



"Without data, science is merely an opinion"

African Open Science Platform

Presented by Ina Smith

Project Manager African Open Science Platform
Academy of Science of South Africa (ASSAf)
UbuntuNetConnect 2017 Conference 2 Nov.

Fake Data, Fake Research

'Fake research' comes under scrutiny

By Helen Briggs
BBC News

27 March 2017 | Science & Environment

f t w e Share



SPL

The long read

The hi-tech war on science fraud

The problem of fake data may go far deeper than scientists admit. Now a team of researchers has a controversial plan to root out the perpetrators

by Stephen Buranyi

One morning last summer, a German psychologist named Mathias Kauff woke up to find that he had been reprimanded by a robot. In an email, a computer program named Statcheck informed him that a 2013 paper he had published on multiculturalism and prejudice appeared to contain a number of incorrect calculations - which the program had catalogued and then posted on the internet for anyone to see. The problems turned out to be minor - just a few rounding errors - but the experience left Kauff feeling rattled. "At first I was a bit frightened," he said. "I felt a bit exposed."

Kauff wasn't alone. Statcheck had read some 50,000 published psychology papers and checked the maths behind every statistical result it encountered. In the space of 24 hours, virtually every academic active in the field in the past two decades

<http://www.bbc.com/news/science-environment-39357819>

<https://www.theguardian.com/science/2017/feb/01/high-tech-war-on-science>



Data sharing: Make outbreak research open access

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

25 February 2015

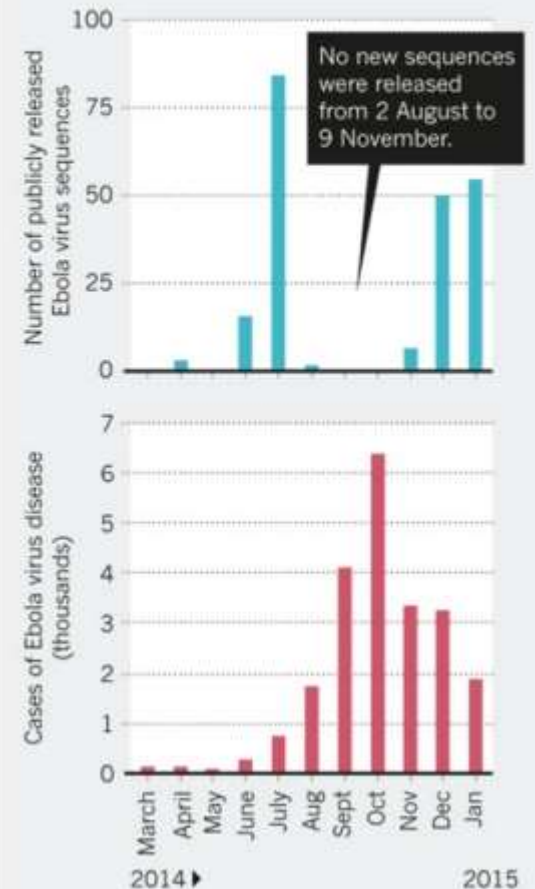


"We urge researchers working on outbreaks to embrace a culture of openness."

<http://www.nature.com/news/data-sharing-make-outbreak-research-open-access-1.16966>

GAPS IN THE DATA

Genome sequences from the West Africa outbreak of Ebola virus were first made publicly available in April 2014. Since 99 genomes were released in July, data sets have been shared sporadically, even though more are known to have been generated.



Sources: Sequences, NCBI/virological.org; Ebola cases, WHO

Trusted Research & Data

- Trust is at the **centre** of the process of science - Wim Hugo
- Trust researchers who have your **best interest** at heart
- Build **new research on existing research/data**
- Data serves as **evidence** for research
- **Accountability** towards funders
- Research & data - **asset**
- Research/data properly **managed & curated**, complete

Fears Researchers Experience

- Getting **scooped**
- **Time & effort, resources** by researcher
- Someone else finding a path-breaking application of the data that researcher hasn't considered – "**competition**"
- Fear of **problems/errors** in the measurement process being exposed
- **Confidentiality/privacy** of respondents - ethics clearance
- **Intellectual Property Rights** – signed away, little understanding, no IP in place

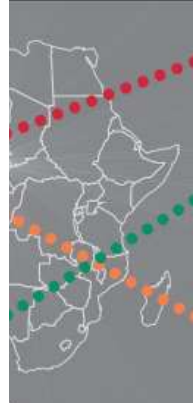
Benefits of Open Data

- Predict **trends** & informed **decisions**
- Drive **development** & improve **livelihoods**
- More **entrepreneurs** – using data in innovative ways, create jobs
- Improve **service delivery**
- Provide **evidence** for research conducted
- Have potentially **far more outcomes when open, higher impact**
- Democratising research & data towards achieving 2030 Sustainable Development Goals, Agenda 2063

What 115 years of data tells us about Africa's battle with malaria past and present

October 17, 2017 7:01am SAST

The prevalence of malaria infection in sub-Saharan Africa today is at the lowest point since 1900.



<http://theconversation.com/what-115-years-of-data-tells-us-about-africas-battle-with-malaria-past-and-present-85482>



The screenshot shows the Open Data Institute (ODI) website. The header includes the ODI logo and navigation links: 'News & blog', 'Get involved', 'Our global network', 'Events', and 'Our fo'. A blue banner with the word 'News' is prominent. The main article is titled 'How open data solves problems in the developing world: from mapping ebola to protecting banana crops', dated '2015-02-12 by Open Data Institute'. It features a list of six bullet points detailing various applications of open data. At the bottom, a short paragraph introduces a report from the ODI.

ODI open data institute

News & blog Get involved Our global network Events Our fo

News

How open data solves problems in the developing world: from mapping ebola to protecting banana crops

2015-02-12 by Open Data Institute

A new white paper launched by the ODI shows how open data can:

- protect banana farmers' livelihoods in Uganda
- help parents to assess school performance in Tanzania
- expose \$62m in potential health savings in Southern Africa
- map the Ebola outbreak to save lives in West Africa
- make aid more effective in Nepal
- engage citizens in policy-making in Nigeria
- build smarter, more responsive cities in Latin America

A report from the Open Data Institute, exploring how world leaders can use and promote open data to tackle global challenges post-2015, has been launched: '[Supporting sustainable development with open data](https://theodi.org/news/new-report-reveals-how-open-data-is-fuelling-problemsolving-in-the-developing-world-from-mapping-ebola-to-protecting-banana-crops)'.



<https://theodi.org/news/new-report-reveals-how-open-data-is-fuelling-problemsolving-in-the-developing-world-from-mapping-ebola-to-protecting-banana-crops>

Protecting banana farmers' livelihoods in Uganda



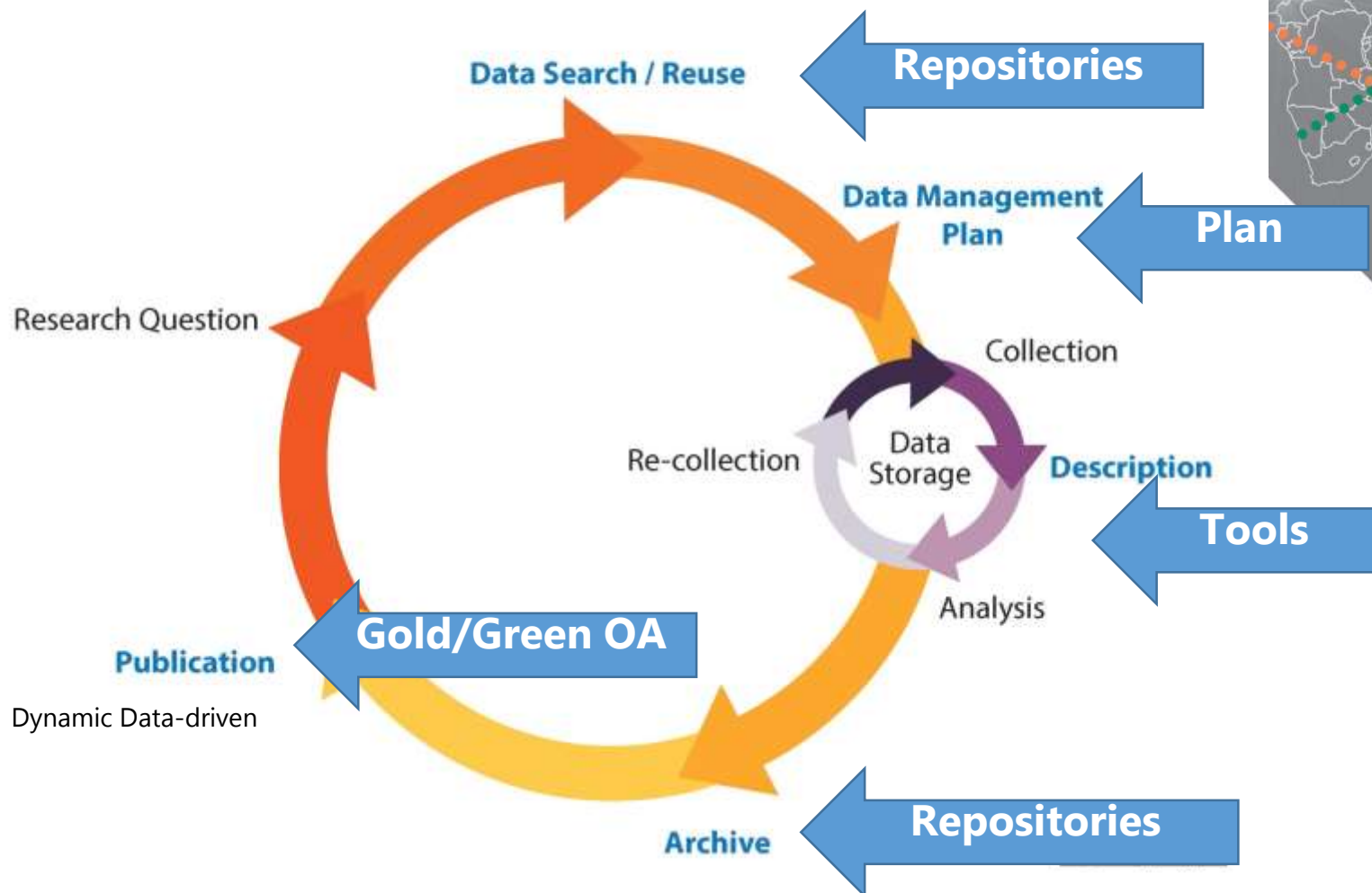
Using maps to increase access to education in Kenya



Monitoring child malnutrition (Uganda)



The Research Data Curation Lifecycle



Original Research Data Lifecycle image from University of California, Santa Cruz
<http://guides.library.ucsc.edu/datamanagement/>

African Open Science Platform

<http://africanopenscience.org.za/>

- Platform = opportunity to engage in dialogue, create awareness, connect all, provide continental view
- Funded by SA Dept. of Science & Technology through National Research Foundation
- 3 years (1 Nov. 2016 – 31 Oct. 2019)
- Managed by Academy of Science of South Africa (ASSAf)
 - Through [ASSAf hosting ICSU Regional Office for Africa](#) (ICSU ROA)
 - Direction from CODATA

Accord on Open Data in a Big Data World



Call to Endorse

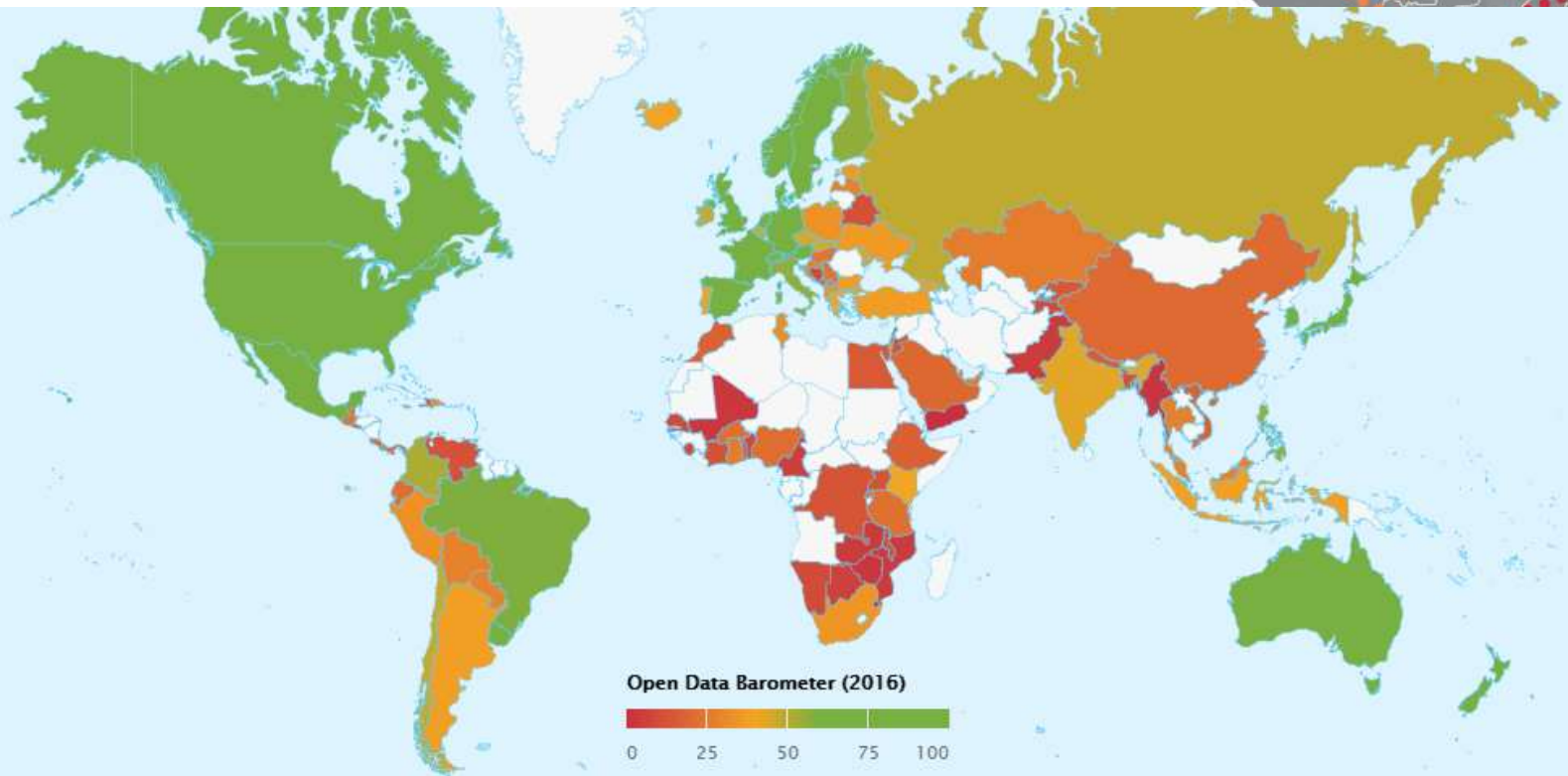
- Proposes comprehensive set of principles
- FAIR Principles
- Values of open data in emerging scientific culture of big data
- Need for an international framework
- Provides framework & plan for African data science capacity mobilization initiative

Key Stakeholders

- Global Network of Science Academies (IAP)
- International Council for Science (ICSU)
- The World Academy of Sciences (TWAS)
- Research Data Alliance (RDA)
- **NRENs** (Internet Service Providers for Education)
- **Association of African Universities** (AAU)
- Network of African Science Academies (NASAC)
- African Research Councils (incl. DIRISA, funders)
- African Universities
- **African Governments**
- Other

Data Repositories on re3data.org





The Open Data Barometer

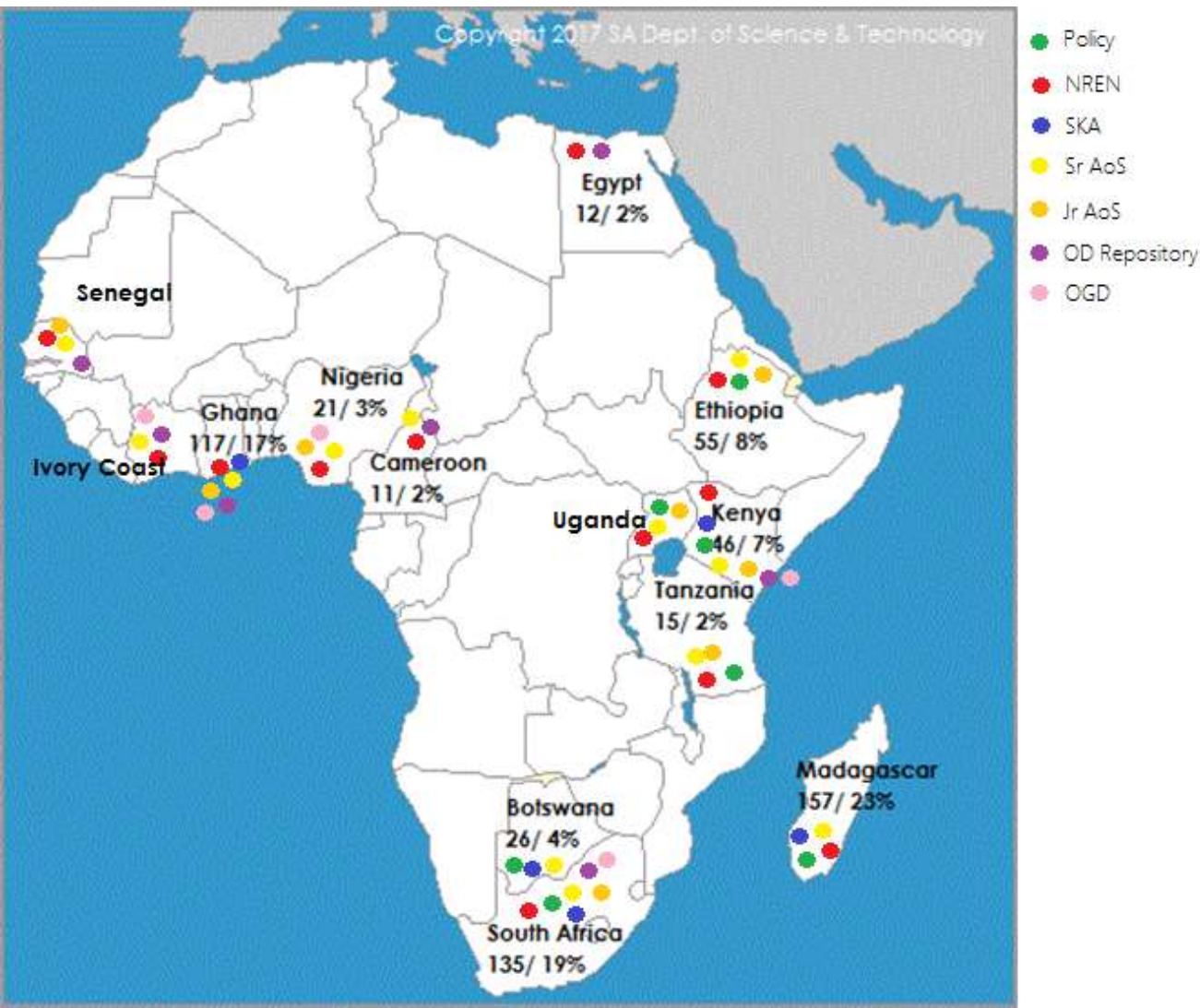
A global measure of how governments are publishing and using open data for accountability, innovation and social impact.

http://opendatabarometer.org/?_year=2016&indicator=ODB

AOSP Focus Areas



African Landscape: Open Science Open Data Activity & Database



National Funder: 2
Science Granting
Councils Initiative: 12
Policy: 7
NREN: 29
SKA: 8 (Botswana, Ghana,
Kenya, Madagascar, Mauritius,
Mozambique, Namibia, Zambia)
Sr AoS: 28
Sr & Jr AoS: 9
OD Repository: 10
OGD: 10

Actions & Deliverables Year 1

- AOSP Side Event to the SFSA 2016 - **Launch**
- AOSP Workshop AAU 2017 (Ghana) - **Policy**
- **Ethiopia** – EAS, ESC, EthERnet 4 Aug. - **Policy**
- Madagascar Meeting & Workshop 2017 – **Policy & Capacity Building & Infrastructure**
- Botswana National Open Data Open Science Forum 2017 – **Policy & Infrastructure**
- UbuntunetConnect Ethiopian Workshop 2017 – **Capacity Building & Infrastructure**
- Database & Networks – Surveys etc

Policy Framework

- Policy provide **guidance** & see to **well-being** of all citizens – political will
- Starts with 1 person – **proposal**
- Identify **key stakeholders**
- Establish **national forum**
- Work with AOSP – workshops/meetings towards roadmap
- Policies to address (also see existing policies, templates):
 - FAIR & GEO Principles
 - Raw vs Processed/other data
 - Licensing
 - Intellectual Property Issues
 - Social & Natural Sciences

Intellectual Property Rights Policy

- Incl. copyright
- WIPO (World Intellectual Property Organization)
- OECD Principles & Guidelines for Access to Research Data from Public Funding
- "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."

<https://ipstrategy.com/2016/12/05/a-new-look-at-intellectual-property-and-innovation-in-africa/>

African Open Science Policies Progress

- **JKUT (Kenya)** Institutional Open Data Policy
- **Uganda** Draft Open Data Policy
- **Madagascar** lobbying for Open Data Policy
- **Ethiopia** – EAS, ESC, EthERNet - Policy
- **Tanzania** draft policy towards Open Access
- Towards a White Paper on Open Research Data Strategy in **Botswana**
- Funder Policy: National Research Foundation (NRF)(SA)
- White Paper on Science, Technology and Innovation in **South Africa**



SA-EU STRATEGIC PARTNERSHIP

THE DIALOGUE FACILITY

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

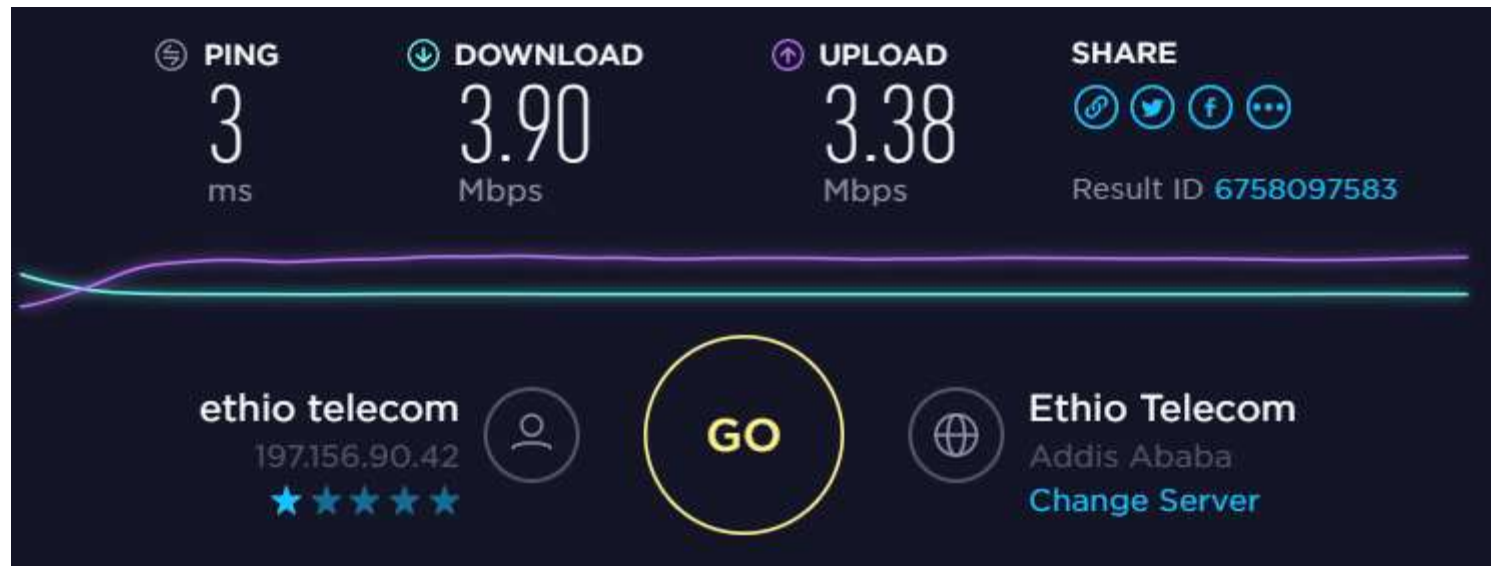
- NRENs – Level 6 Elaborated Service Offering

[An NREN Capability Maturity Model – Duncan Greaves](#) (2015, Tertiary Education Network)

- 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
- SKA to generate 35,000-DVDs-worth of data every second - “the whole world wide web every day” (Fanaroff)
- Internet of Things, Videos, Webinar, Online Learning



Internet Speed



Infrastructure Framework

- Continental, National, Regional, Community, Institutional
- Roadmaps needed
- Centres for High Performance Computing, Supercomputers
- “Some of the early stages in NREN formation cannot take place at all if the policy or regulatory environment is unduly hostile. In such cases, **policy and regulation** are themselves the spaces in which intervention is required.” – Duncan Greaves (2015)

In Case You're Wondering, Here's Why Data Is So Expensive in South Africa

Blame limited spectrum and expensive infrastructure.

© 22/08/2017 16:30 SAST | Updated 22/08/2017 16:31 SAST



Research ICT Africa is a regional ICT policy and regulation think tank active across Africa and the Global South

[Home](#) [1GB Basket](#) [Value for Money Index](#) [OECD Basket](#) [Research ICT Africa Website](#)

Cheapest price for 1GB basket in Africa by country

- OECD Basket: 1GB data based on prepaid data top ups or bundled top ups

- All Operators

- US Dollars (USD)

<http://www.researchictafrica.net/pricing/ramp.php>

Search...	2014			2015				2016				2017	
Countries	Q2 ↕	Q3 ↕	Q4 ↕	Q1 ↕	Q2 ↕	Q3 ↕	Q4 ↕	Q1 ↕	Q2 ↕	Q3 ↕	Q4 ↕	Q1 ↕	Q2 ↗
Egypt	-	-	-	-	13.83	3.59	3.18	3.13	2.82	2.82	1.74	1.41	1.25
Tunisia	-	8.66	8.21	2.59	5.16	5.11	4.99	4.94	4.83	4.54	1.77	1.76	2.04
Guinea	-	-	-	-	-	7	6.66	3.95	4.01	3.35	3.3	3.24	2.2
Tanzania	7.92	7.86	7.53	5.64	6.25	5.98	6.03	5.98	0.89	2.39	2.99	2.29	2.27
Ghana	4.7	4.29	4.72	4.38	3.75	3.9	3.96	3.86	3.91	2.54	2.46	2.27	2.34
Rwanda	11.92	11.78	11.77	4.42	4.39	4.31	4.13	4.06	4.03	3.91	3.73	3.06	2.43
Mozambique	6.39	6.58	5.08	4.78	4.44	4.03	3.62	3.35	2.87	2.27	2.12	2.27	2.71
Uganda	5.85	5.79	5.57	5.24	3.97	3.45	3.49	3.55	3.62	3.59	3.44	3.37	2.82

OS Repository Software

- Take ownership, responsibility, develop skills – towards empowerment, reduced dependency, sustainability
- Examples:
 - Dataverse
<https://dataverse.org/>
 - CKAN
<https://ckan.org/>
 - DKAN
<https://getdkan.org/>
 - Nesstar
<http://www.nesstar.com/software/publisher.html>



<https://www.coretrustseal.org/about/>

Data Repositories vs Social Media

- Social media sites/3rd party software:
 - Connect researchers sharing interests
 - Marketing data
 - Sites belong to third parties – and data
- Repository:
 - Supports export/harvesting of metadata
 - Offers long-term preservation
 - Non-profit – no advertisements
 - Uses open standards and protocols
 - Copyright

Capacity Building Framework

- Engineers, Statisticians, Data Scientists, Librarians, Data Curators, Researchers, System Administrators, Policymakers, Auditors, Data Centre Managers, Data Architects – Wim Hugo
- Different skills for different categories of data workers
- Existing workshops presented
- Tertiary curricula need to adapt more rapidly
- Never too early to learn to work with data, program

Recommended Data Skillset

- FAIR & GEO data principles, data citation, version control, licensing, persistent Identifiers: DOIs, ORCID IDs
- Data Repositories
- Command Line Interpretation
- Software Development, Programming
- Data Organisation
- Data Cleaning
- Data Management & Databases
- Data Analysis & Visualisation (incl. programming)
- Machine Learning
- Artificial Neural Networks
- Research Data Management Planning

Research Data Management Plan



<https://github.com/DMPRoadmap>



Research Proposal
Ethics Committee
Funder
Data Server &
Repository
Etc.

**Biosignature Suites: Using Connections
between Microbes and Minerals to
understand Biogenic Carbonates**

A Data management plan created using the DMPTool

Creator(s): DMP dmpcurator

Affiliation: University of California, Office of the President

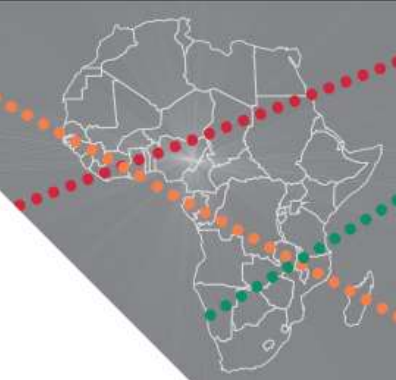
Last modified: May 30, 2014

Copyright information: The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creators as the source of the language used, but using any of their plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal.



Sections	Questions
Data Collection	<ul style="list-style-type: none">- What data will you collect or create?- How will the data be collected or created?
Documentation and metadata	<ul style="list-style-type: none">- What documentation and metadata will accompany the data?
Ethics and Legal Compliance	<ul style="list-style-type: none">- How will you manage any ethical issues?- How will you manage copyright and Intellectual Property Rights (IPR) issues?
Storage and Backup	<ul style="list-style-type: none">- How will the data be stored and backed up during the research?- How will you manage access and security?
Selection and preservation	<ul style="list-style-type: none">- Which data should be retained, shared, and/or preserved?- Are any restrictions on data sharing required?
Responsibilities and Resources	<ul style="list-style-type: none">- Who will be responsible for data management?- What resources will you require to deliver your plan?

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.

DEC 3, 2016 @ 12:13 PM 23,975 VIEWS

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: LPI 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

Enquiries regarding remuneration and benefits: Human Resources Client Service Centre on 021 808 2753

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

- Mechanisms that acknowledge publication of datasets and to promote data sharing
- Funder requirements changing
- 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

Closing Remarks

- Collaborate & learn from one another – strength in diversity, collaboration, connectedness
- Lifelong self-learning towards empowerment
- Take ownership & collect/curate data in ethical way
- Trusted & valid data managed in trusted way
- Downloaders vs Uploaders
- Exploit data for the benefit of society (Min Naledi Pandor)
- Tell the African story, in an African way

Thank you!

Ina Smith

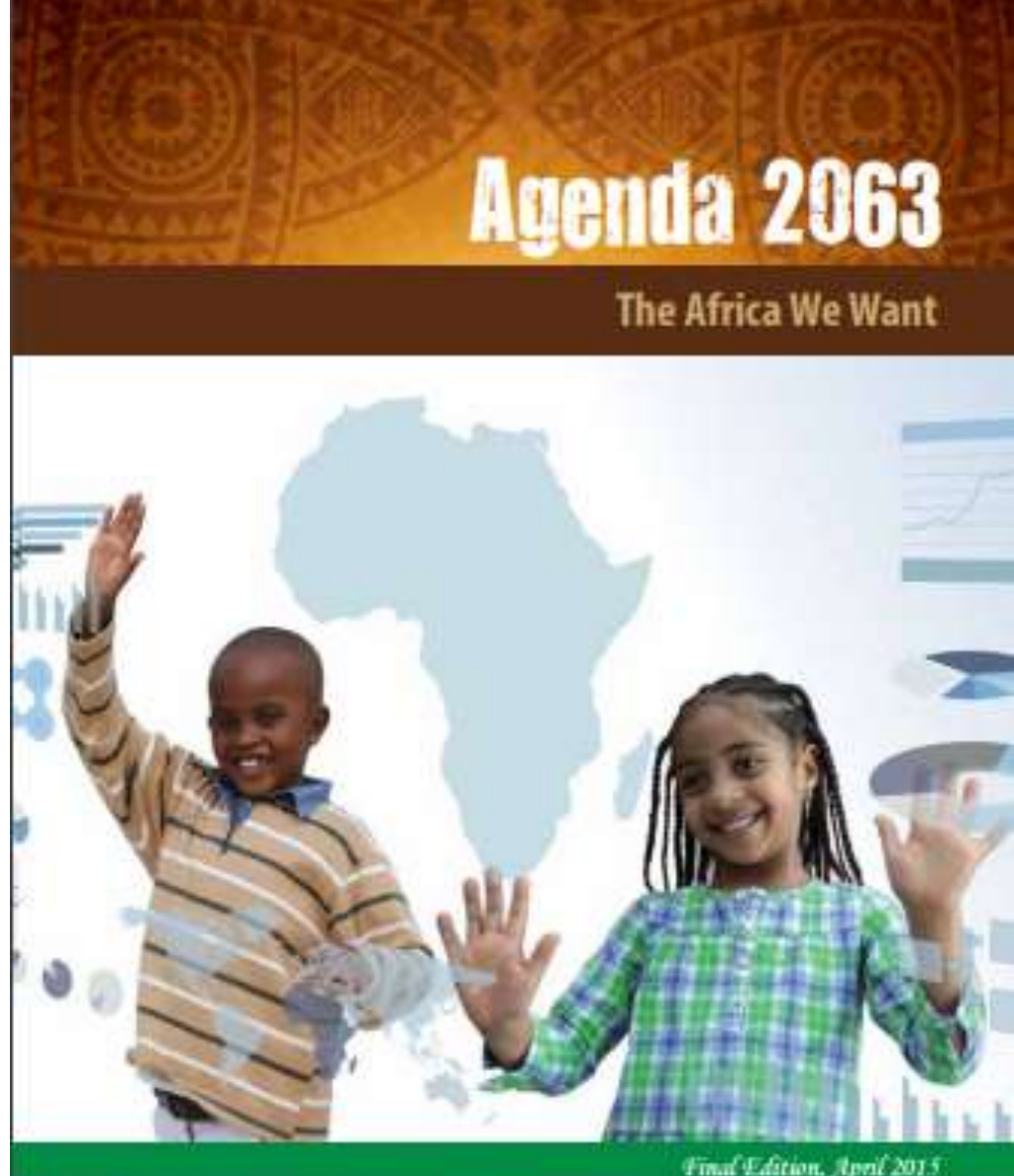
ina@assaf.org.za

Susan Veldsman

susan@assaf.org.za

<http://africanopenscience.org.za>

@AOSP_Official



Tim Berners-Lee

"Let us create the **Web We Want**,
for the **World We Want**".