up an institutional data repository requires institutional commitment and funding – from the library, information technology services, the office of research, records and archives services, the institutional quality management unit, the institutional ethics committee chair, research chairs, heads of research units and centres, and the centre for postgraduate studies. The roles and responsibilities for each should be further clarified, and Monitoring and Evaluation practises put in place.

Interoperable Open Source (OAI-PMH compliant) data repository software is recommended – the applications can be downloaded from the WWW, and installed on a server/in the cloud.

Application Software in Support of RDM

Example **data repository software** to manage the data during the sharing phase of the research data lifecycle, include:

- Invenio - https://invenio-software.org/
- Dataverse - https://dataverse.org/
- DSpace - https://duraspace.org/dspace/ and DSpaceCRIS https://dspace-cris.4science.it/handle/123456789/15
- World Bank Technology Options - http://opendatatoolkit.worldbank.org/en/technology.html

For preservation purposes, in a **dark archive**, consider:

- LOCKSS - https://www.lockss.org/
- CLOCKSS - https://clockss.org/

Data Management Plans and templates to consider:

- See *Annexure A* for a data management plan in Word format.
- DMPonline - https://dmponline.dcc.ac.uk/
- DMPTool - https://dmptool.org/

[3] https://www.coretrustseal.org/

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

- DMPRoadmap - https://github.com/DMPRoadmap
- NECDMC - https://library.umassmed.edu/resources/necdmc/dmp
- Also see http://www.dcc.ac.uk/resources/data-management-plans

Following a visualisation of possible RDM services during various stages of the research data lifecycle.

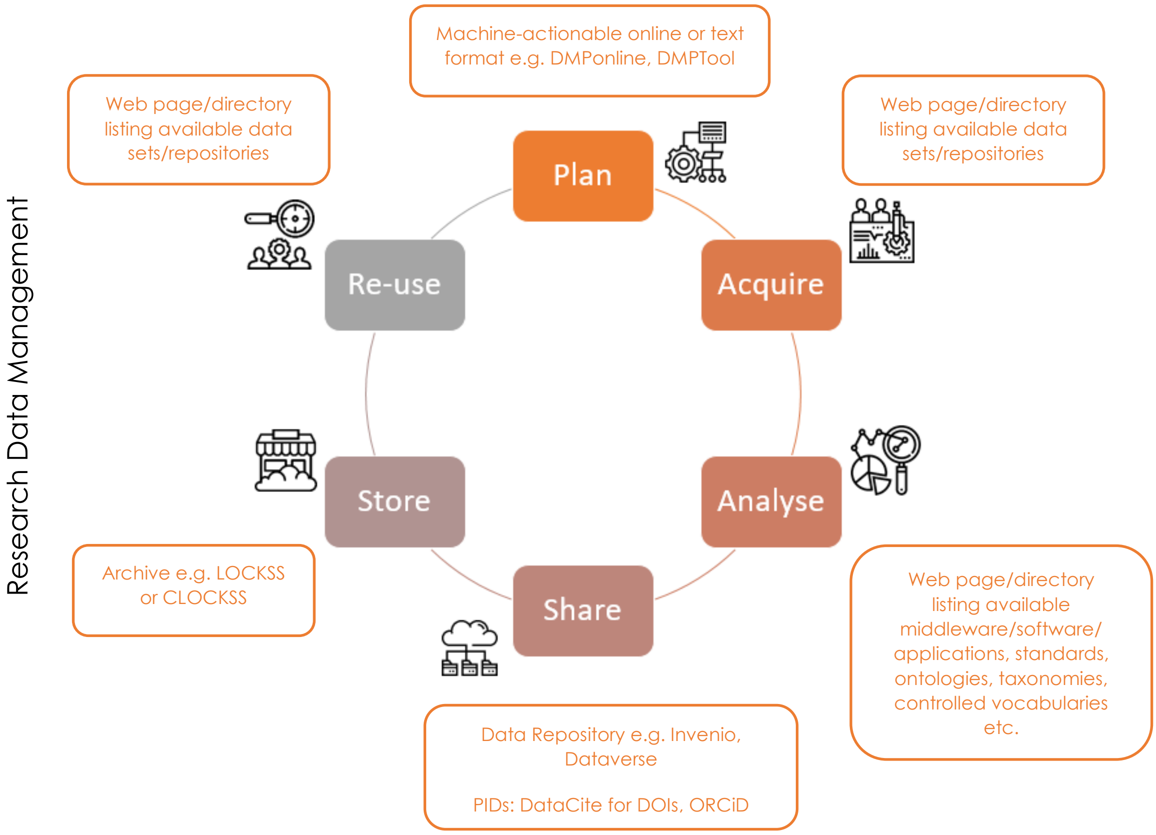

Research Data Management
Machine-actionable online or text format e.g. DMPonline, DMPTool
Web page/directory listing available data sets/repositories
Web page/directory listing available data sets/repositories
Plan
Re-use
Acquire
Store
Analyse
Share
Archive e.g. LOCKSS or CLOCKSS
Web page/directory listing available middleware/software/ applications, standards, ontologies, taxonomies, controlled vocabularies etc.
Data Repository e.g. Invenio, Dataverse
PIDs: DataCite for DOIs, ORCiD

Figure 2. RDM services during various stages of the research data lifecycle

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

Annexure A: Checklist for a Data Management Plan

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant.
Grant Reference Number	Enter grant reference number if applicable.
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	**Questions to consider**: - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI/Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI/Researcher ID	E.g. ORCiD http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details.
Date of First Version	Date the first version of the DMP was completed.
Date of Last Update	Date the DMP was last changed.
Related Policies	**Questions to consider:** - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) or Data policy? - Does your funder have a Research Data Management or Data policy? - Are there any formal standards that you will adopt? **Guidance:** List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.
Data Collection	
Type, format, volume of data	What data will you collect or create? **Questions to consider:** - What type, format and volume of data? - Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse? **Guidance:** Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.
Data Collection Method	How will the data be collected and created? **Questions to Consider:** - What standards or methodologies will you use? - How will you structure and name your folders and files? - How will you handle versioning? - What quality assurance processes will you adopt? **Guidance:** Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

Documentation and Metadata	
Accompanying Material	What documentation and metadata will accompany the data? **Questions to consider:** - What information is needed for the data to be read and interpreted in the future? - How will you capture / create this documentation and metadata? - What metadata standards will you use and why? **Guidance:** Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.
Ethics and Legal Compliance	
Ethical Issues	How will you manage any ethical issues? **Questions to consider:** - Have you gained consent for data preservation and sharing? - How will you protect the identity of participants if required? - How will sensitive data be handled to ensure it is stored and transferred securely? **Guidance:** Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.
Intellectual Property Rights	How will you manage Intellectual Property Rights (IPR) issues, e.g. copyright? **Questions to consider:** - Who owns the data? - How will the data be licensed for reuse? - Are there any restrictions on the reuse of third-party data? - Will data sharing be postponed / restricted e.g. to publish or seek patents? **Guidance:** State who will own the IPR (incl. copyright) of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on IPR or copyright. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.
Storage and Backup	
Storage and Backup Procedures	How will the data be stored and backed up during the research? **Questions to consider:** - Do you have sufficient storage or will you need to include charges for additional services? - How will the data be backed up? - Who will be responsible for backup and recovery? - How will the data be recovered in the event of an incident? **Guidance:** State how often the data will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes. If you choose to use a third-party service, you should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.
Access and Security	How will you manage access and security? **Questions to consider:** - What are the risks to data security and how will these be managed?

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.

Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

	- How will you control access to keep the data secure? - How will you ensure that collaborators can access your data securely? - If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems? **Guidance:** If your data is confidential (e.g. personal data not already in the public domain, confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.
Selection and Preservation	
Selection Criteria	Which data should be retained, shared, and/or preserved? **Questions to consider:** - What data must be retained/destroyed for contractual, legal, or regulatory purposes? - How will you decide what other data to keep? - What are the foreseeable research uses for the data? - How long will the data be retained and preserved? **Guidance:** Consider how the data may be reused e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.
Long term Preservation	What is the long-term preservation plan for the dataset? **Questions to consider:** - Where e.g. in which repository or archive will the data be held? - What costs if any will your selected data repository or archive charge? - Have you costed in time and effort to prepare the data for sharing / preservation? **Guidance:** Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.
Data Sharing	
Sharing	How will you share the data? **Questions to consider:** - How will potential users find out about your data? - With whom will you share the data, and under what conditions? - Will you share data via a repository, handle requests directly or use another mechanism? - When will you make the data available? - Will you pursue getting a persistent identifier for your data? **Guidance:** Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the reuse of your data.
Restrictions	Are any restrictions on data sharing required? **Questions to consider:** - What action will you take to overcome or minimise restrictions? - For how long do you need exclusive use of the data and why? - Will a data sharing agreement (or equivalent) be required? **Guidance:** Outline any expected difficulties in sharing data with acknowledged long-term value, along with causes and possible measures to overcome these. Restrictions may be due to confidentiality, lack of consent agreements or IPR, for example. Consider whether a nondisclosure agreement would give sufficient protection for confidential data.

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

Responsibilities and Resources	
Data Manager	Who will be responsible for data management? **Questions to consider:** - Who is responsible for implementing the DMP, and ensuring it is reviewed and revised? - Who will be responsible for each data management activity? - How will responsibilities be split across partner sites in collaborative research projects? - Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners? **Guidance:** Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.
Resources	What resources will you require to deliver your plan? **Questions to consider:** - Is additional specialist expertise (or training for existing staff) required? - Do you require hardware or software which is additional or exceptional to existing institutional provision? - Will charges be applied by data repositories? **Guidance:** Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.

Table 1. Checklist for a data management plan (DCC, 2013)

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa.
Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.

Bibliography

Burnett, P. (2013). *What is the role of a librarian in Research Data Management?* INASP Blog. Available at: http://blog.inasp.info/research-data-management-role-librarians/ (Accessed on 12 May 2018).

Chiware, E., & Mathe, Z. (2015). Academic libraries' role in Research Data Management Services: a South African perspective. *South Afr. J. Libr. Inf. Sci. 81, 1.* Available at: http://sajlis.journals.ac.za/pub/article/view/1563 (Accessed on 12 May 2018).

DCC. (2013). *Checklist for a Data Management Plan.* V.4.0. Edinburgh: Digital Curation Centre. Available at: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf (Accessed on 12 May 2018).

Mutusiak, K.K. & Sposito, F.A. (2017). Types of research data management services: An international perspective. *Proceedings of the Association for Information Science and Technology Banner.* 54 (3), pp 754-756. Available at: https://doi.org/10.1002/pra2.2017.14505401144 (Accessed on 12 May 2018).

Nhendodzashe, N. & Pasipamire, N. (2017). *Research data management services: are academic libraries in Zimbabwe Ready? The case of the University of Zimbabwe Library*. Available at: http://library.ifla.org/1728/ (Accessed on 12 May 2018).

Pinfield, S., Cox, A.M., & Smith, J. (2014). Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLOS ONE* 9, e114734. Available: at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114734 (Accessed 2018 May 12)

Research Data Management Forum. (2008). *RDMF2: Core Skills Diagram*. Available at: http://data-forum.blogspot.com/2008/12/rdmf2-core-skills-diagram.html (Accessed on 12 May 2018).

Whyte, A. & Tedds, J. (2011*)*. Making the Case for Research Data Management. *DCC Briefing Papers*. Edinburgh: Digital Curation Centre. Available at: http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm (Accessed on 12 May 2018).

© 2019. SA Dept. of Science and Technology, National Research Foundation, Academy of Science of South Africa. Please cite as follows: Academy of Science of South Africa (2019), *African Open Science Platform (AOSP) Research Data Management Framework: fostering a culture of Open Data within African National Systems of Innovation.* doi: http://dx.doi.org/10.17159/assaf.2019/0050

This work is licensed under a Creative Commons Attribution 4.0 International License.