

African Open Science Platform

Infrastructure Services Toward Data Sharing

Benjamin Aribisala

Dean of Science and Professor of Computer Science,
Lagos State University, Nigeria
Benjamin.Aribisala@lasu.edu.ng

Outline

- ❖ Background
 - ❖ Open Science and Open Data
 - ❖ Africa Open Science Platform (AOSP)
 - ❖ What is ICT Infrastructure Framework?
 - ❖ Why ICT Infrastructure Framework?
 - ❖ Expected Content of the Framework
 - ❖ Progress so far
-

Background

There have been increased concerns on **repeatability** in scientific research

A recent study¹ attempted to replicate 18 articles on **microarray gene expression analyses**

Results

- ❖ Reproduced in principles- 2
- ❖ Reproduced with discrepancies - 6
- ❖ Not reproducible - 10

Reasons for Failure

- ❖ Data not available
- ❖ Methods unclear
- ❖ Software not available
- ❖ Contradictory results

More than repeat and reproduce²

- ❖ Repeatability - Change Time
- ❖ Reproducibility – Change Scientist
- ❖ Replicate- Scientist and Location
- ❖ Reuse – Different experiment

1. Ioannidis et al., 2009. Repeatability of published microarray gene expression analyses

2. Drummond C, 2009, Repicability is not Reproducibility: Nor is it Good Science

Background

Collaboration is the key

- ❖ To solve the common problems facing humans
- ❖ To ensure repeatability, reproducibility, replication and re-usability

Sharing

- ❖ data, ideas, software, infrastructure, etc

We also need to be open

Open Science and Open Data

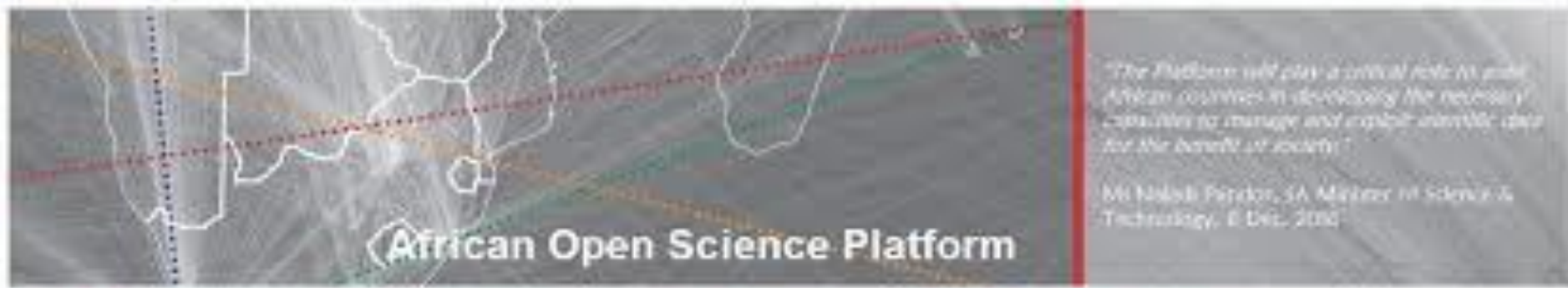
“**Open Science** is the practice of science in such a way that others can **collaborate and contribute**, where research data, lab notes and other research processes are **freely available**, under terms that enable **reuse, redistribution and reproduction** of the research and its underlying **data and methods**.”

- FOSTER Project, funded by the European Commission.

Open data is any data that is **freely** available to everyone to **use and re-use without restrictions** from copyright, patents or other mechanisms of control.

Free	reusable
collaborate	methods
Data	knowledge

Africa Open Science Platform (AOSP)



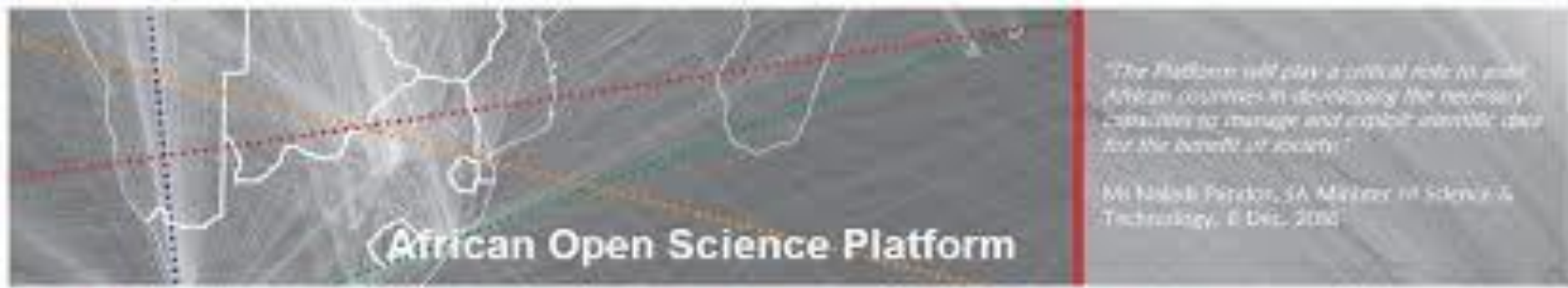
African Open Science Platform

🌐 **AOSP** is an initiative in Africa that promotes the values of open data in the emerging scientific culture of big data.

🌐 Goals of AOSP

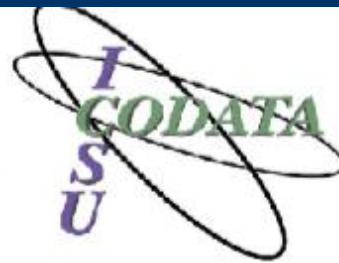
- To develop a standard **open data policy** for Africa
- To develop a robust **capacity building** plan in open science
- To **advocate** open data and open science in Africa
- To **coordinate** series of actions channelled towards promoting openness in science

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



African Open Science Platform

- Funded by the National Research Foundation (NRF) (SA Dept. of Science and Technology)
- Directed by CODATA (ICSU)
- Managed by Academy of Science of South Africa (ASSAf)
- Through ASSAf hosting International Council for Science (ICSU) Regional Office for Africa



ICSU and CODATA

- 🌐 ICSU: International Council for Science – consists of 17 interdisciplinary bodies e.g. CODATA <http://www.icsu.org/>
- 🌐 CODATA: Committee on Data for Science and Technology
<http://www.codata.org/>
- 🌐 **Mission:** Strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.

ICT Infrastructure Framework

What is ICT Infrastructure Framework?

- ❑ ICT Infrastructure encompasses all the **devices, networks, protocols and procedures** that are employed in the telecoms or information technology fields to foster interaction amongst different stakeholders

 - ❑ Physical and Virtual Resources that support
 - the flow
 - storage
 - processing and
 - analysis of data

 - ❑ A Framework is a **basic structure or guideline** underlying a system or concept.
-
-

ICT Infrastructure

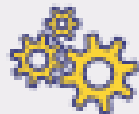
Infrastructure Solutions



ICT Automation



ICT Virtualisation



IaaS/PaaS



Backup and Archiving

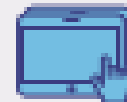
End User Devices and Platforms



Notebooks



Desktops and Workstations



Smart Devices

Server Infrastructure



Storage



Servers



Databases and Middleware



Web Servers

Data Networks



LAN/WAN



WiFi



VoIP

Data Centres



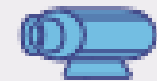
Server and Data Switchboards



UPS



Access Control Systems



CCTV

Why ICT Infrastructure Framework?

- 🌐 Researchers need data (FAIR Data)
 - 🌐 Good data management require good infrastructure
 - 🌐 Open Science and Open Data require infrastructure sharing
 - 🌐 Most Africa Nations have poor budgetary allocation for ICT infrastructure
 - 🌐 There is no guideline on minimum requirement for ICT infrastructure to support open science
 - 🌐 There is no coordination of resources and investment in shared infrastructures
-
-

Expected Content of the Framework

- The main goal of the project is to develop ICT Infrastructure Framework
 - To support Open Science and Open Data
 - To propose a generic guidelines on what **stakeholders** need to do to **make the sharing of research data a reality**.
 - Framework to cover the following
 - 1. Internet connectivity & role of NRENs
 - 2. Hardware & Software requirements
 - 3. Data Lifecycle (collection, processing, analysis, identifiers, citation, curation)
-

Attributes of the Framework

- ❑ Framework to be service driven
 - Start with a service in mind and then develop the framework for it.
 - Services vary from one region to the other
 - A fairly generic or common services could be the focus
 - ❑ Framework should be focused on promoting open science and open data.
 - ❑ Framework to enable the **stakeholder/provider** offer the service(s) that will meet the needs of the **user**.
 - Requirements – e.g internet, hardware, software and data lifecycle.
-
-

Internet Connectivity

- ❑ What is a good Campus Network ?
 - ❑ Bandwidth requirement
 - ❑ Network policy
 - ❑ Role of NRENs
 - ❑ Role of Government and institutions
-
-

Research and Education Network (NREN)

- ❑ The main purpose of a NREN is to provide high value-added services to support the needs of the research and education communities within a country e.g. internet, resource sharing, etc
 - ❑ NREN is organized under a top-bottom layered model
Global/International connectivity;
 - Global / International connectivity;
 - Regional RENS e.g. WACREN
 - National REN
 - Campus networks e.g. LANrks
 - ❑ To recommend to governments and institutions
 - Join NREN
 - Have a good campus network
 - ❑ Identify Requirements for joining NREN
 - ❑ Define a good campus network
-
-

Hardware and Software Requirements

Identified Services

1. Federated identity
 2. Data repository
 3. Cloud services
 4. Data Transfer Service e.g. Globus, Aspera
 5. Prioritised suite/portfolio of services
-
-

Federated Identity

"Every robust collaboration is built on trust. Federated Identity Management allows any organization to express their local trust relationships with faculty, staff, and students in a way that can be understood and leveraged by the global R&E community." Lucy Lynch (NSRC)

- 🌐 Federated Identity is about using a local privilege at global level
 - 🌐 There is a need to promote and Recommend Federated Identity to all Governments and R&E community
 - 🌐 **The Framework will contain the requirement for an institution/ individual to have Federated Identity**
-
-

Data Repository

- Need to recommend to all institutions to have repositories
 - At least repository for library
 - With Federated Identity , the repositories are accessible from other institutions
 - Recommended technologies for building a repository
 - The Framework will define what makes a good data repository
-
-

Cloud Computing

- 🌐 Cloud computing refers to a number of networked servers
 - to process
 - store and
 - deliver services using the internet
 - 🌐 Provides avenue for the creation and use of “on-demand” computing services which can be grouped as
 - Infrastructure as a service (IaaS)
 - Platform as a Service (PaaS)
 - Software as a service (SaaS)
-
-

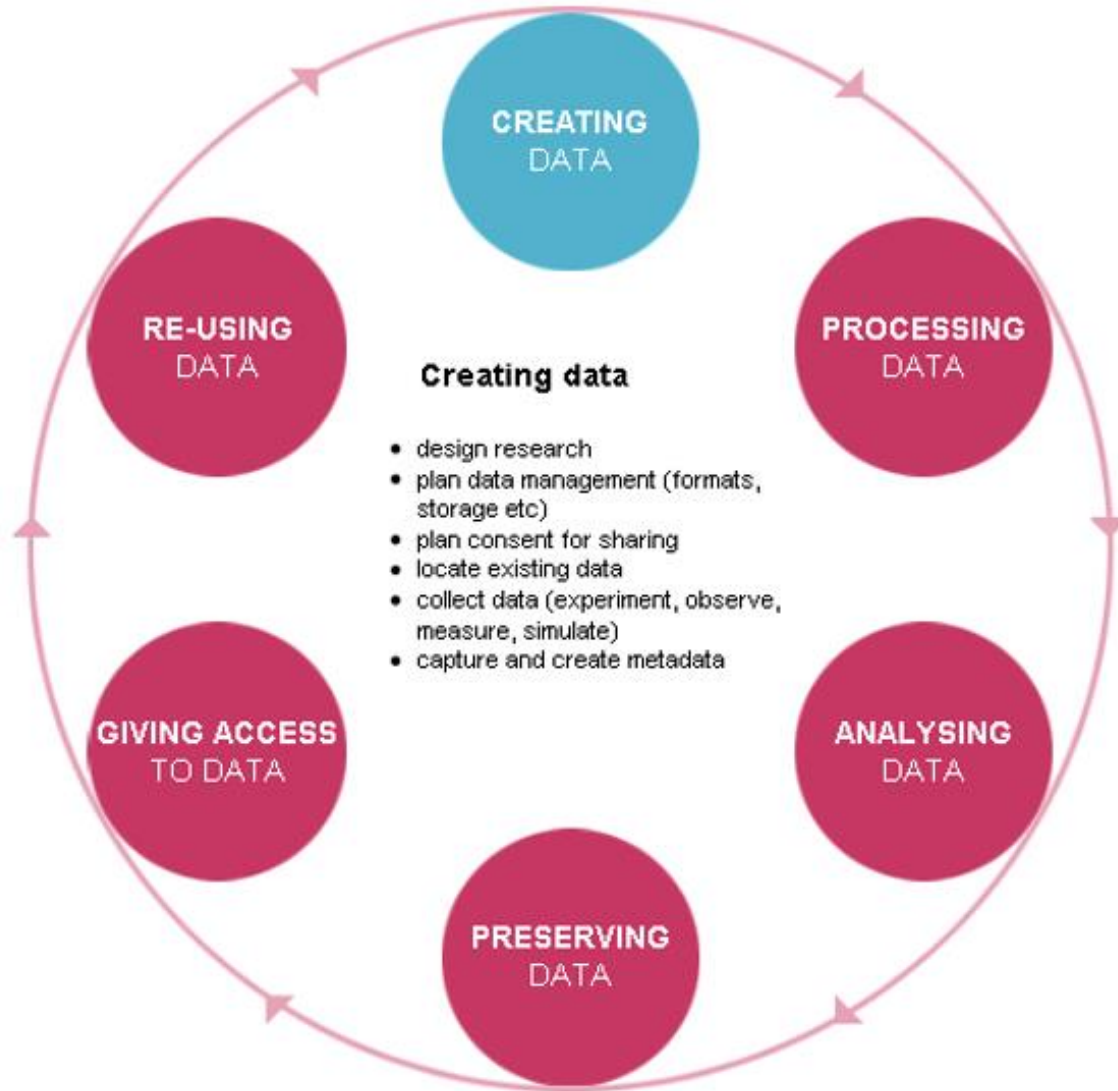
Cloud Computing

- 🌐 The minimum services a research cloud should provide include
 - controlled user access,
 - a variety of open stacks and specifications,
 - good end user documentation,
 - networking,
 - access to storage (shared or private),
 - facilitate meta-data searching of the storage,
 - able to accommodate the use of containerization technology,
 - have data management and governance policies in place.
 - 🌐 Cloud service can be recommended for institutions
 - 🌐 Any institution with a cloud service could be called smart
 - 🌐 **Framework will contain the requirements for offering cloud services**
-

Data Transfer Services

- A good data transfer service must be
 - fast
 - use protocols that do not interfere with network operations
 - No significant degradation over long haul networks.
 - The framework will state the requirements for good data transfer
 - It will also contain the meaning of fast in terms of transfer speed
-
-

Data Lifecycle



Appreciation

1. Ousmane Moussa
 2. Sumir Panji
 3. Anwar Vahed
 4. Renier van Heerden
 5. Jean Ganglo Cossi
 6. Leon Staphorst
 7. Bruce Becker
 8. Susan Veldsman
 9. Ina Smith
 10. Nozuko Hlwatika
-