NICIS – DIRISA Data Intensive Research Support

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

science & technology
Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA

CSIR
our future through science

April 2020

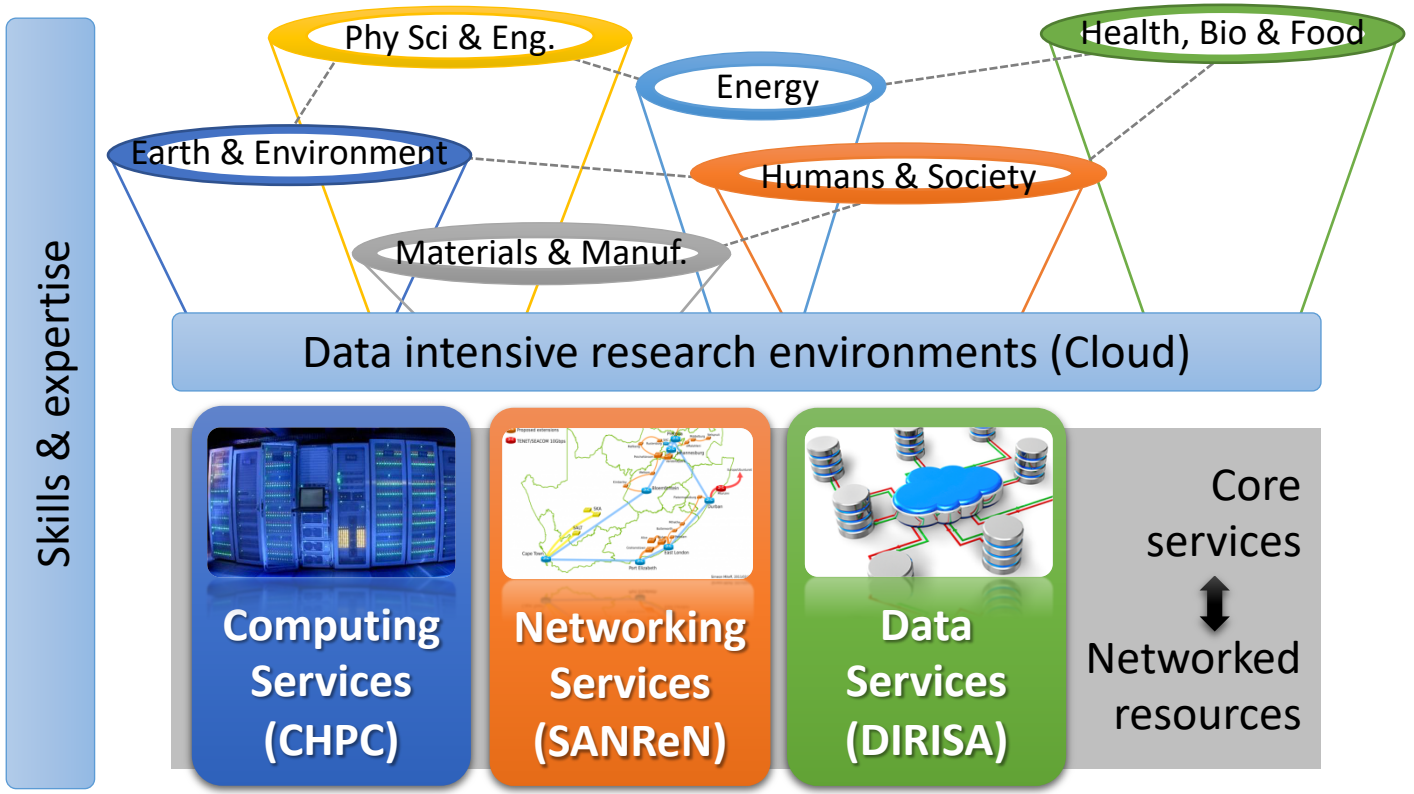
Transboundary Water Collaboration

NICIS Strategic Objectives

- Sustain a world class and relevant national integrated cyberinfrastructure system for Science and Technology
- Enable and promote eScience in South Africa
- Position South Africa to take part in, host and lead large scale global research and science projects (e.g. SKA, CERN experiments, etc.).
- Provide thought leadership to South Africa's evolving cyberinfrastructure strategy and activities, and facilitate uptake of cyberinfrastructure.
- Foster the development of human capacity in cyberinfrastructure and its application, and contribute to the transformation of this sector.

The National Integrated Cyberinfrastructure System (NICIS)

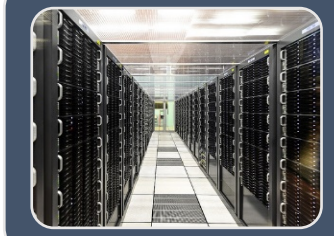
- Federated physically distributed cyber platform for e-research
- Overarching coordination supporting national strategy
  - **National (Tier1)**
  - Regional (Tier2)
  - Institutional (Tier3)
- Priority & cross-cutting domains



DIRISA Objectives

Advocate and coordinate

- Coordinate initiatives
- Stakeholder engagement



Build data infrastructure

- Build and maintain Tier 1 nodes and services
- Start T2 domain nodes

nicis

NATIONAL INTEGRATED CYBER INFRASTRUCTURE SYSTEM

DIRISA

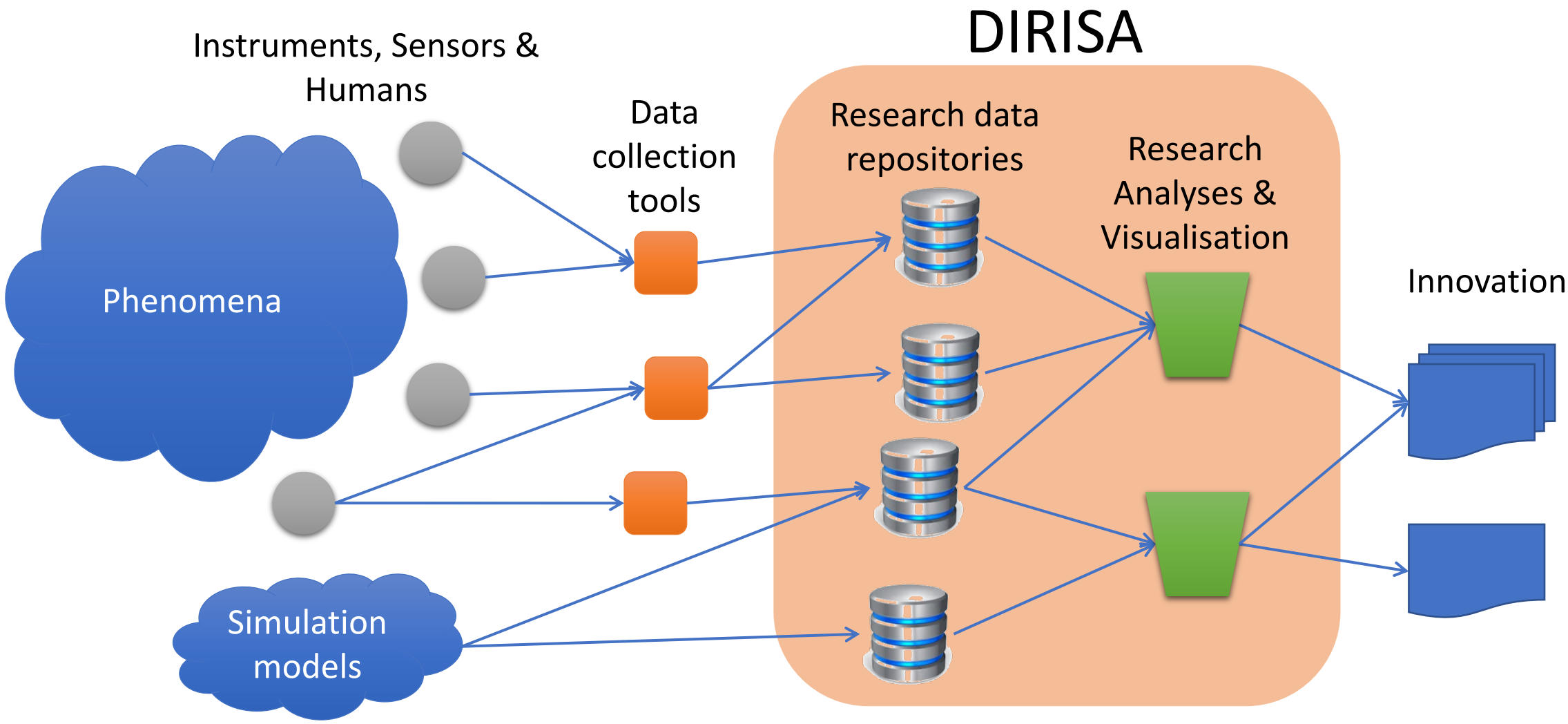
Manage research data

- Deploy RDM services
- Develop user policies

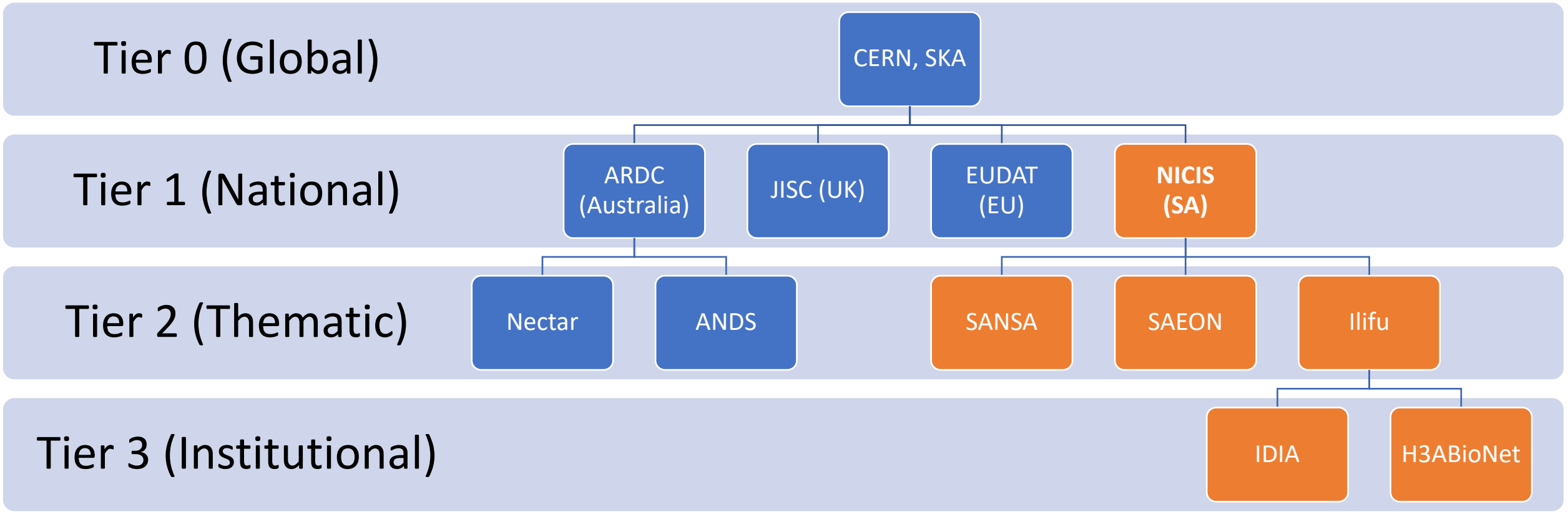
Develop capacity and skills

- Support e-Research programmes
- Coordinate training workshops

Services for the Research Data Value Chain

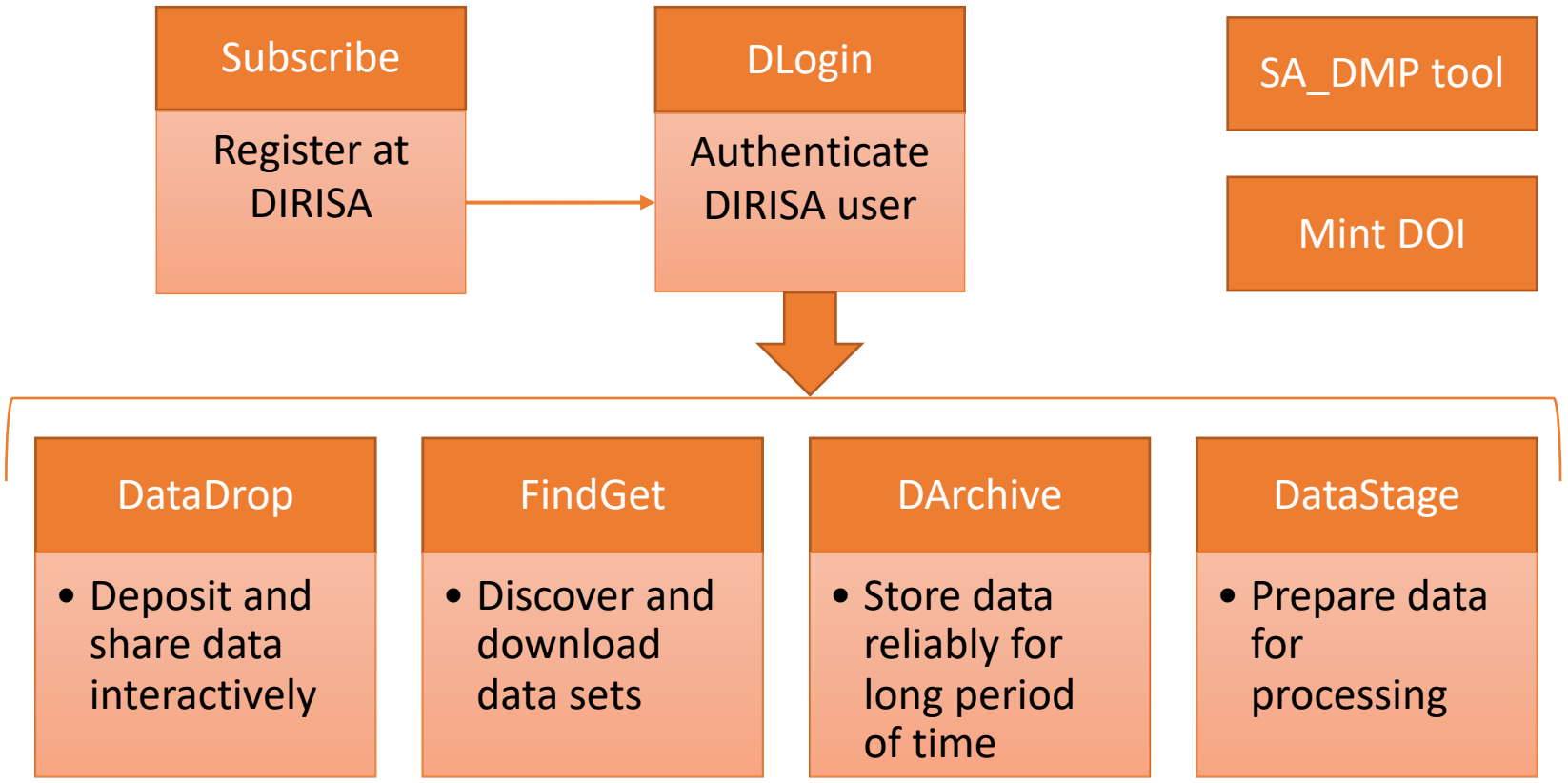


South African National Research Data Commons



South African National Research Data Commons

Phase 1: Research Data Management



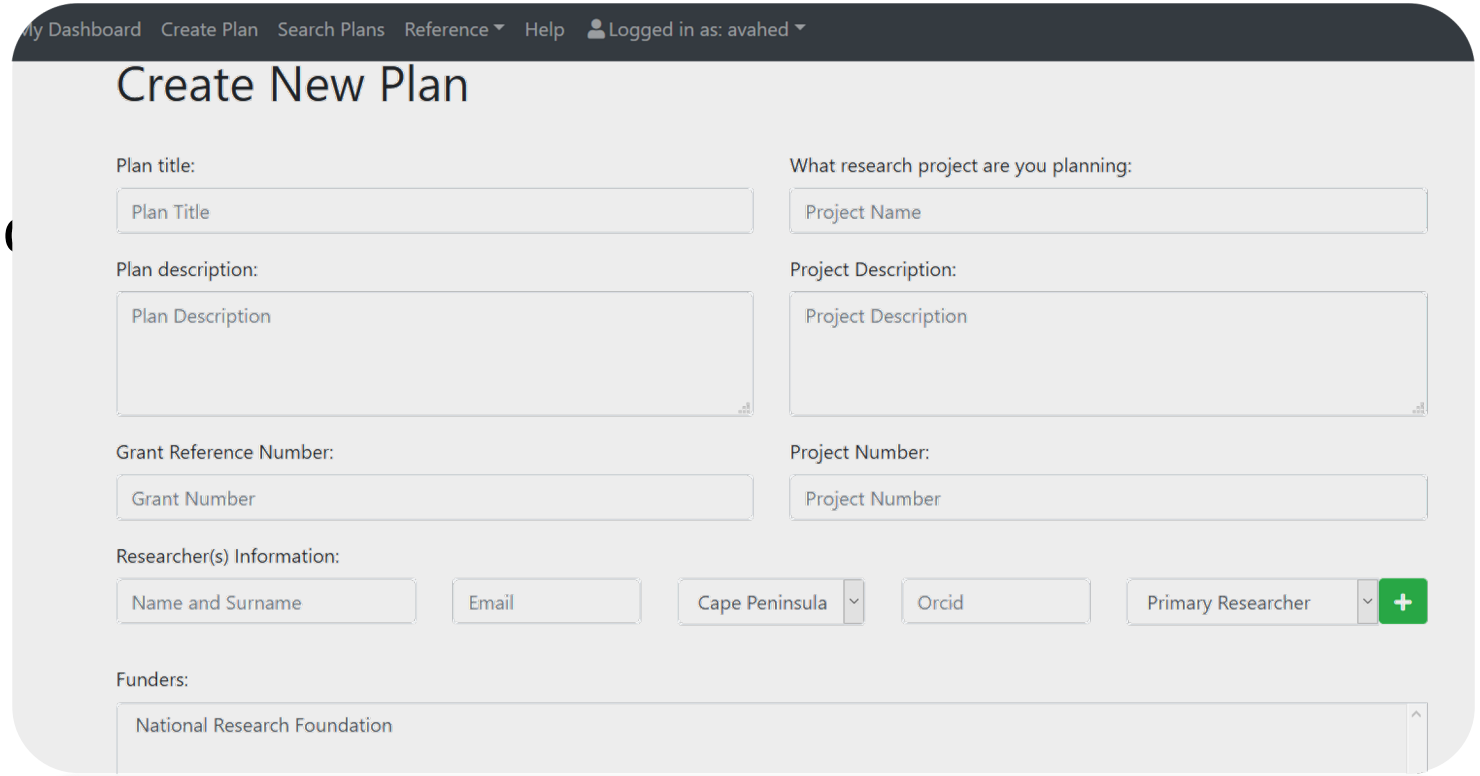
Phase 2: Collaborative Research Environments

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

Data Management Planning: SA_DMP Tool

Create data management plans: https://secure.dirisa.ac.za/SADMPTool/

- Funder requirements
- Data quality and preservati
- Visibility and discovery
- Asset management
- Publication provenance
- Attribution: citable data

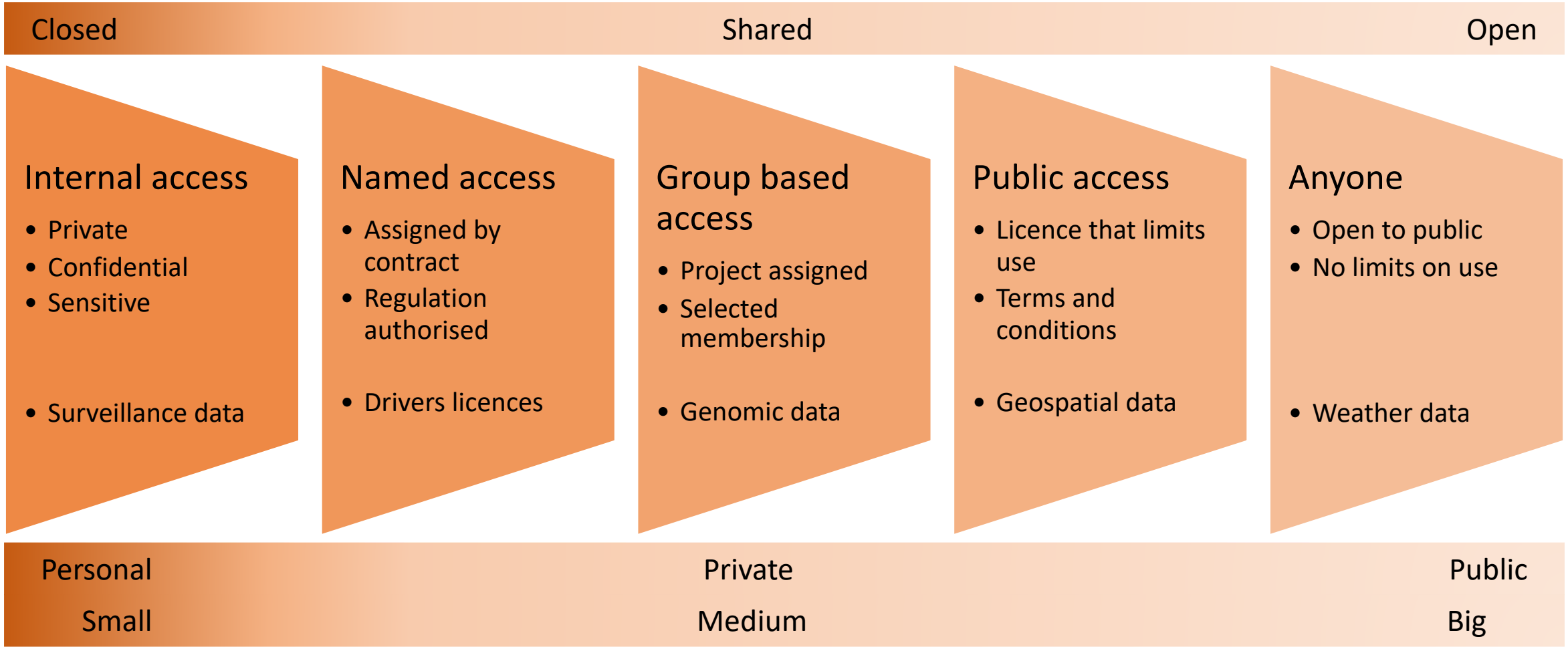


Models of service

- Data storage as a service
  - Registered user has 20GB (default) cloud storage for online interactive access
  - User determines data sharing model
- Infrastructure as a service
  - User has access to a virtual research environment (VM, Docker, Jupyter notebook…)
  - User arrangements
- Data archiving as a service
  - Research data with metadata uploaded to DIRISA long-term (5 – 10 yrs) archival storage facility
  - CHPC archived data outputs

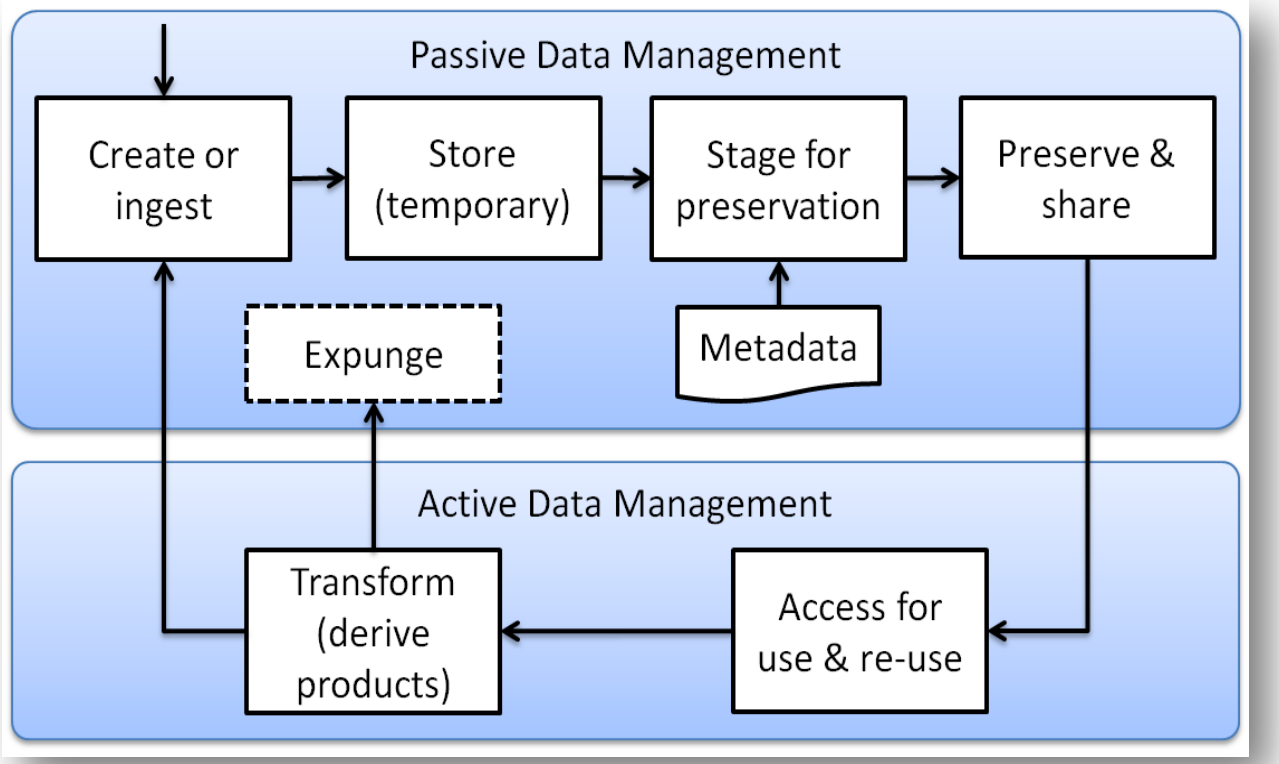
Data Access Model: Open by Default

“As open as possible; as closed as necessary”

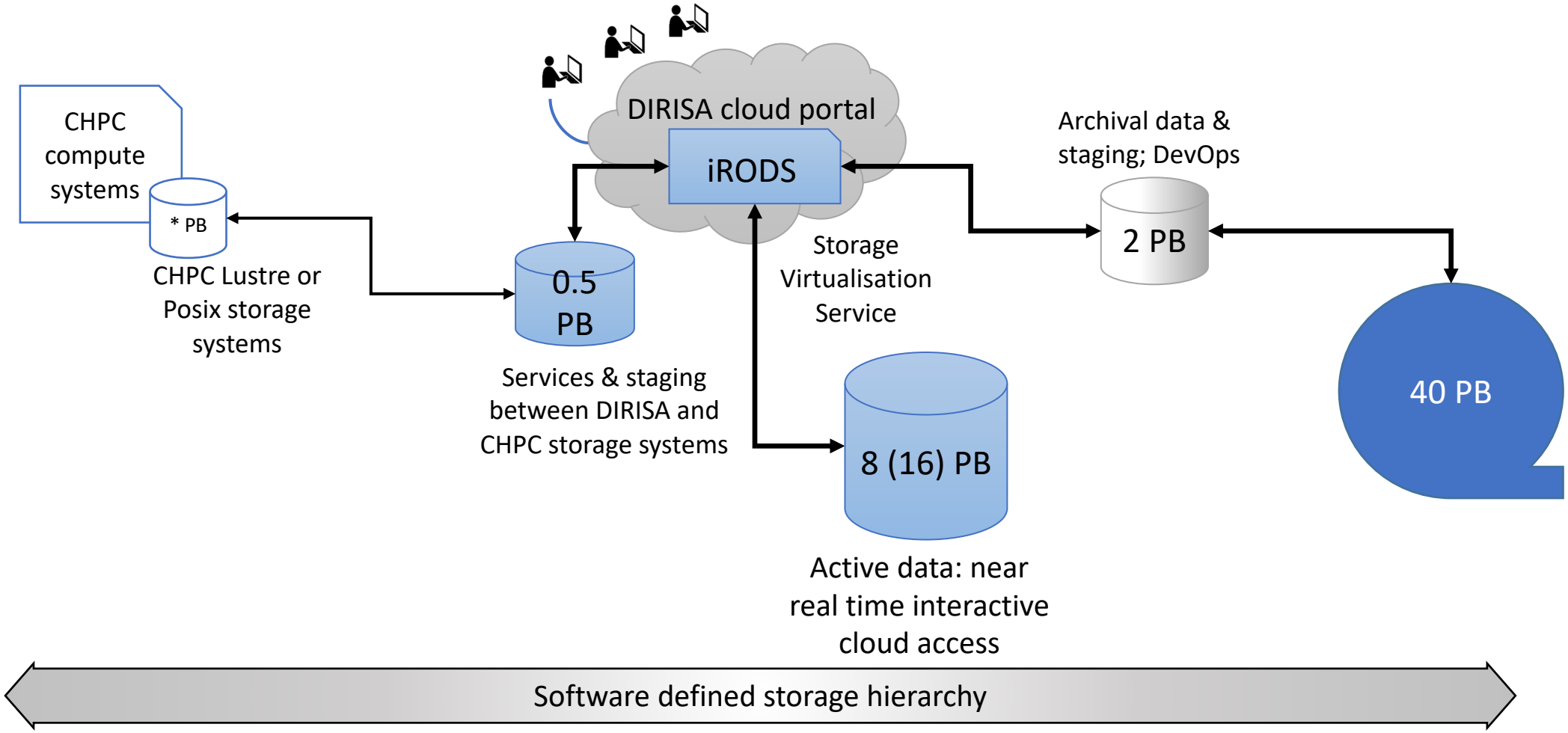


Activities

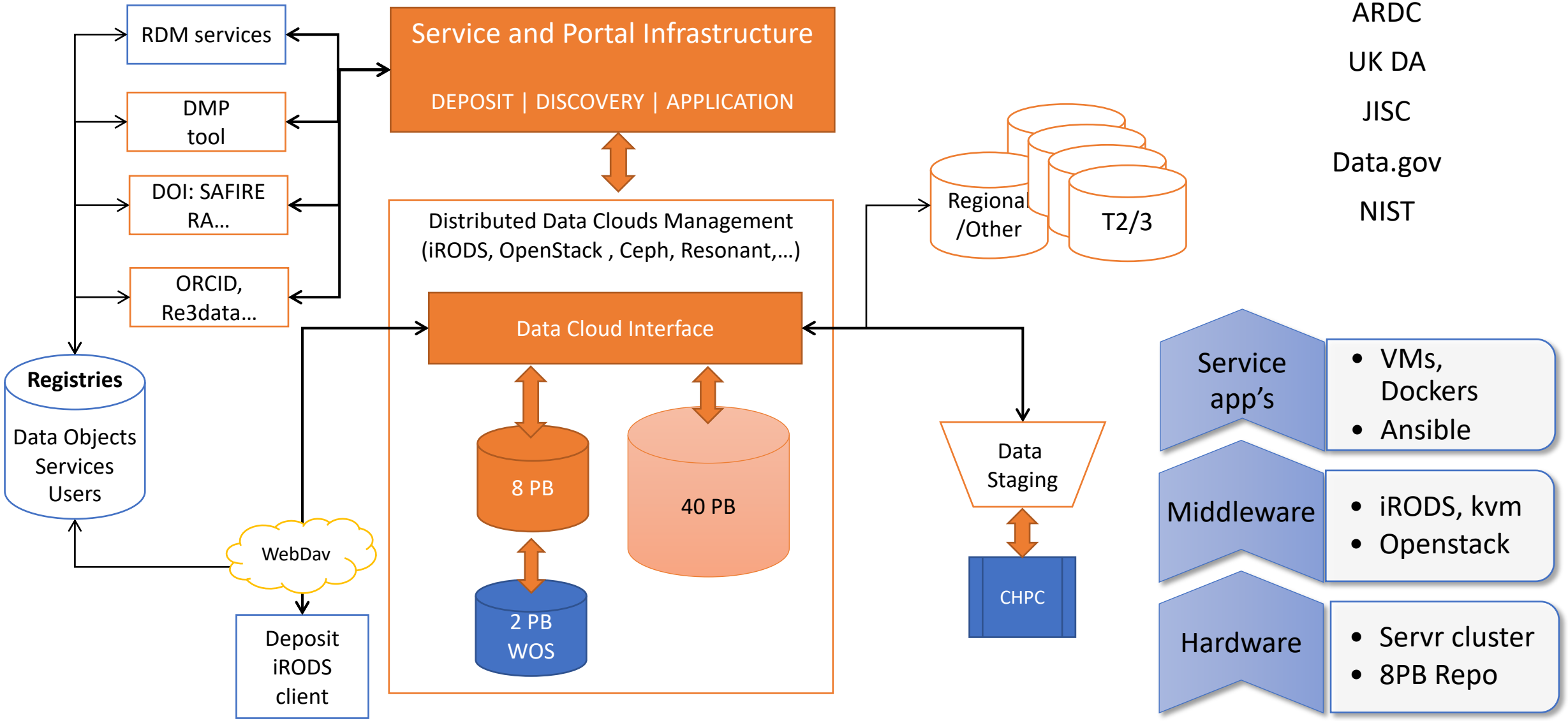
- Data Infrastructure and services (federated data commons)
  - Online (“Hot”) 8 PB & Archival 40 PB (“Cold”) repositories
  - Cloud-based access, upload and discovery services
- Research Data Management
  - Policies: Security & Regulation (POPIA)
- Capacity & expertise
  - Masters in eScience
  - Coordinating training
- Advocacy & outreach
  - Local: DSI, DTPS, USAf, ASSAF, NRF
  - Global: RDA, DCC, CODATA, WDS, SKA
- Coordination & strategy
  - National Big Data strategy
  - AOSP; SADC Cyberinfrastructure



Tier 1 Conceptual Architecture



High Level Architecture



Exemplars

EUDAT

ARDC

UK DA

JISC

Data.gov

NIST

Conclusion: what is needed?

- Data repository
  - Volume/s and growth (Data sets)?
  - Access (Security)?
  - Latency (Bandwidth)?
  - Storage period?
- Analytical environment/s
  - Analytical tools?
  - Services hosting (Web, REST, Databases)?
  - Access (Group/s)?
  - Compute requirements?
  - In-memory (RAM) requirements?

DIRISA does not have domain expertise

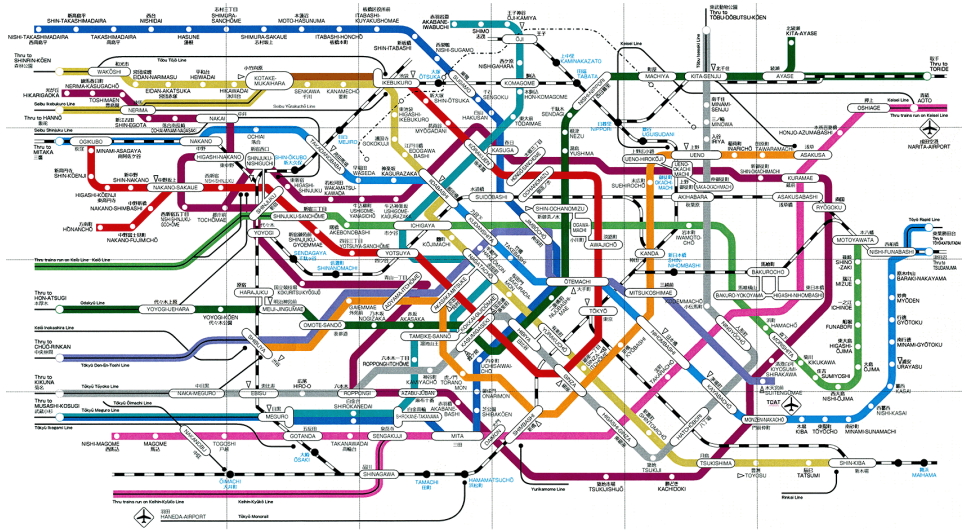
Thank you

Anwar Vahed

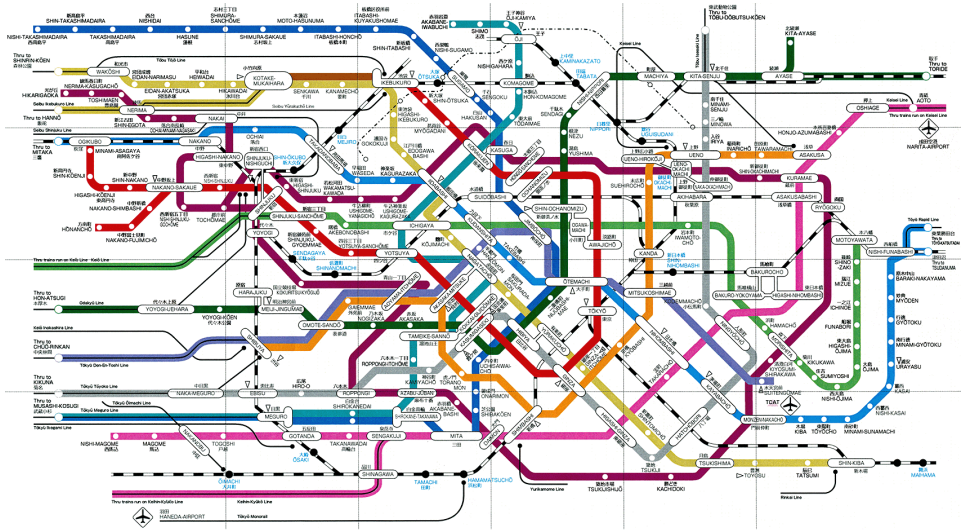
NICIS – DIRISA

avahed@dirisa.ac.za

Where are we now?



Where are we now?



Data Attribution

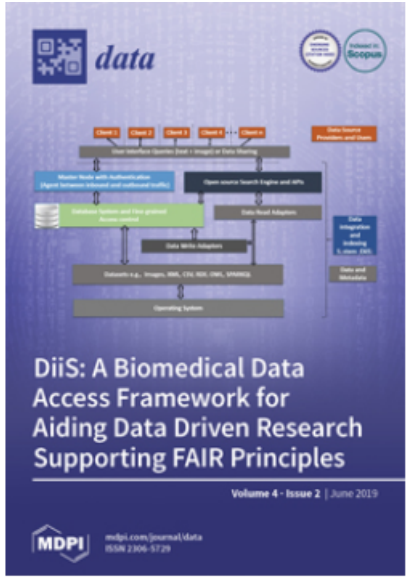
F1000Research
AN OPEN SCIENCE JOURNAL FOR LIFE SCIENTISTS

Journal of
open psychology data

open health data

Journal of
open archaeology data

A peer-reviewed open-access journal
Biodiversity
Data Journal
Making your data count! ISSN 1314-2828 (online)



Data in Brief

Editors-in-Chief: Hao-Ran

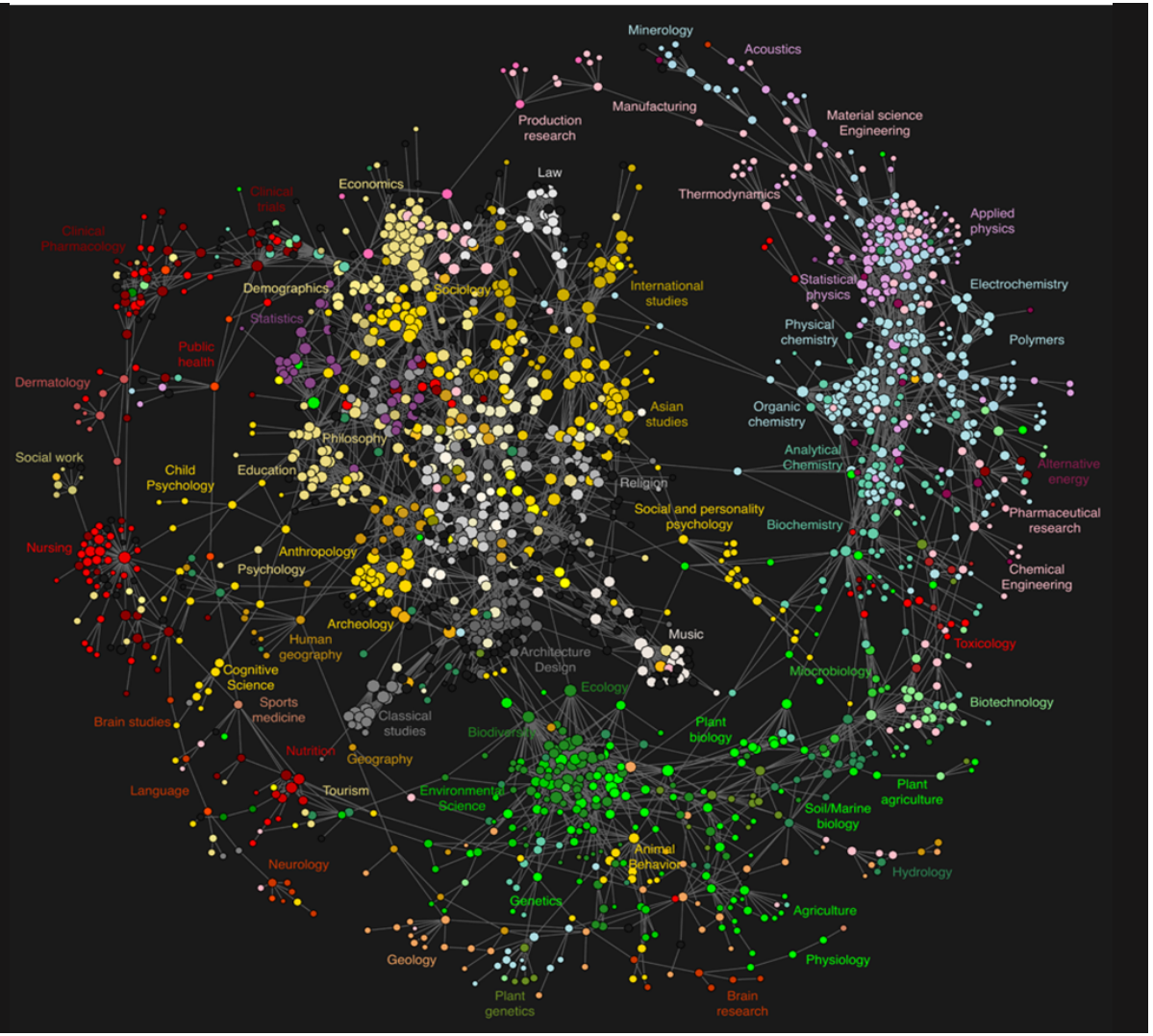
> View Editorial Board

ISSN: 2352-3409

health dat

Data Connects Disciplines

- Physical & Chemical Sciences
- Biological Sciences
- Medicine & Health
- Engineering & Manufacturing
- Environmental & Earth Sciences
- Social Sciences & Humanities
- Languages
- Education
- Business & Economics
- Law
- Social media

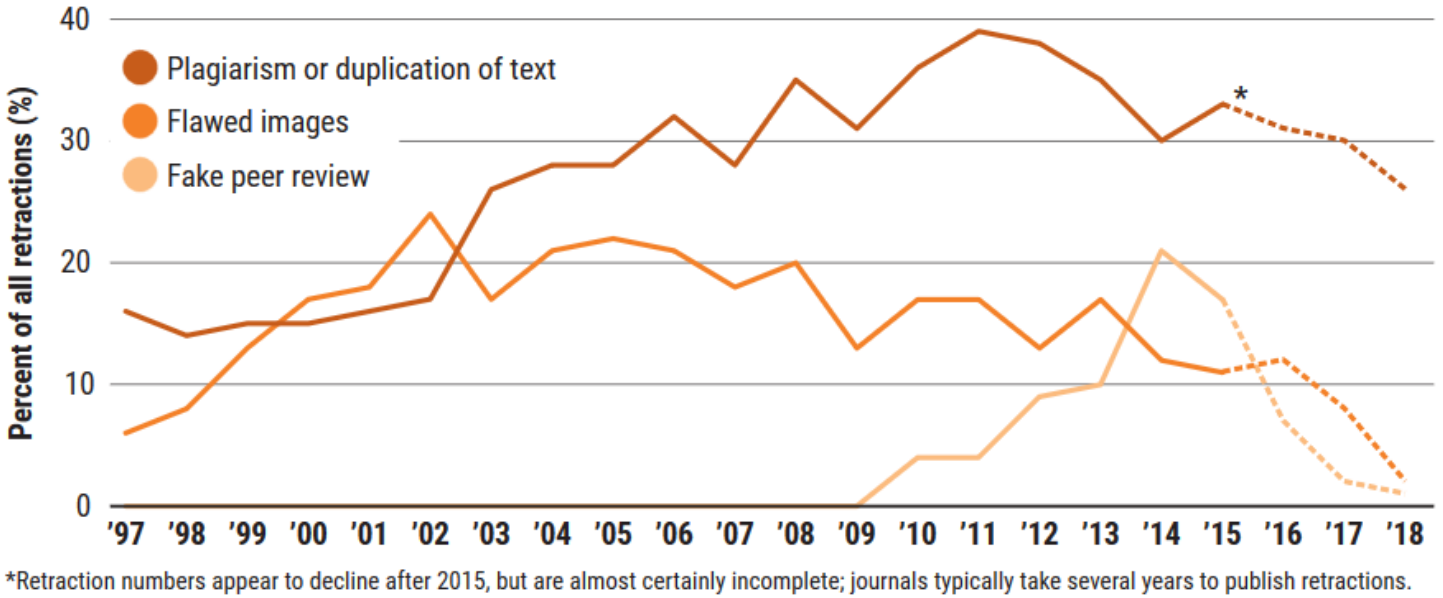


Publication Provenance

Retraction Watch

Tracking retractions as a window into the scientific process

Kyoto University suspends first author of retracted Kumamoto quake paper



*Retraction numbers appear to decline after 2015, but are almost certainly incomplete; journals typically take several years to publish retractions.

nature
International journal of science

Retractions

Retraction | 11 April 2019
Retraction Note: KAT5 tyrosine phosphorylation couples chromatin

Retraction | 20 February 2019
Retraction Note: A homing system targets therapeutic T cells to brain

Improving Return on Data

Skills and expertise

e-Research Environments

Data Management Services

National Data Infrastructure