





A national initiative of the Department of Science and Innovation and implemented by the CSIR.



DIRISA for Open Data and Open Science

Stellenbosch University Symposium 13 October 2021

Dr Anwar Vahed NICIS - DIRISA

Outline

- About NICIS & DIRISA
- Open Data, Open Science, Open Minds
- Global Open Science
- South African Open Science Cloud (SAOSC)
- South African Open Data Cloud
- A Tiered Data Infrastructure
- Open Data Management and Open Access

National Integrated Cyberinfrastructure System (NICIS)

- Common services
 between any two Service
 Areas
- Tight integration among Areas.
- Development of cross-cutting skills internally and within the community



Data Intensive Research Initiative of South Africa (DIRISA)



Advocate & coordinate

- Stakeholder engagement
- National R&D projects



Build data infrastructure

Build and maintain Tier 1 node and servicesStart T2 domain nodes



NATIONAL INTEGRATE

DIRISA



Manage research data

- RDM services
 - Policies & guidelines



Develop capacity and skills

- Support e-Research programmes
- Coordinate training & workshopa

Open Data, Open Science, Open Minds

"...outputs of publicly funded research are publicly accessible with no or minimal restriction" [OECD, 2015] Increased efficiency & return Cost-benefit analysis: FAIR vs not FAIR

- Multiplier effect & spill-over (COVID-19 R&D a good example)
- Transparency, integrity & quality
- Citizen science & awareness



Global Open Science

- OECD, CODATA, UNESCO, WDS, Global Science Forum
- Underway or emerging
 - European Open Science Cloud (EOSC)
 - China Science and Technology Cloud
 - Australian Research Data Commons (ARDC)
 - Malaysian Open Science Platform (MOSP)
 - African Open Science Platform (AOSP)
 - Latin America, Korea, Canada's NDRIO and Germany's NFDI
- Global Open Science Cloud (GOSC)



Towards a UNESCO Recommendation on Open Science

Building a Global Consensus on Open Science

South African Open Science Cloud (SAOSC)

- SA-EU Open Science Framework Report (2018)
- Dialogue with EU (EOSC)
- SAOSC: Build on existing resources
 - SA Research Cloud (NICIS)
 - SA research data infrastructure (DIRISA)
 - Develop policies and governance structures
 - Federate into other local resources, and internationally into Africa & Global Open Science Cloud (GOSC)
 - Advocate

Global & National Data Commons



South African Research Data Commons (SA Research Data Cloud)

Phase 1: Research Data Management



Phase 2: Collaborative Research Environments

> My analytics & workflows My data sets and outputs My communities

Virtual Machines Jupyter Notebook

Data Access: Open by Default

FAIR: "As open as possible; as closed as necessary"

Closed	Shared			Open
Internal access Private Confidential Sensitive 	 Named access Assigned by contract Regulation authorised 	Group based access • Project assigned • Selected membership	 Public access Licence that limits use Terms and conditions 	AnyoneOpen to publicNo limits on use
Surveillance data	Drivers licences	Genomic data	Geospatial data	Weather data
Personal	Private			Public
Small	Medium			Big

SA Open Research Cloud

OLab **T** Experimental, prototyping and development of bespoke services and modules for local conditions Sandboxes for CEPH, iRODs, POCs, etc



Pilot Cloud

- Verification, Alpha and
- Beta testing of services and applications

Pre-production Cloud

Self-developed cloud providing failover, to be moved into full production

Production Cloud

Operational services supporting public **T** sector C () User cloud services for compute- and data- intensive research

> COVID. **Bioinformatics**, Climate Change,...

DIRISA Data Cloud Portal



Things to do

Time lined implementation roadmap:

- Policies and governance
- Interoperability arrangements, & Digital Object management
- Design and expand cyberinfrastructure
- Funding
- Develop skills
- Communicate and advocate

Is it urgent? Yes, challenges are global and surprisingly common

Concerns About Open Science

- "This is my data and research..." [Accreditation & Helicopter Science]
- Increase in digital divide between researchers in high-income and low-income countries
- Confusion between "fast" and "open" science
- Confusion between "open" and "private"
- Reliability of data, models, methods and analysis disseminated



NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM



Thank you

Anwar Vahed avahed@csir.co.za





