

DIRISA

Data Intensive Research Initiative of South Africa



Data Intensive Research in South Africa

Galaxy Africa

3-5 April 2018

Accelerating Data Intensive Research

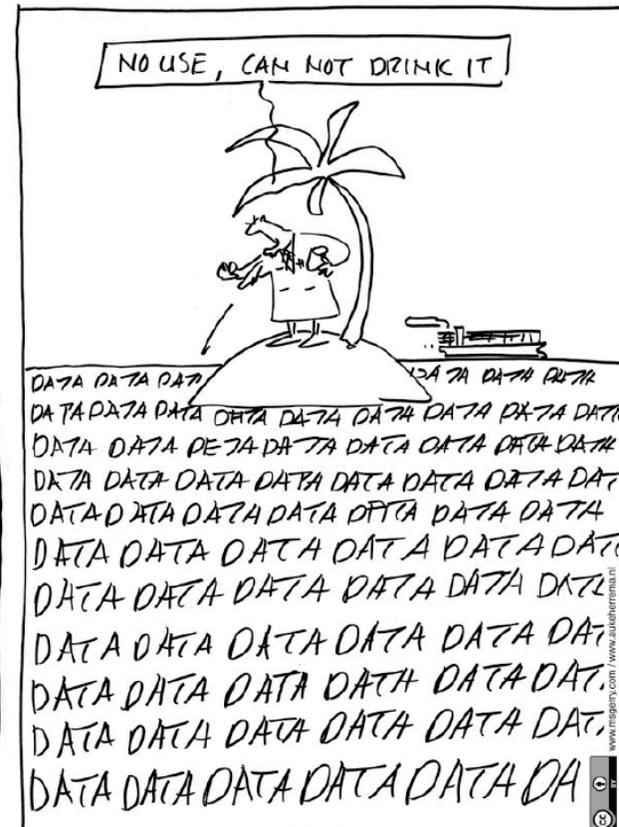
“We need to get greater value (benefit, impact) from our investments in data”



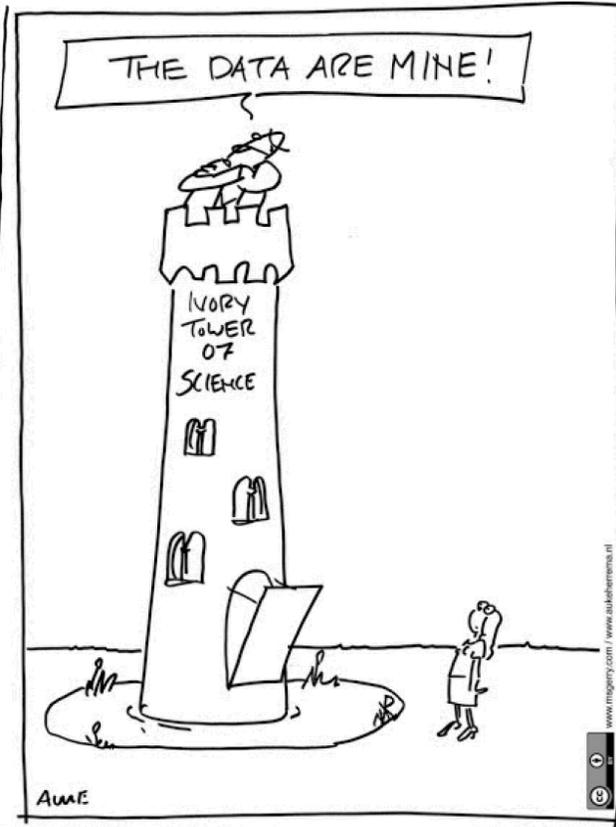
REUSABLE DATA



DATA SHARING



DATA OCEAN

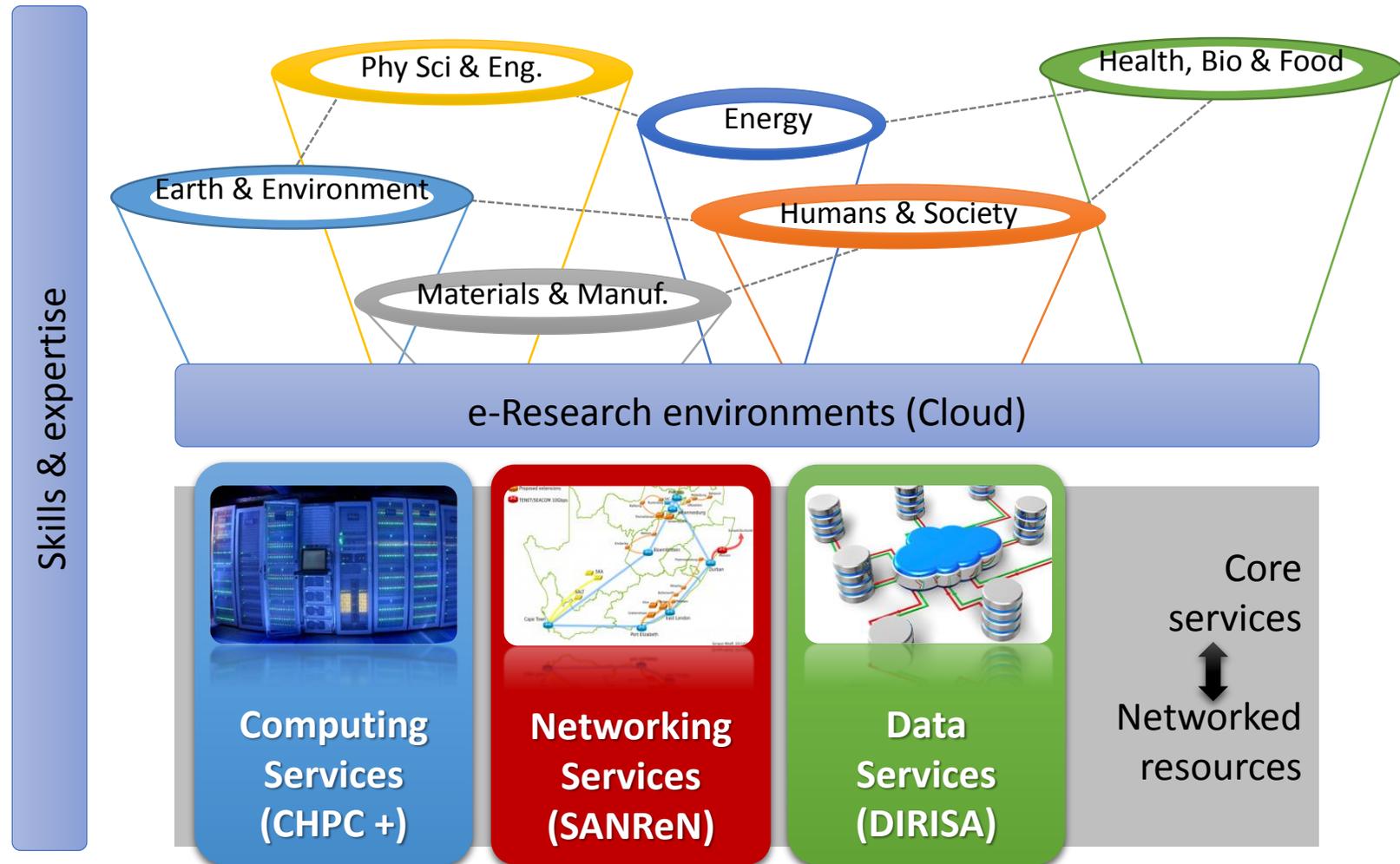


SCENE FROM THE PAST?

National Integrated Cyberinfrastructure System

National Integrated Cyberinfrastructure System (NICIS)

- Advanced Integrated cyber platform offering services for
 - HPC
 - Data
 - Networking
- Priority science and education
- Overarching coordination implemented by CSIR



DIRISA Objectives



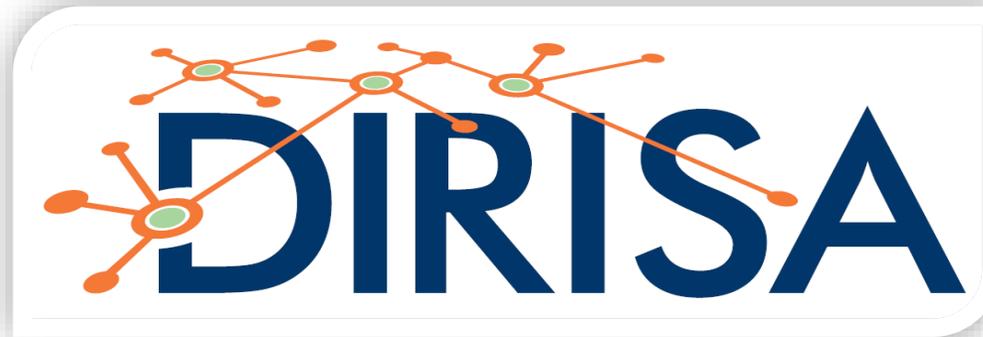
Advocate and coordinate

- R&D initiatives
- Stakeholder workshops



Build national data infrastructure

- Build and maintain Tier 1 nodes and services
- Start Tier 2 domain nodes



Research data management

- DMP tool
- PID service
- User policies and practices

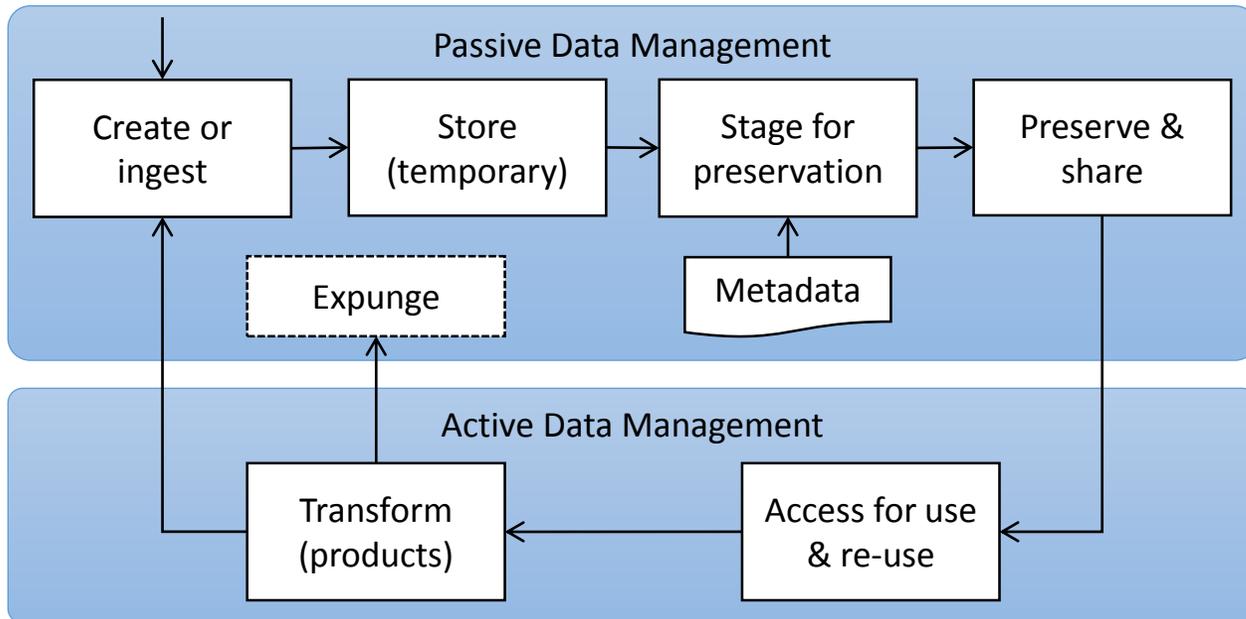


Develop human capital and skills

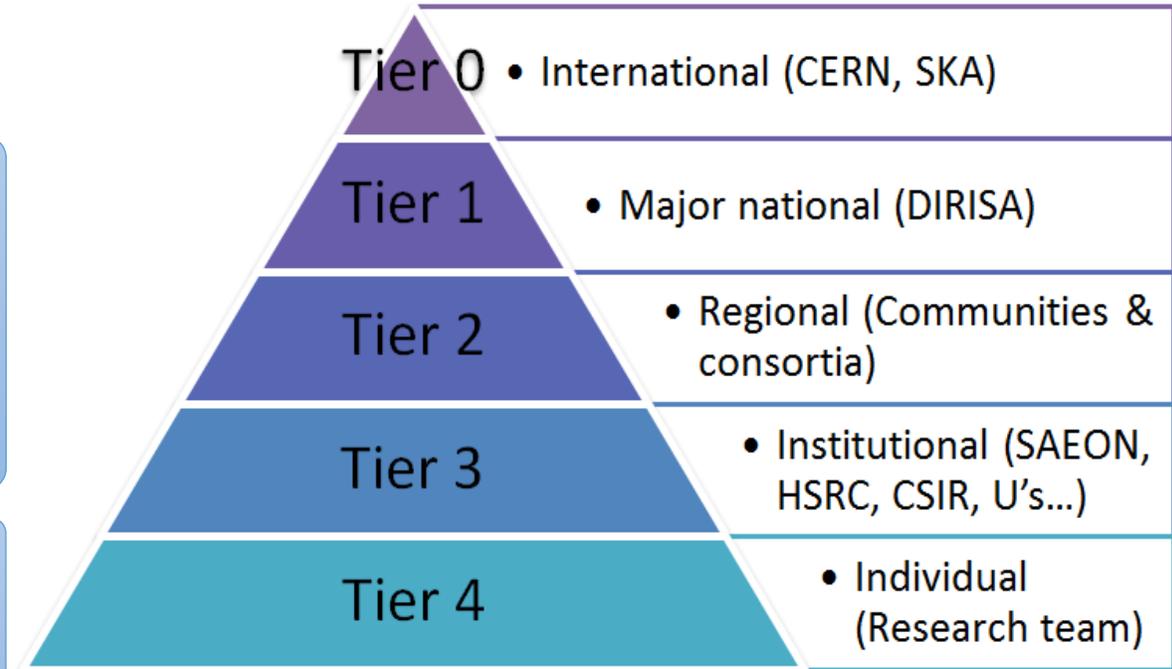
- e-Science postgraduate programs
- Conferences and training workshops

National Data Infrastructure

- “I just want to store/preserve my data (reliably)”
- “I just want to share my data (in a controlled way)”
- “I just want to process my data”

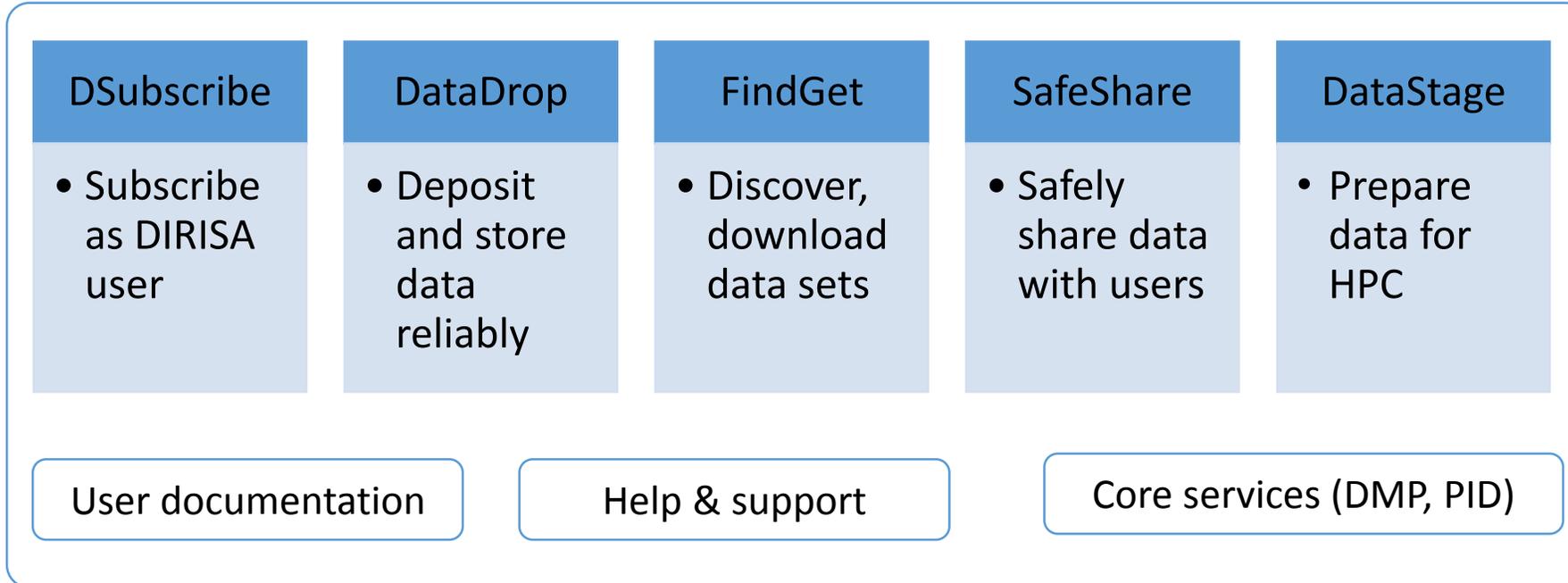


Tiered Data Infrastructure



South African National Data Infrastructure (SANDI)

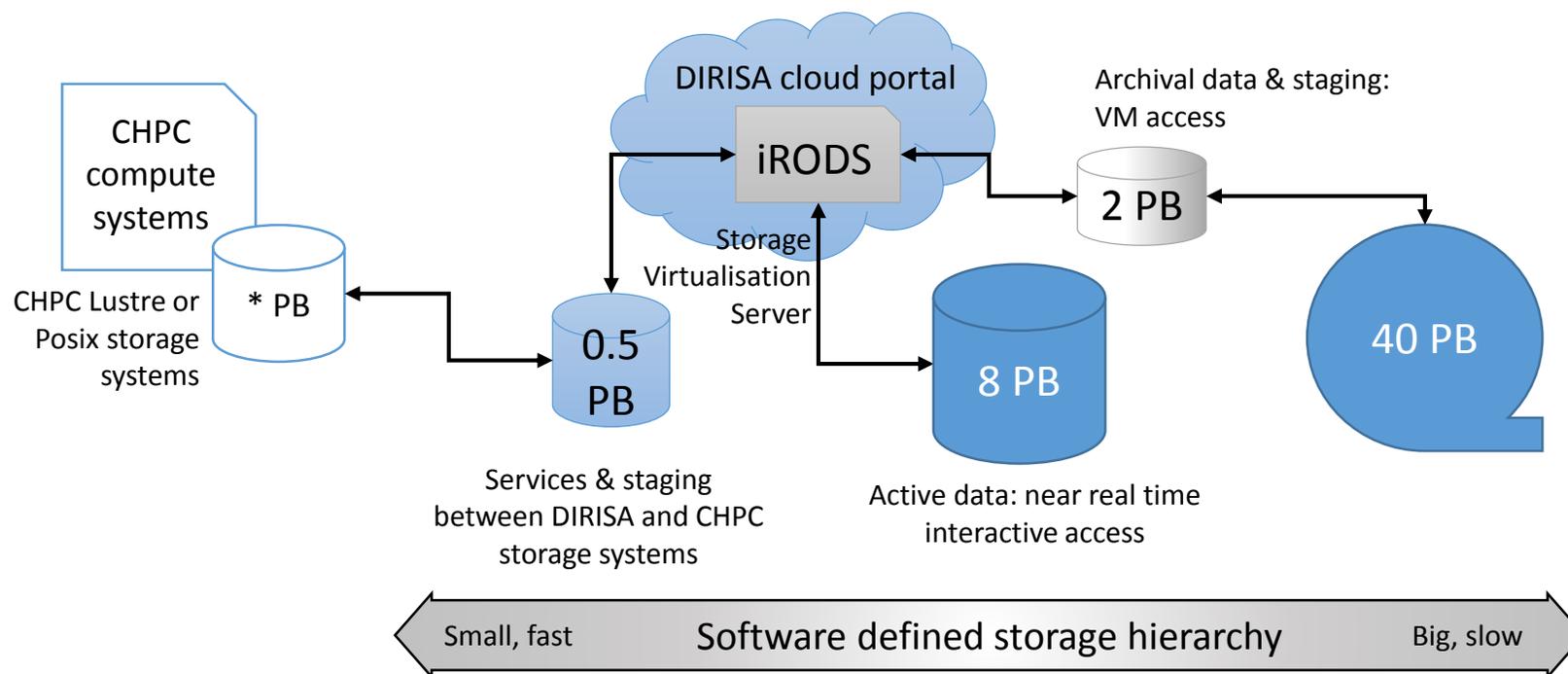
Phase1 Research Data Management



Phase2 Collaborative Research Environments

- My data management plans
- My workflows
- My data sets and outputs
- My communities

Architecture



Requirements

- Open (FAIR) Data & Open Science
- Federated locally and globally ("One-stop-shop" catalogue)
- Certified as Trusted Repository
- Linked to funder systems
- Suite of services for RDM and data analytics

Exemplars

EUDAT
ANDS
JISC
Data.gov

Currently

1. Data Infrastructure and services
 - DIRISA Tier 1 (8PB) store
 - Regional Tier 2 Node in Western Cape (ILIFU)
2. Capacity development
 - National e-Science Masters (NEPTTP)
 - Data Science training (Data carpentry, RDM courses and workshops)
3. Data management
 - PID Allocation: (Handle and DOI registries)
 - DMP Tool
 - DIRISA portal single sign-on
 - Policies across data life cycle
4. Outreach and coordination
 - Conferences and workshops
 - Africa Open Science Platform
 - Big Data strategy

Western Cape Regional Tier 2 Data Node (ILIFU)

Consortium of UCT (Lead), UWC, SU, CPUT, SARAQ, [Sol Plaatje]

Focus: Astronomy & Bioinformatics

- Establish shared data infrastructure supporting multi-institutional and multidisciplinary data intensive research
- Accelerate human and infrastructure capacity building in data intensive research
- Build on and strengthen existing initiatives
- Encourage national and international collaborative participation (academia & industry)
- Harness co-investment in cash or in kind from public and private partners

ILIFU for Data-intensive Biology

- Shared access to data (promote reuse)
- Collaborative multi-institutional research
- Educational resource
 - Postgraduate student research
 - Research data management
 - Promote Open (FAIR) Data & Open Science principles
- Shared licensing for analytical tools
- RDM: Figshare services
- Hardware platform expanded for bioinformatics

Thank you

DIRISA SERVICES

The Data Intensive Research Initiative of South Africa makes the following tools and services available to the research community.



USER SUBSCRIPTION

The DIRISA User Subscription is used to allow user's to authenticate themselves before using DIRISA services. Authentication is currently granted through SAFIRE. However, in the future, more federated identities will be managed.



SA DMP TOOL

The South Africa Data Management Planning (SA DMP) Tool is based on the DCC Online implementation. This tool is used to produce a formal document that outlines how a researcher will handle research data during the research process and after the project is completed.



DATA DEPOSIT TOOL

Research data is a critical part of the research process and while there are several cloud-based storage solutions to store data, DIRISA provides a reliable, persistent and easily accessible resource to safely share data. The development of a cloud-based data deposit tool is underway and will be available to subscribed users.



DIGITAL OBJECT IDENTIFIER

Data sets at DIRISA and trusted repositories are persistently and uniquely identifiable through a Digital Object Identifier (DOI). Apply for a DOI to be allocated to your data set through one of several such existing systems.

www.dirisa.ac.za

dirisa@csir.co.za

