NATIONAL INTEGRATED
CYBER INFRASTRUCTURE SYSTEM

Big Data Strategy for Research, Development and Innovation

NITHeCS

21 August 2023

science & innovation
Department:
Science and Innovation
REPUBLIC OF SOUTH AFRICA

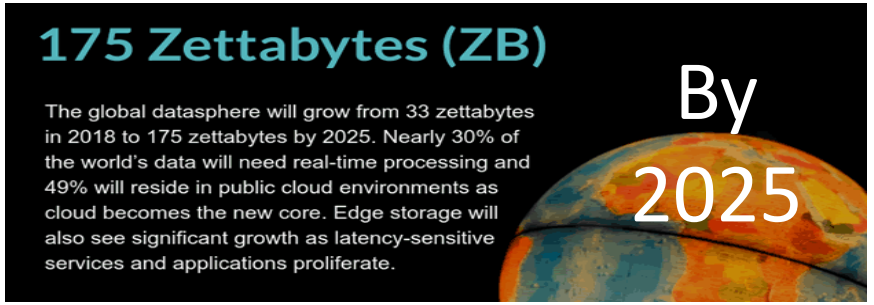
CSIR
Touching lives through innovation

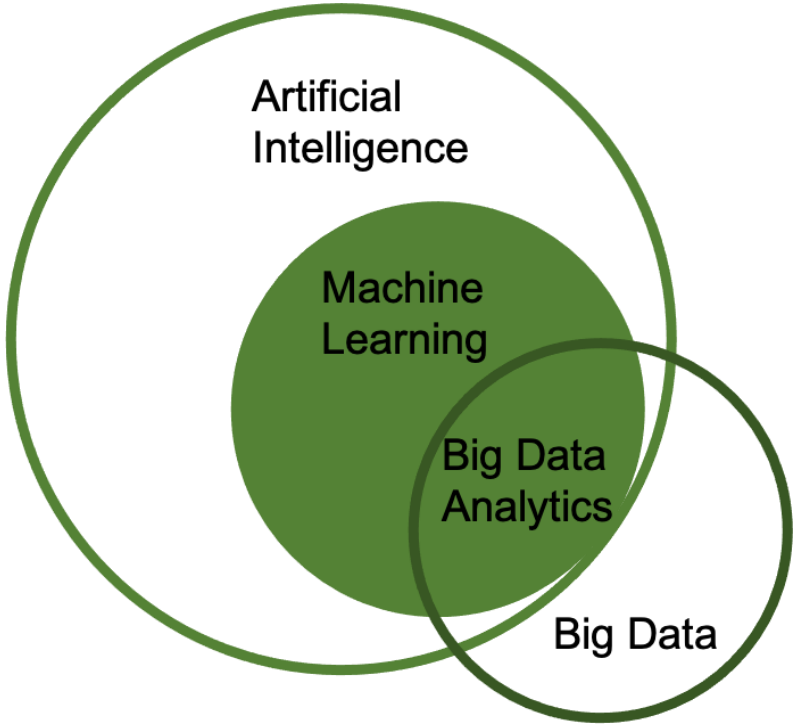
Outline of Strategy

1. National landscape
2. Process
3. Purpose and Scope
4. Vision and Principles
5. Objectives

Implementation

1. Framework of actions [18]
2. Indicators of success [16]
3. So far…





NICIS
NATIONAL INTEGRATED CYBER INFRASTRUCTURE SYSTEM
DIRISA

Process

Draft document

- Draft with contributions from international experts
- NIST, EUDAT, ARDC, JISC,…

Input from stakeholders

- Received input from stakeholders and DST
- CSIR, NICIS, SKA, MIIA,… (Annexure B)

National Research Big Data Strategy workshop

- Workshops. Number of delegates: 98 delegates
- University DVCs; Research Councils; Pvt Sector, Researchers

Publish strategy for Key Stakeholders input

- Publish strategy for key stakeholders' comment
- Regular reviews to maintain relevance with changing landscape

A synthesis of

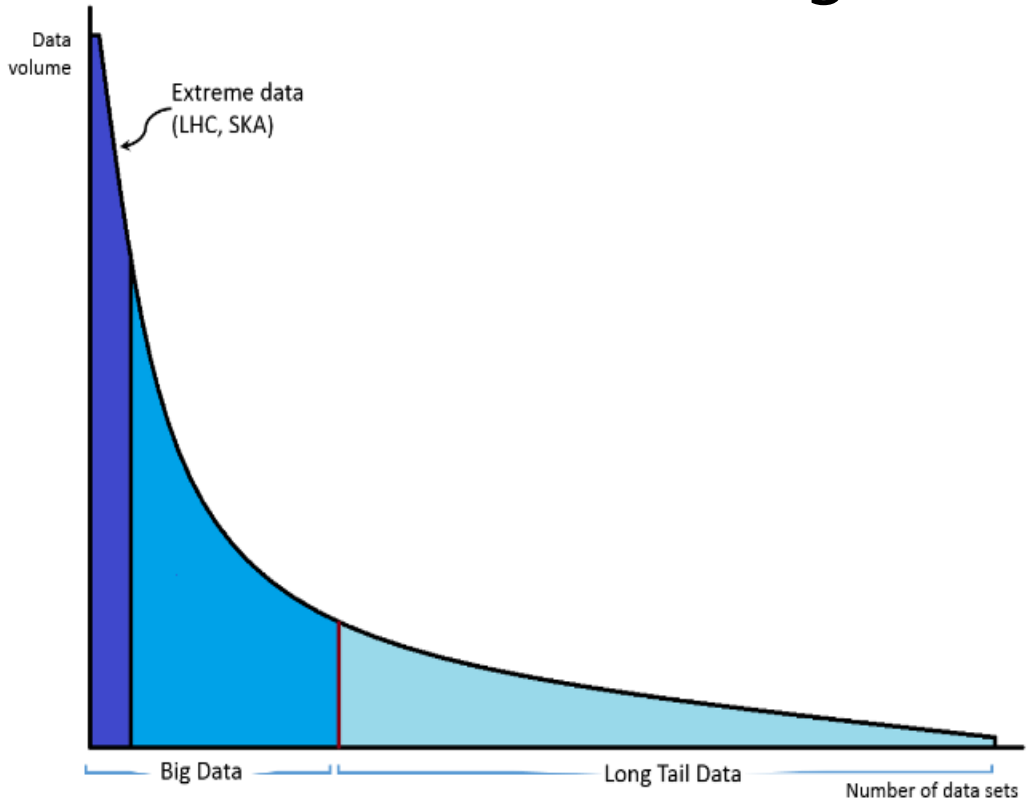
- Expert opinion
- Inputs from key stakeholders
- Strategies of other countries

Guided by

- National priorities
- Local landscape
- Opportunities

Big Data definition

"Data that have acquisition, storage, transmission, processing or visualisation requirements that exceed the capabilities of conventional technologies"



Attributes

Variety
- Multiple different formats

Velocity
- Transmitted or generated at high speeds

Value
- Costly or impossible to recreate. Historical

Volume
- Huge amounts of generated or observed

Veracity
- Accuracy and integrity

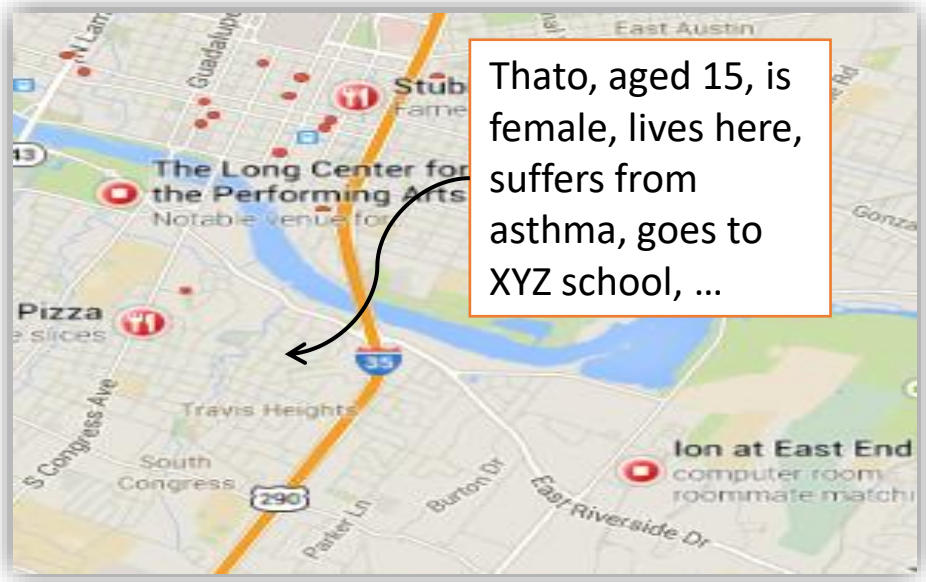
Big Data

Challenges

- Scale; Heterogeneity; Complexity

Data acquisition or generation	Data preparation and pre-processing	Data analysis and mining	Data visualisation and curation	Value derivation and impact
• Structured/unstructured • Sensors and events • Social media • Real-time data streaming	• Validation • Security and privacy • Metadata annotation • Data models	• Machine learning • Analytics • Data fusion • Simulation and modelling	• Aggregation • Publication • Data linking • Data preservation	• New knowledge and actionable insights • New or improved services • Better decisions and planning

- Human capital
  - Management and Science
- Cyberinfrastructure
  - Compute; Storage; Transmission
- Ethics and Privacy
  - Security; Policy; Regulation



Opportunities

Computer Science (AI-ML, IoT,…); Astronomy; Biosciences; Environmental Sciences; Particle Physics; Health; Earth Observation

Social Sciences and Humanities; Education; Law

- Indigenous Knowledge Systems
- Manufacturing & Industry
- Healthcare
- Transport
- Water and Energy
- Finance and commerce
- Retail
- Disaster Management
- Food security
- e-Government

Purpose and scope

- To realise economic, social, scientific and industrial benefit of Big Data for South Africa.
- Focus is on ***research*** big data. But necessary to engage private sector, government departments as well as continental and global big data stakeholders

In support of national strategic goals:

- ✓ A more competitive industry; a more capable state; improved service delivery; expert and skilled human capital;
- ✓ Informs and guides funders, research community and public sector on planning, policy and investment in Big Data;
- ✓ Provides private sector with a national view of priorities in the Big Data domain, allowing business and industry to align

Vision and Principles

A Big Data ecosystem deriving from large data sets, knowledge that leads to innovation that accelerates socio-economic growth and positions South Africa to be competitive in the 21st century big data economy

Principles

- **Inclusivity**. Engagement with *all* key stakeholders
- **Privacy**. Policies *regulate* the ethical use of big data and preserve privacy
- **Governance**. There are clearly defined roles, responsibilities and oversight
- **FAIR**. Data is "*as Open as possible and as closed as necessary*"
- **CARE**. Collective benefit with Authority to control with Responsible and Ethical use

Objectives

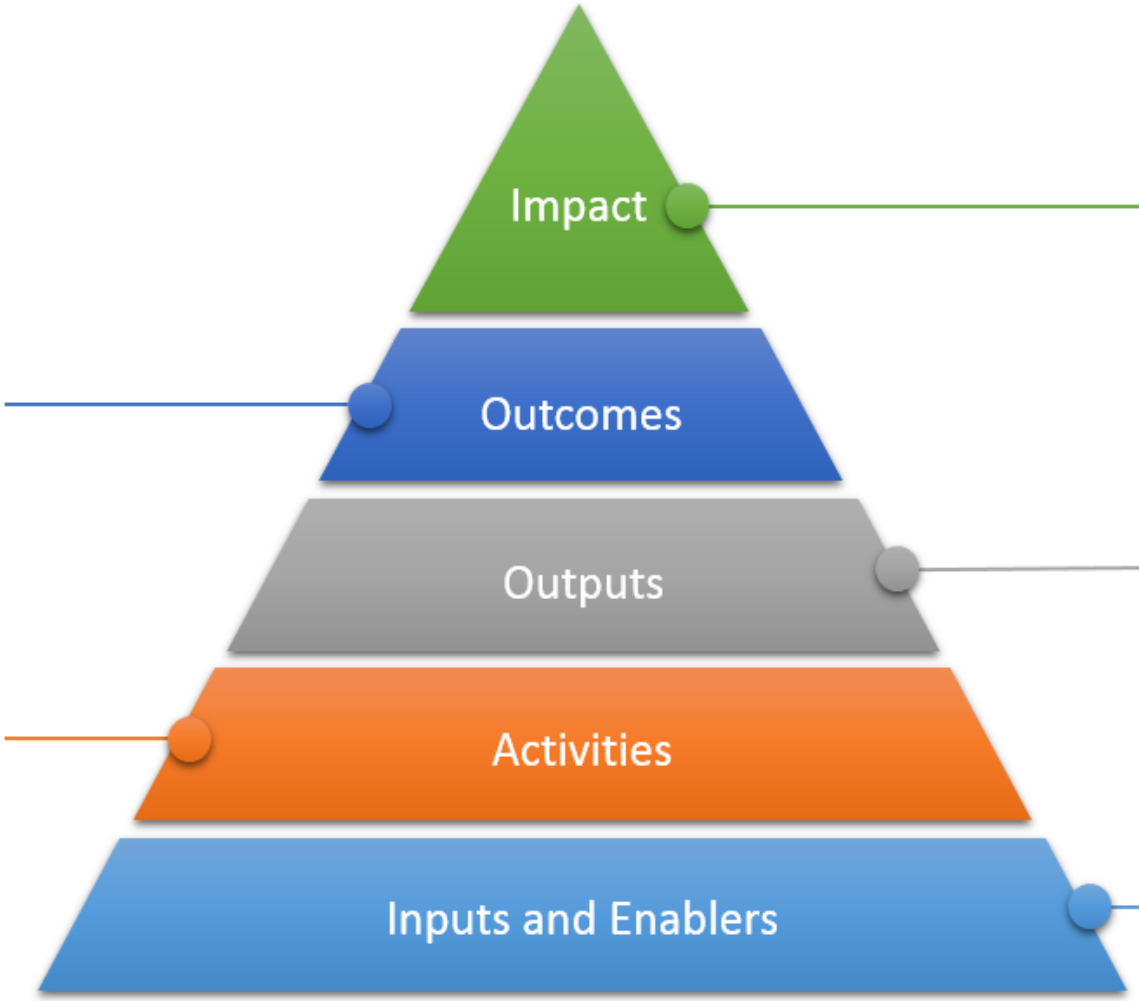
1. **Human Capital Development**. To develop the next generation of R&D expertise and skills to extract knowledge from big data
2. **Cyberinfrastructure**. To provide a sustained and advanced national cyberinfrastructure with allied services
3. **Collaboration.** To foster a thriving and collaborative ecosystem of Big Data RD&I that links government, academia and the private sector
4. **Data Governance**. To develop and adopt standards and policies ensuring compliance with privacy, ethical and legal regulations
5. **Overarching coordination**. To maintain an overarching and coherent national approach to investments

- Framework of Actions [18] categorised under objectives
- Indicators of Success [16]

Value and impact

Sound, Open data governance; Leading edge expertise; Innovative and relevant research; Science strategy achievements; Maximised value from investments in data

Develop infrastructure and services; Implement policies and standards; Develop skills and expertise; Coordinate activities; Advocate and promote data intensive research



Achievement of NSI, NDP goals; World class Grand Science; Innovation supporting knowledge economy; Globally competitive industry and socio-economic development

Policies and frameworks; Advanced technology services and tools; Capacity development programmes; Coordinated multi- disciplinary research and Grand Science

Skilled and expert human Capital; Technology Infrastructure; Funding; Stakeholders

Implementation: Framework of Actions [18]

- **Fast-track human capital development** by establishing an advanced RD&I institute (virtual - like Turing Institute?)
- Incentivise triple-helix opportunities and **identify collaborative "flagships"**
- **Sustained investment** in Big Data cyberinfrastructure to accommodate research across all disciplines
- Develop standards, **policies**, frameworks and **code of conduct** to regulate ethical use of (Big) Data
- NICIS is the implementing agency and provides national cyberinfrastructure
- In line with STI White Paper, establish overarching governance structure (or augment an existing one)

Indicators of Success [16]

- Number and extent of collaborative industrial big data projects
- Number of Masters, Doctoral and postdoc Big Data researchers
- Number and diversity of big data research programmes offered by institutions;
- Contributions to the transformation of the digital economy & 4IR
- Use of big data in public services and for evidence-based decision making and planning
- Recognition in terms of international partnerships and scientists attracted to South Africa

Implementation Plan

- Survey of existing activities
  - HCD
  - R&D
  - Resources (expertise and ICT)

to develop call for potential collaborative R&D project/s (in identified priority areas

- Discussions with HEIs and organisations/initiatives with similar purpose

to achieve level of coherence and advocate

- Upgrade NICIS platform

Thank you

The strategy is at
https://www.csir.co.za/documents/bdpublicationpdf

NICIS-DIRISA:
A National Research Data Infrastructure

DIRISA Overview

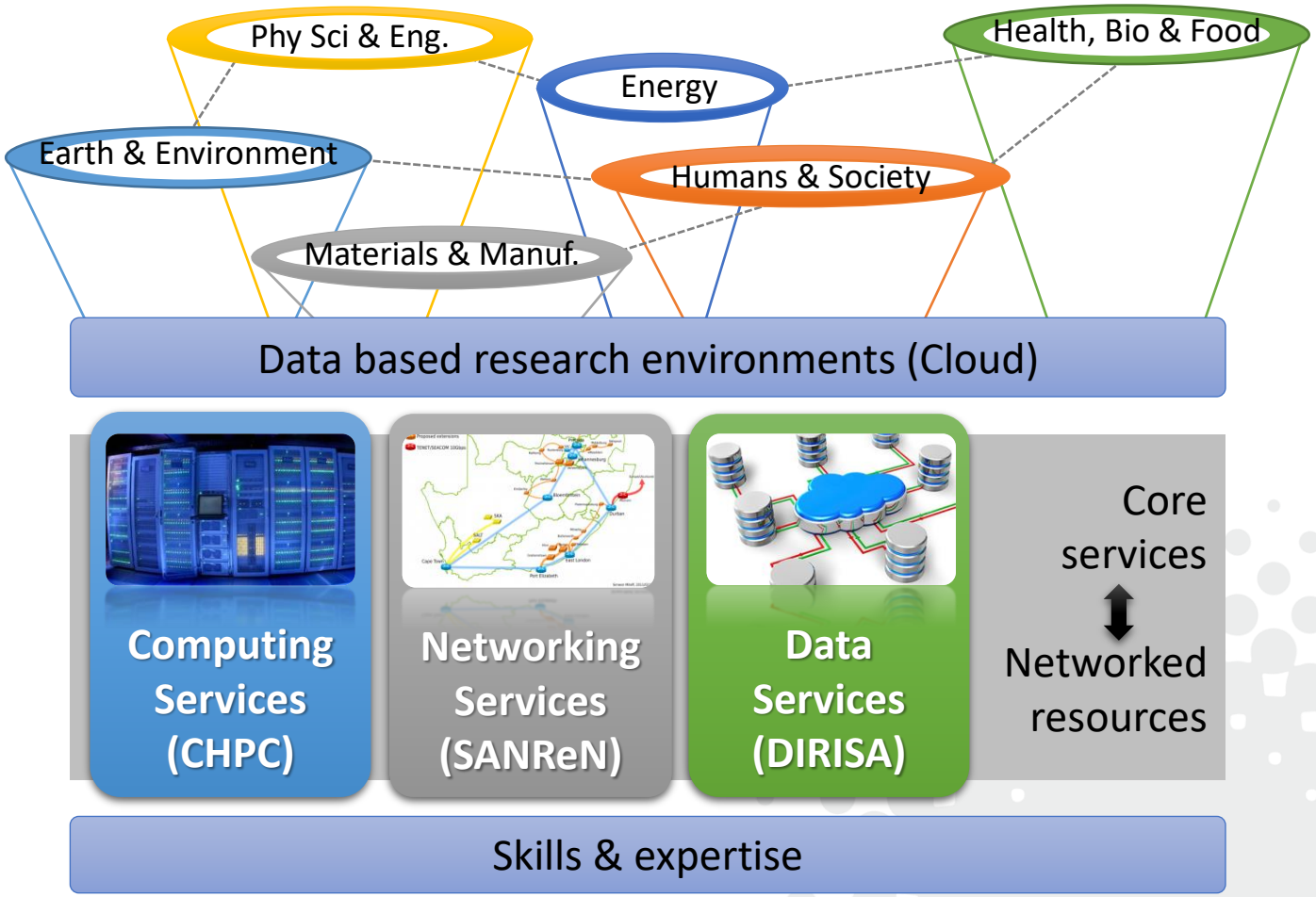
A national initiative of the Department of Science and Innovation and implemented by the CSIR.

About DIRISA & NICIS

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
DIRISA

DIRISA
A national initiative enabling and supporting data driven research

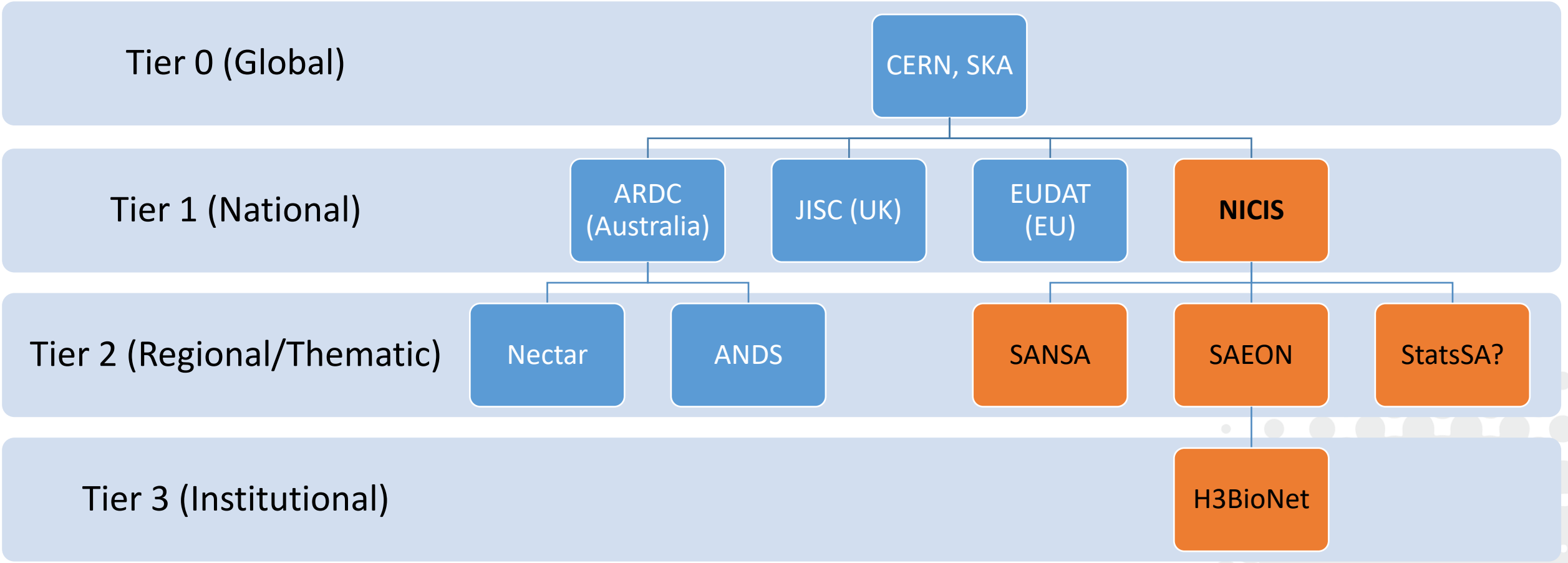
"Researchers deposit, find and access relevant data in the DIRISA Data Commons. They share, reuse and combine data from other domains with their own research"



A national initiative of the Department of Science and Innovation and implemented by the CSIR.
science & innovation
Department: Science and Innovation
REPUBLIC OF SOUTH AFRICA
CSIR
Touching lives through innovation

South African National Data Commons

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
DIRISA



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

science & innovation
Department:
Science and Innovation
REPUBLIC OF SOUTH AFRICA

CSIR
Touching lives through innovation

DIRISA Drivers: "*We must get more value from our investments in [research] data*" [Ex. Min. N Pandor, 2016]

- National Development Plan; STI Decadal Plan; 4IR Implementation Plan; National Data and Cloud Policy; Open Science Framework Report
- Data is key to Research; key to the Digital Economy:
  Technologies: 4IR, AI, Machine Learning, Internet of Things, Block Chain, etc.
- Data represents our heritage

SAEON: Environmental
SANSA: Earth Observation
SAAO: Astronomy
iThemba labs, Meerkat,…
Academia, SOEs & research councils:
Universities, HSRC, SAWS, CSIR, ARC, WRC…
Govt departments &
StatsSA, Energy, Health, Water,…

SKA. € 2 billion to 2022
€ 650 million for Phase 1
SA so far: R2 billion

DIRISA Objectives and Activities

Build research data infrastructure

- National research data repository
- Cloud services for RDM and research

- Online/Active “Hot” storage. 8PB
- Archival/Passive “Cold” repo. 20PB
- Services. DMP tool; DOI minting

Develop skills and expertise

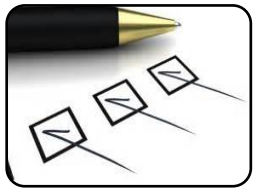
- Postgraduate and e-Research R&D
- Training & workshops

- National eScience Masters
- Data Science Summer School,…

Advocate and coordinate

- Research projects and HCD
- Stakeholder engagement

- Global: RDA, CODATA, AOSP, GOSC
- Local: DSI, NRF, USAf, DCDT,
- Annual Research Data Workshop



Strategic input

- Strategies and frameworks
- Policies and guidelines

- National Big Data strategy
- Open Science Policy Framework

Data Access Model: Open by Default