

# A national research data management strategy for South African universities: the dilemma of collective action

Dr Dale Peters

Director: UCT eResearch  
University of Cape Town



UNIVERSITY OF CAPE TOWN  
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD



# Outline

Scramble for Africa  
National Integrated Cyberinfrastructure System (NICIS)  
Dilemma of collaboration





# Scramble for Africa

Data is the new oil



UNIVERSITY OF CAPE TOWN  
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD





DATA IS THE NEW OIL





S

# DATA IS THE **NEW** OIL

but do you have the resource to refine it?

[successflow.co.uk](http://successflow.co.uk)



UNIVERSITY OF CAPE TOWN  
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD





Higher education in a democratic South Africa faced huge challenges – primarily the need to achieve greater equity, efficiency and effectiveness within institutions and across the system.

[globalriskinsights.com](http://globalriskinsights.com)



UNIVERSITY OF CAPE TOWN  
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

UCT  
**eResearch**  
ACCELERATING RESEARCH



# National Integrated Cyberinfrastructure System (NICIS)

First infrastructure grant



UNIVERSITY OF CAPE TOWN  
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD



# National Integrated Cyberinfrastructure System (NICIS)

---

**CALLS FOR PROPOSALS FROM REGIONAL AND NATIONAL CONSORTIA**

**A regional Tier 2 data node and a national e-science teaching platform**

ICT-based research or e-research is impacting a broad spectrum of disciplines across the natural, medical, human and social sciences and beyond, and changing the nature of research. The South African National Integrated Cyberinfrastructure System is an initiative supported by the Department of Science and Technology and implemented by the CSIR. Each call is for a self-organised, consortium of higher education institutions and qualified partners to submit proposals.

---

## **CALL 1: ESTABLISHMENT OF A REGIONAL TIER 2 DATA NODE**

Proposals are invited from consortia of institutions (higher education, science councils, national facilities and the private sector) to establish and host a regional Tier 2 data node that will form part of the national tiered research data infrastructure. Key deliverables of the node will be the establishment of multi-institutional consortia with shared data infrastructures supporting multi-disciplinary data intensive research, accelerated human capital development and infrastructure capability building.







# ILIFU

“Cloud technology has the capacity to democratise RDM services. This not only empowers individual researchers, giving them real control over their data, but also allows distributed organisations to work together as one.”

## AFRICAN RESEARCH CLOUD (ARC)

The ARC, a collaboration between UCT, the Inter-University Institute for Data-Intensive Astronomy (IDIA) and North West University, is the prototype for a cloud-based service to researchers working in data-intensive disciplines. Established in 2016, the ARC is testing different models of data management, storage and transfer through radio astronomy and genomics projects.

“The initiative is a first for Africa, and will be a real benefit to researchers on the continent,” says Sakkie Janse van Rensburg, UCT’s executive director of Information Communication Technology Services (ICTS).

## SOUTH AFRICAN DATA-INTENSIVE RESEARCH CLOUD (SADIRC)

Given the success of the ARC prototype, the next step is the expansion of the ARC to include a greater number of research institutes, including both universities and organisations such as SKA South Africa and the South African National Space Agency (SANSA). A memorandum of understanding was in development at time of writing, which will formally constitute SADIRC.

In time, it is hoped, SADIRC will expand to offer access to storage for massive data sets, as well as the tools and software to properly collaborate on, analyse and visualise the data – to all South African researchers, including those based at our most under-resourced institutions.

## THE ILIFU PARTNERS



## ILIFU

The establishment of the ARC meant that UCT was perfectly placed to lead a consortium of institutions in the Western Cape province of South Africa to put in a bid to the National Integrated Cyberinfrastructure System (NICIS), supported by the Department of Science and Technology. The goal of this bid was to build a data-intensive research facility in the Western Cape that would cater explicitly to the needs of researchers working in astronomy and bioinformatics. The bid was successful; and today, this project is known as 'Ilifu' ('cloud', in isiXhosa).

Ilifu will receive funding from the Department of Science and Technology for a period of three years. It will bring together the existing infrastructure and expertise of the various partner institutions, and build on that to create a hub for data-intensive research systems, platforms and tools in the Western Cape. A further mandate for Ilifu is the development of a research data management system (see the following chapter of this publication).

Within the three-year funding period, Ilifu is set to continue as a self-sustaining facility.

## THE SUM IS GREATER THAN ITS PARTS

While the investment in infrastructure is segmented, the offering itself is greater than the sum of its parts. Working together, Ilifu, the ARC and (in time) the SADIRC will provide researchers access – through an online portal – to the entire tiered infrastructure system, as one entity (see the National Integrated Cyberinfrastructure System box).

A researcher will thus be able to log in to the online portal, from any device in any location, to access the stored data sets and run the necessary programs to analyse and visualise the data.

## NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM

The NICIS is a national initiative of South Africa's Department of Science and Technology. The strategy defines different tiers of research infrastructure. Tier III provides institutional infrastructure, and tier II is regional – and, as is the case with Ilifu, involves the collaboration of several universities. Tier I refers to national-level infrastructure.



# The dilemma of collaboration



UNIVERSITY OF CAPE TOWN  
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

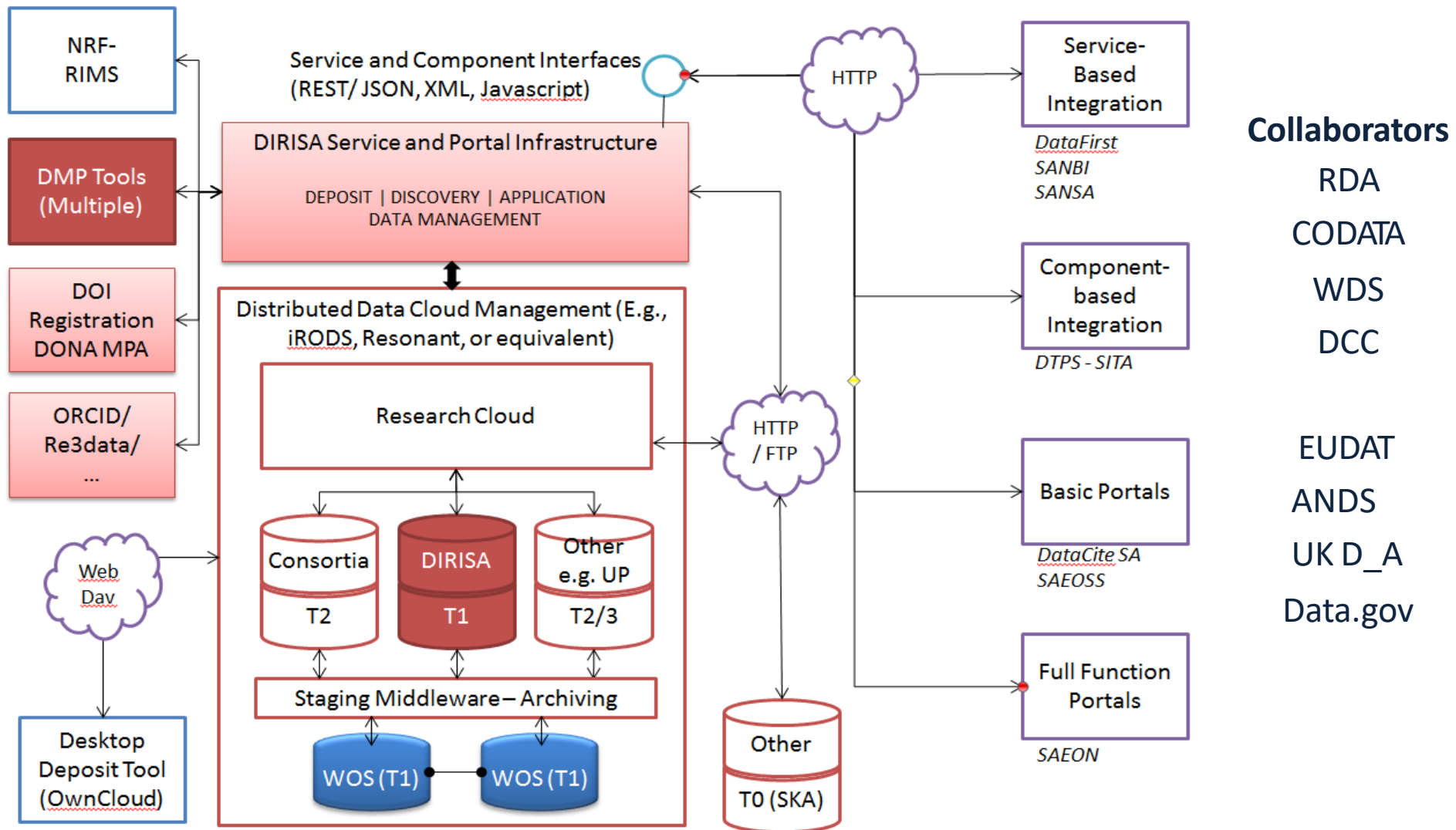






# DIRISA

Data Intensive Research Initiative of South Africa





## National Non-Traditional Research Output Portal



AUTOPOPULATED

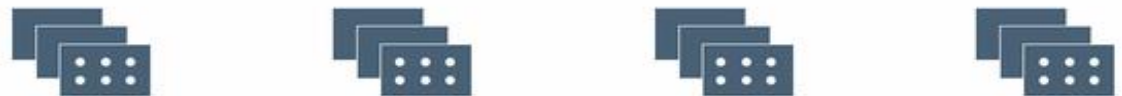
## Region or subject specific 'Data' Portals



*e.g. genomics.funderx.com*

AUTOPOPULATED

## Portals for each grant code



normalisation and mapping to Funded Grant Code



and Institution using GRID, UberResearch and FundRef

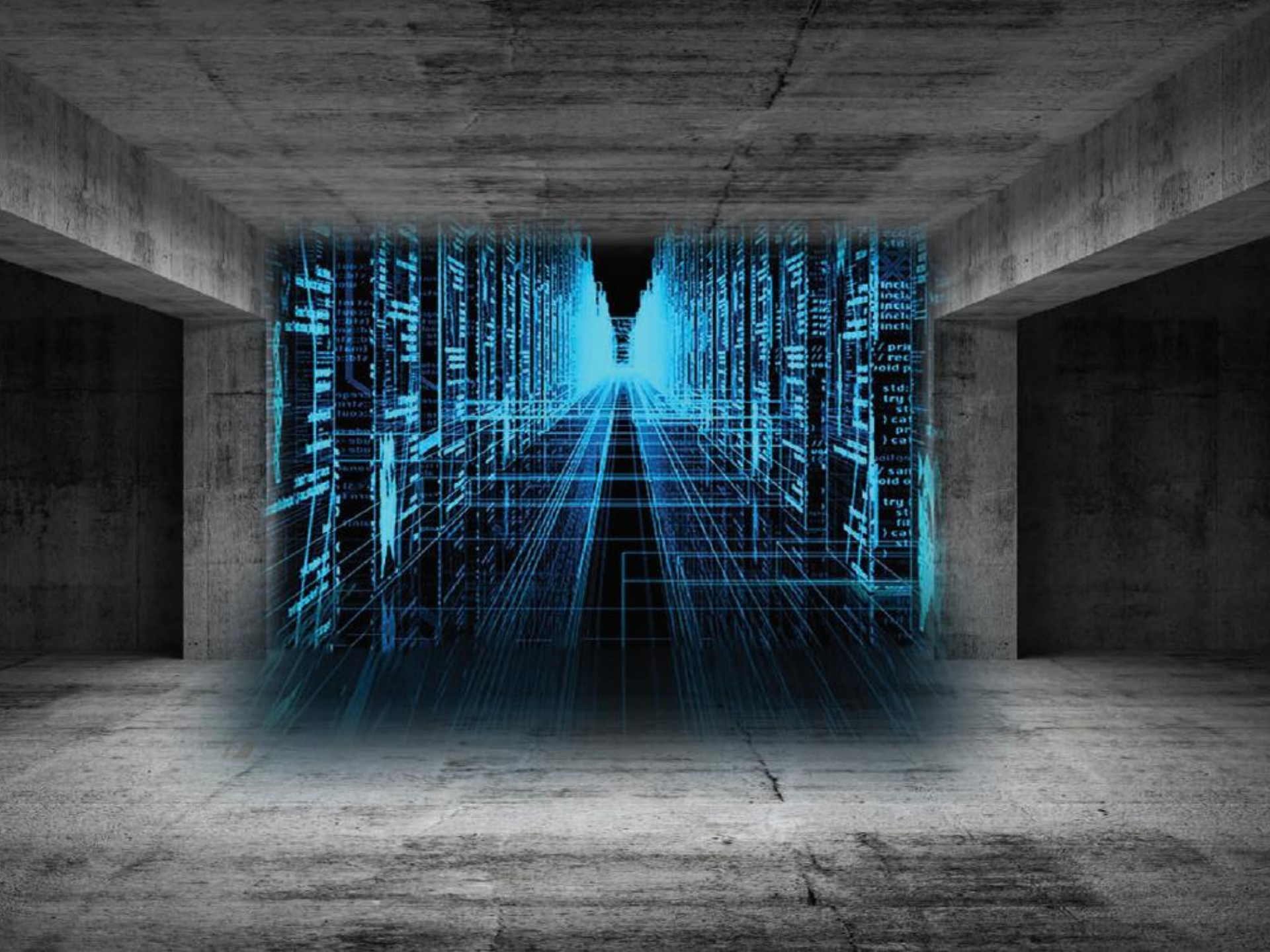
Funded Researcher Uploads

Academic Publisher Datasets  
from e.g. SpringerNature,  
ACS, Wiley, etc.

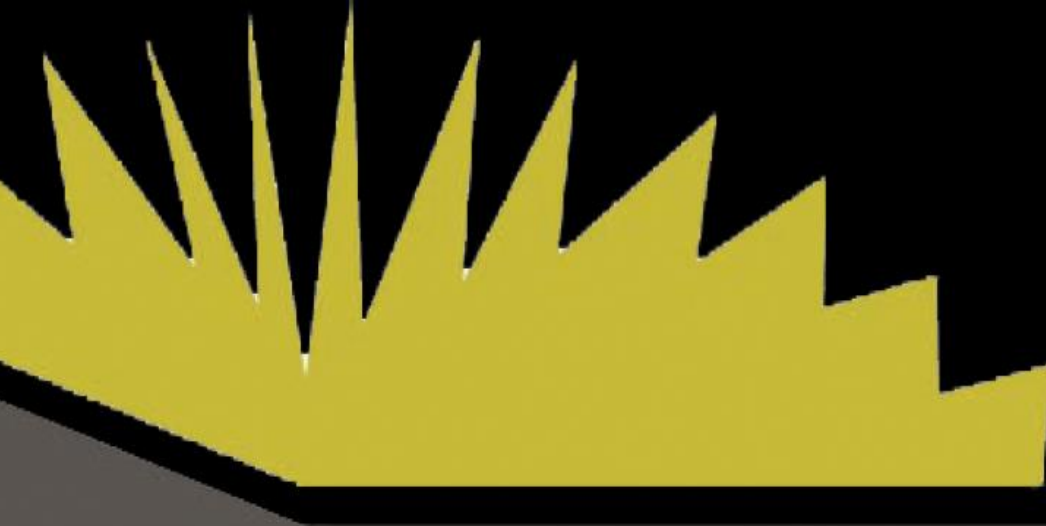
Migration or ingest of  
data from other  
**Funded Repositories**, or  
**Institutional Data Repositories**

Migration or ingest of data  
from **popular repositories**  
e.g. Figshare, Dryad, Zenodo









Cloud technology has the capacity to democratise our research infrastructure needs, empowering individual researchers, giving them real control over their data, but also allowing distributed organisations to work together as one



Meeting the needs of  
**big data  
science**

Further information: <http://www.eresearch.uct.ac.za/>