Perspectives from the African Open Science Platform (AOSP)

Presented by Susan Veldsman
Director: Scholarly Publishing
ASSAf
About AOSP

• Funded by the National Research Foundation (NRF) (SA Dept. of Science and Technology)
• 3 years (1 Nov. 2016 – 31 Oct. 2019)
• Open Data: Where are we? Where do we want to be? What do we need to get there?
• Managed by Academy of Science of South Africa (ASSAf)
  • Through ASSAf hosting ICSU Regional Office for Africa (ICSU ROA)
  • Direction from CODATA
About ASSAf

- **Recognise** scholarly achievement & excellence
- **Mobilise members** in the service of society
- Conduct **systematic & evidence-based studies** on issues of national importance (ASSAf OA Repository)
- Promote the development of an **indigenous system** of South African research
- Publish science-focused **journals** (SAJS, QUEST, SciELO SA)
  - Training in Open Journal Systems (OJS)
  - Criteria for high quality OA journals
  - Ambassador for Directory of Open Access Journals (DOAJ)
• Develop productive **partnerships** with national, regional and international organisations to build capacity within the National System of Innovation (NSI)

• Create diversified **sources of funding** for sustainable functioning and growth of a national academy

• **Communicate** with relevant stakeholders
  • Association of African Universities (AAU) DATAD-R harvester of OA repositories
  • Evaluation instrument – harvesting IRs adhering to criteria for best practice (ISO 16363, Data Seal of Approval, WDS etc.)
African Academies of Science
AOSP Governance

- Advisory Council (Chair: Prof Khotso Mokhele)
  - Terms of Reference
- Technical Advisory Board (Chair: Prof Joseph Wafula)
  - Terms of Reference
- Working Group
  - Project Team
  - Project Administrator
Key Stakeholders

- Global Network of Science Academies (IAP)
- International Council for Science (ICSU)
  - Regional Office for Africa (ROA)
  - Committee on Data for Science and Technology (CODATA)
  - World Data System (WDS)
- The World Academy of Sciences (TWAS)
- Research Data Alliance (RDA)
- Association of African Universities (AAU)
• Network of African Science Academies (NASAC)
• African Research Councils (incl. DIRISA, funders)
• African Universities
• African Governments
• NRENs (Internet Service Providers for Education)
• Other
Progress re Openness in Africa

• Open Institutional Research Repositories ([Webometrics](https://webometrics.info) - 74)
• Open Educational Resources (E.g. UCT OER)
• Massive Open Online Courses (MOOCs)
• Open Access Journals (DOAJ) (11 SA universities hosting their own)
Open Data Repositories (re3data - 16)
Data Repositories vs Social Media

• Social media sites:
  • Connect researchers sharing interests
  • Marketing data
  • Sites belong to third parties

• Repository:
  • Supports export/harvesting of metadata
  • Offers long-term preservation
  • Non-profit – no advertisements
  • Uses open standards and protocols
  • Copyright
Africa Supporting Open Science

- **OA2020** (2017)
- **Dakar declaration on Open Science in Africa** (2016)
- **Open Science for the 21st century A declaration of ALL European Academies** (2012)
- **Salvador Declaration on Open Access Cape Town Declaration** (2010)
- **Cape Town Open Education Declaration** (2008)
- **Kigali Declaration on the Development of an Equitable Information Society in Africa**
What is Open Science/ Open (Research) Data?
Open Science Defined

“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 and Collaborative Science in Development Network project, funded by IDRC
Research data

Information collected using specific methods for a specific purpose of studying or analysing.

"Data" implies accompanying metadata (e.g. precise definitions of quantities, equations of interrelationships, scientific units of measurement, error analysis, etc.)

In experimental sciences the data is all the information required to repeat the experiment and the resulting data reported from that experiment.

• https://blogs.ch.cam.ac.uk/pmr/2010/07/25/pp01-what-is-scientific-data/
• www.yourdictionary.com/scientific-data
Social Science?

- For **social science**, **data** is generally numeric files originating from **social** research methodologies or administrative records, from which statistics are produced. It also includes, however, more **data** formats such as audio, video, geospatial and other digital content that are germane to **social science** research.
From commitment to action

OGP was launched in 2011 to provide an international platform for domestic reformers committed to making their governments more open, accountable, and responsive to citizens.
Open Data, Open Science & Research Lifecycle (Foster)
Why African Open Science/ Open Data?
Data sharing: Make outbreak research open access

Nathan L. Yozwiak, Stephen F. Schaffner & Pardis C. Sabeti

25 February 2015

“We urge researchers working on outbreaks to embrace a culture of openness.”
"What 115 years of data tells us about Africa's battle with malaria past and present" —
Weekend reads: Fake scientists; fake research; major evils of modern research

with one comment

The week at Retraction Watch featured the story of a graduate student who fought back after being caught in the middle of a fraud case, and the retraction of a hotly debated paper from Nature Cell Biology. Here’s what was happening elsewhere:

- Borat, the prominent scientist: Making up names and CVs is one of the latest tricks to game scientific metrics, our co-founders write in Nautilus.
- “The new owner of two prominent chains of Canadian medical journals is publishing fake research for cash, and pretending it is genuine.” (National Post)
- “[S]elective reporting, selective citing, and flaws in quality assurance and mentoring are the major evils of modern research.” (Research Integrity and Peer Review)
- Grad students and postdocs are frequently the scapegoats in cases of fraud, and they have little in the way of recourse in the current system, note our co-founders in STAT.
- When it comes to failing to publish clinical trial data, Canadian institutions are among the worst. (Darryl Hol, CBC News)
In many African countries, intellectual property protection is undeveloped, ineffective, expensive and unenforced and in some African countries there exists uncertainty on protection of IP and the threat of innovation being stolen away from inventors.

Difficulties with sharing data

• One of the challenges of sharing data is to provide enough information about
  – Context – Measurement process (Metadata)
• Plus the data must be stored in a way that it is “discoverable”
• Must adhere to standards an protocols
• All of this costs time and effort
• Fear of getting scooped
• Fear of someone else finding a path-breaking application of the data that one hadn’t thought of
• Fear of problems/errors in the measurement process being exposed
• Confidentiality/privacy of respondents - Ethics clearance
• IPR
Accord on Open Data in a Big Data World

- Values of open data in emerging scientific culture of big data
- Need for an international framework
- Proposes comprehensive set of principles
- FAIR Principles
- Provides framework & plan for African data science capacity mobilization initiative
- Proposes African Platform

Call to Endorse
4 Focus Areas

- Policy
- Infrastructure
- AOSP
- Capacity Building
- Incentives
Policy Framework

- **JKUT (Kenya)** Institutional Open Data Policy
- **Uganda** Draft Open Data Policy
- **Madagascar** Lobbying for Open Data Policy
- Towards a White Paper on Open Research Data Strategy in **Botswana**
- **White Paper on Science, Technology and Innovation in South Africa**
- Funder Policy: National Research Foundation (NRF)(SA)
- OECD Principles & Guidelines for Access to Research Data from Public Funding
30 Nov – 1 Dec 2017

The Department of Science and Technology (DST), in partnership with the European Commission, cordially invite you to a SA-EU Open Science workshop in order to develop a South African Open Science Policy Framework.
Infrastructure Framework

• NRENs –
  • Richly connected at high speed to many other networks/resources
  • Value added services e.g. grid & cloud computing resources, user controlled light paths, videoconferencing, federated identity services, and more
  • Deep culture of collaboration
Infrastructure Framework

• Roadmaps needed

• Centres for High Performance Computing
African NRENs

- **Cluster 1**: Southern and Easter Africa - UbuntuNet Alliance
- **Cluster 2**: Western and Central Africa - WACREN
- **Cluster 3**: North Africa -- Asren

Source: https://www.geant.org/News_and_Events/PublishingImages/WorldBank_Map2.jpg
AfricaConnect2 supports the development of high-capacity internet networks for research and education across Africa. It builds on existing networks in Eastern & Southern Africa and North Africa and will extend connectivity into West and Central Africa.

By collaborating with AfricaConnect2, the pan-European GÉANT network strengthens Europe's links with the African continent and provides African research and education communities with a gateway for global collaborations.

Through capacity building and advanced network connectivity and services, AfricaConnect2's mission is to contribute to sustainable development and a more inclusive Information Society across the whole African continent. Find out more >>

https://www.africaconnect2.net/Pages/Home.aspx
Data Prices Are Too Expensive For Most. You Should Care.

#dataMUSTfall, please.

<table>
<thead>
<tr>
<th>Country</th>
<th>Cost per 1GB (USD) of mobile data in Q2 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>0.89</td>
</tr>
<tr>
<td>Egypt</td>
<td>2.82</td>
</tr>
<tr>
<td>Mozambique</td>
<td>2.87</td>
</tr>
<tr>
<td>Ghana</td>
<td>3.91</td>
</tr>
<tr>
<td>Nigeria</td>
<td>4.85</td>
</tr>
<tr>
<td>Rwanda</td>
<td>5.03</td>
</tr>
<tr>
<td>South Africa</td>
<td>5.27</td>
</tr>
</tbody>
</table>
“Construction of the SKA is due to begin in 2018 and finish sometime in the middle of the next decade. **Data acquisition will begin in 2020, requiring a level of processing power and data management know-how that outstretches current capabilities.** Astronomers estimate that the project will generate **35,000-DVDs-worth of data every second.** This is equivalent to “the whole world wide web every day,” said Fanaroff.

The project is investing in **machine learning** and **artificial intelligence software tools** to enable the **data analysis.** In advance of construction of the vast telescope - which will consist of some 250,000 radio antennas split between sites in Australia and South Africa - SKA already employs more than 400 engineers and technicians in **infrastructure, fibre optics** and **data collection.**”

Africa participation

- Botswana
- Ghana
- Kenya
- Madagascar
- Mauritius
- Mozambique
- Namibia
- Zambia
Capacity Building Framework

• Data collector vs data user vs data manager

Therefore the following are core aspects to capacity building:
• Research Data Management Planning
• Repositories
• Command Line Interpretation
• Software Development
• Data Organisation
• Data Cleaning
• Data Management & Databases
• Data Analysis & Visualisation (incl. programming)
Challenges re Capacity Building

Students Say They Are Not as Tech Savvy as Educators Assume

By Jenny Abamu  Jun 20, 2017
How to define the evolving role of data scientist

Data science is a hot new career, but companies still aren't sure how to best use these employees. Here's how to hire data scientists with a clear strategy in place.

Robert Half Top Ten Technology Jobs In 2017: Data Scientists, Big Data Expertise In High Demand

Systems Administrator: Linux systems
(Ref. BIB/185/0517)

The incumbent will be responsible for the technical administration of the Linux-based software applications in the Library and Information Service, while liaising with ICT and Library staff to research and implement new technologies for scholarly communication and research support services.

Duties:
- Providing first-level application support for niche Linux-based systems for scholarly communication and research support;
- Assisting Library staff in application configuration, data extraction, and reporting;
- Liaising with the institutional ICT Division on systems administration and related support incidents;
- Researching, installing, configuring, testing and maintaining Linux-based software applications;
- Overseeing the scripting of automation routines to streamline back office processes;
- Maintaining documentation of systems configuration and workflow processes.

Requirements:
- A three-year tertiary IT qualification, or a minimum of five years' relevant work experience in Linux systems administration;
- Demonstrated competencies in the management of the Linux operating system on command-line level (CLI);
- Experience in database management software such as MySQL or PostgreSQL;
- Experience in data manipulation and analysis;
- Good communication skills.

Recommendations:
- LPIC certification;
- Working level knowledge of R, Python, or related data science tools;
- Knowledge of descriptive metadata schemas and bibliographic standards.

Commencement of duties: 1 September 2017 or as soon as possible thereafter

Closing date: 25 June 2017

Enquiries regarding this position: Mr. Wouter Klapwijk on 021 808 4378

The University will consider all applications in terms of its Employment Equity Plan, which acknowledges the need to diversify the demographic composition of the staff corps, especially with regard to the appointment of suitable candidates from the designated groups.
Incentives Framework

- Funder requirements changing
- Mechanisms that acknowledge publication of datasets and to promote data sharing
- How do we deal with difficulties in sharing data—what are the solutions
- Why is sharing essential
- How do we make sharing successful
- How do we lay the fears down and ensure buy-in
AOSP Series of Webinars on Incentives
October 2017

Dr Louise Bezuidenhout
Institute for Science, Innovation and Society (InSIS)
University of Oxford
25 October 2017, 10:00-11:00 SAST

Prof Martin Wittenberg
Director DataFirst, Univ. of Cape Town, SA & Winner of the 2016/2017 NSTF-South32 Data for Research Award by advancing the availability, management and use of data for research by an individual or an organisation
26 October 2017, 10:00-11:00 SAST

Veerle Van den Eynden
Producer Relations & Research Data Manager
UK Data Archive
University of Essex
27 October 2017, 10:00-11:00 SAST
Actions & Deliverables Year 1

• AOSP Side Event to the SFSA 2016 - Launch
• AOSP Workshop AAU 2017 (Ghana) - Policy
• Madagascar Meeting & Workshop 2017 – Policy & Capacity Building
• Botswana National Open Data Open Science Forum 2017 - Policy
• UbuntunetConnect Ethiopian Workshops 2017 – Policy & Infrastructure
• Database & Networks – Surveys etc
Actions & Deliverables Year 2

- Capacity Building – 3x AOSP Workshops
  - Batch/Grid Computing: Condor, HTCondor, SGE, PBS/Torque, LSF, SLURM
  - Computing Workflow: DAGMan, Pegasus, Makeflow
  - Data Repositories: Figshare, Zenodo, re3data.org, Dataverse, DSpace etc.
- RDM: DMPRoadmap
- Command Line Interface Interpretation: UNIX Shell, bash (incl. editors e.g. Nano, Emacs, Vim, etc.)
• Software Development: **Git, GitHub**, Mercurial
• Data organization: Spreadsheets
• Data Cleaning: OpenRefine
• Data Management & Databases: **SQL, Hadoop** (for really large datasets), MySQL, PostgreSQL
• Data Analysis & Visualisation: **R, R Markdown, ggplot2**, Python, MATLAB, C, C++
• Machine Learning: **R, Python**
• Artificial Neural Networks: **R, Python, Tensorflow, deep convolutional networks**
Closing Remarks

• Collaborate & learn from one another – strength in diversity
• Take ownership & collect/curate data in ethical way
• Downloaders vs Uploaders
• Trusted & valid data managed in trusted way
• Exploit data for the benefit of society (Min Naledi Pandor)
• Tell the African story, in an African way
Thank you
Susan Veldsman
susan@assaf.org.za

Ina Smith
ina@assaf.org.za
http://africanopenscience.org.za
@AOSP_Official

Copyright 2017 SA Dept. of Science and Technology
License: CC-BY