DIRISA Update
Trusted Digital Repositories

A. Vahed

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

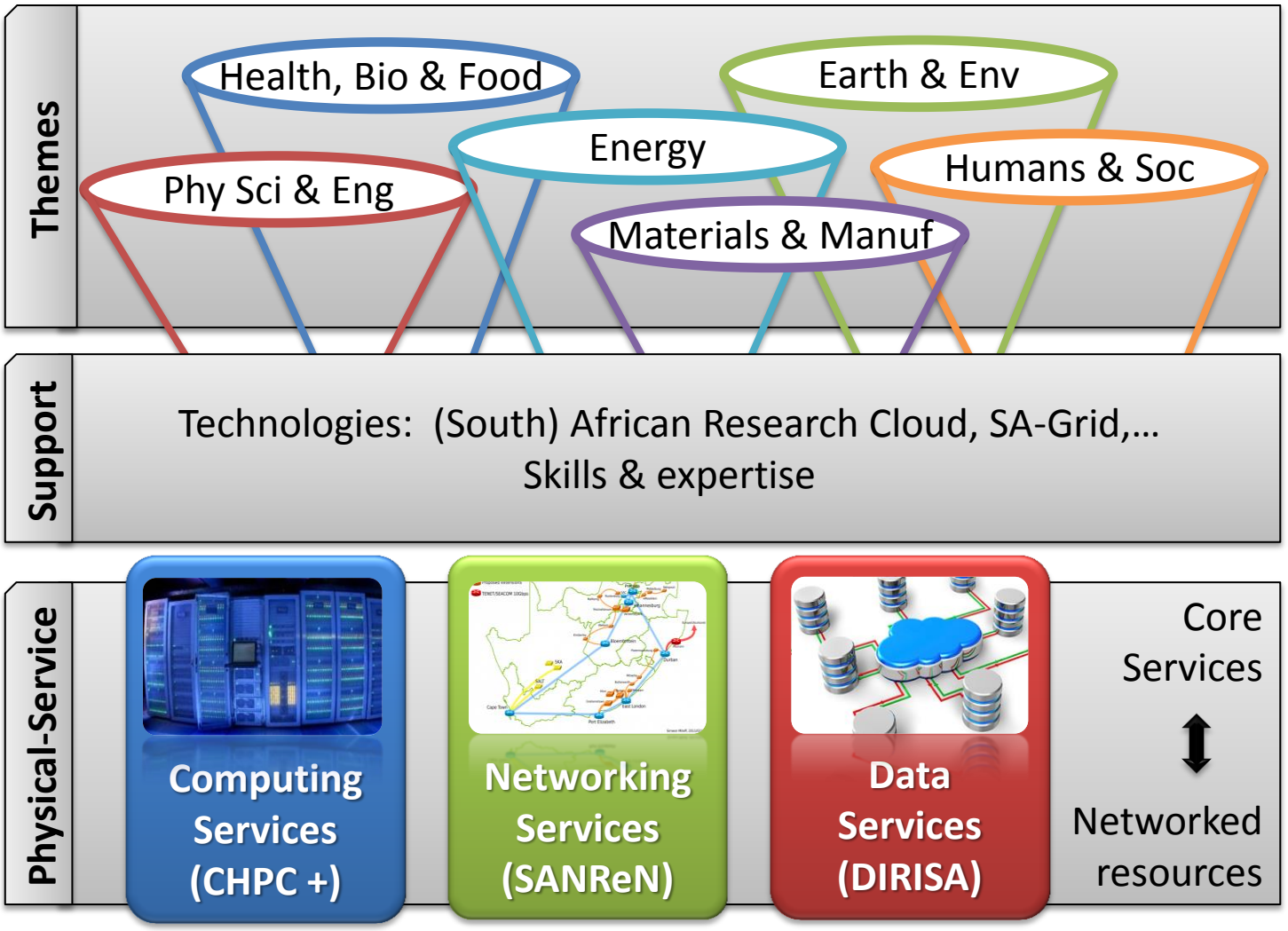
National Integrated Cyberinfrastructure System (NICIS)

Support national science strategy:

Accelerate creation of vibrant, sustained research ecosystems:

Integrate & coordinate e-infrastructure:

South African Research Infrastructure Roadmap (**SARIR**)



DIRISA

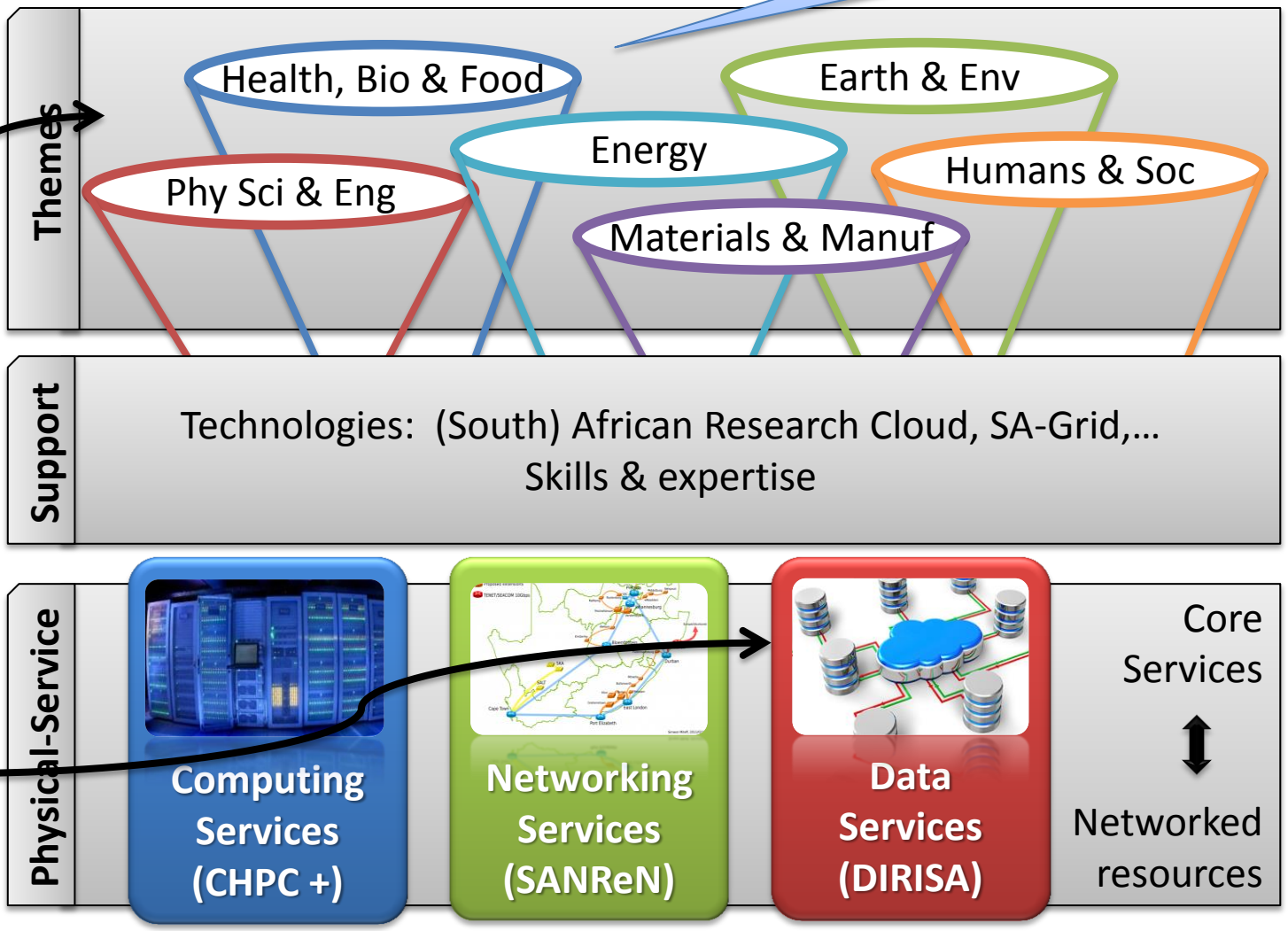
SARIR: South African Research Infrastructure Roadmap

eResearch environments:

"… be the leading organisation … that advocates for and implements data initiatives

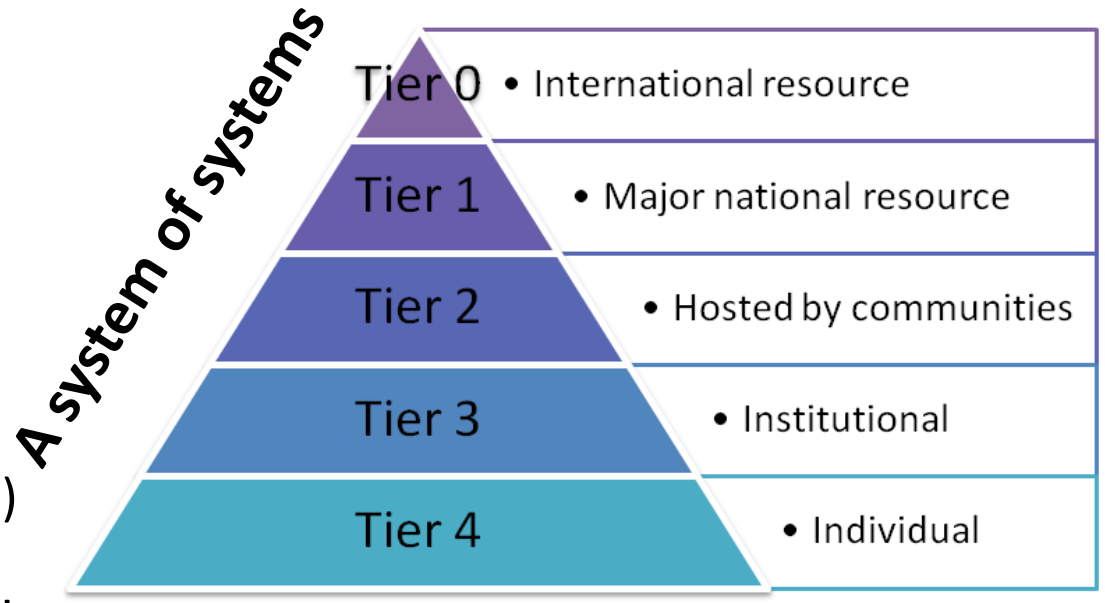
Data services:

"... define, design and provide entire range of data services across the data lifecycle"



DIRISA: so far…

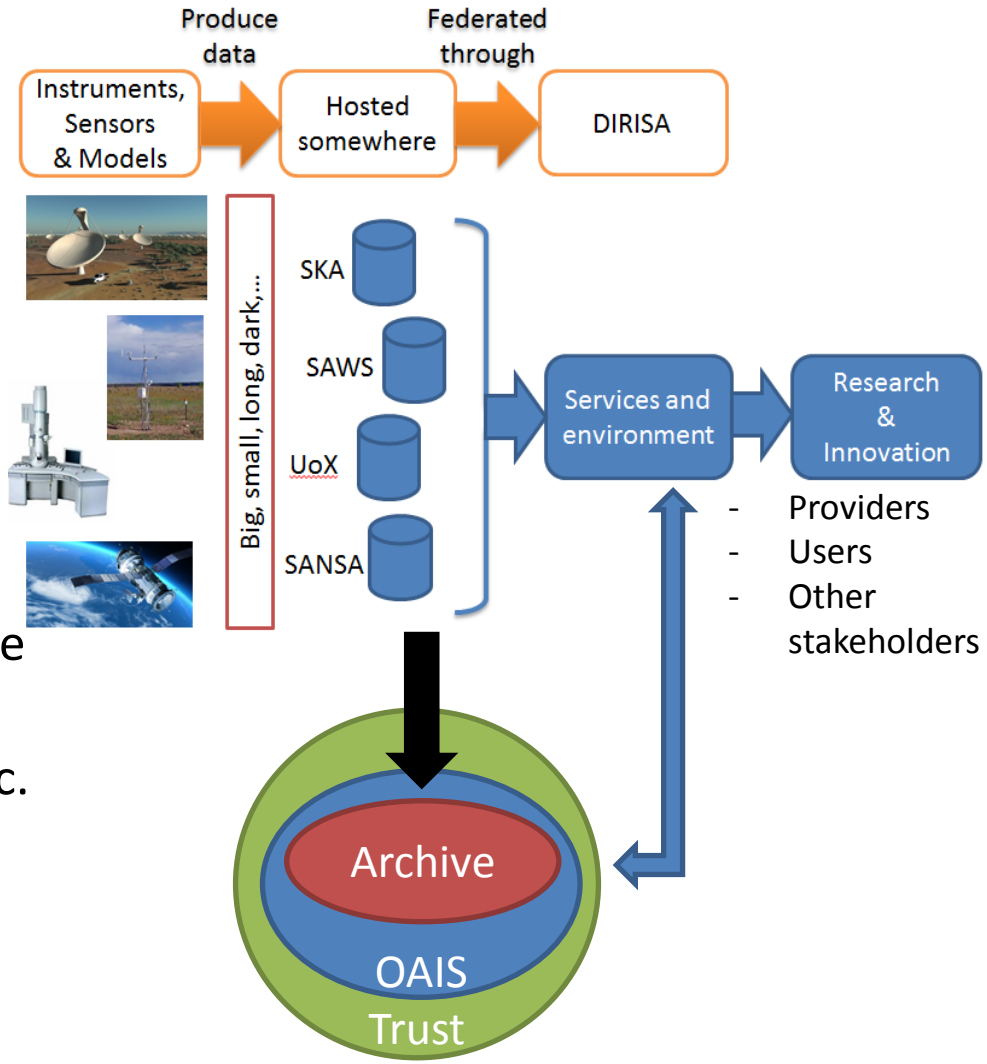
- Federated tiered national data infrastructure
  - Implement T1: specifications
  - Start up T2: call
- Data management
  - RDM strategy
  - Priority policies
- Services
  - DMP & DOI
  - Data deposit
- Skills & expertise
  - Data Science (research) coursework MSc
- Stakeholder engagement
  - SADA



Trusted Repositories: Why bother?

More Value!

- Changing nature of research
  - Collaboration (silo to shared)
  - Automated processing
- Provenance and data loss
  - Repeatability
- Quality assurance
  - Innovation beyond primary use
- We're losing important data
  - Dropbox, Amazon, G-Drive, etc. are NOT Trusted Repositories
- Recognition & credit
  - Citation & Altmetrics



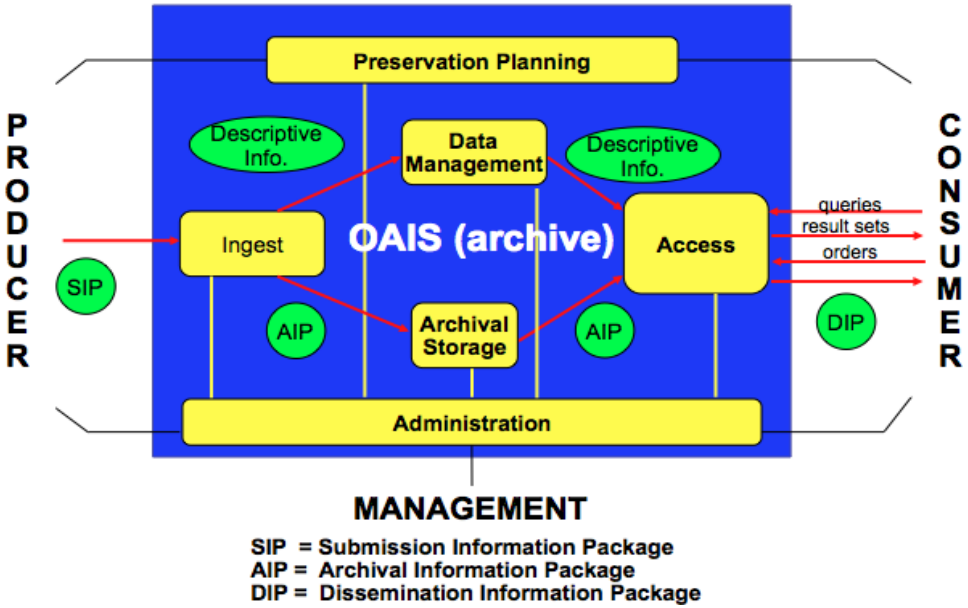
About a Trusted Digital Repository

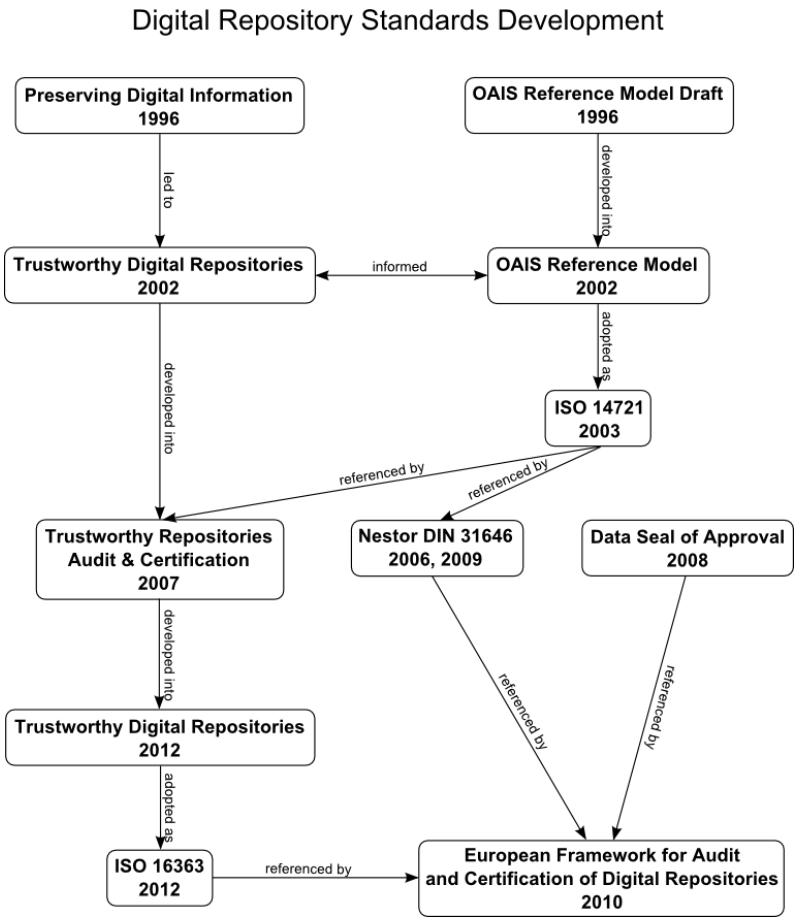
"provide **reliable, long-term access** to managed digital resources to its designated community, now and into the future."

RLG-OCLC Report 2002

- Responsible for long-term maintenance of digital resources (administrative, organisational, financial, technological, system security, procedural)
- Ongoing access and security
- In accordance (compliant) with common convention and standards (OAIS reference model)
- Established evaluation methodologies to audit and measure performance (policies & practices)
- Identity, integrity and quality is open and explicit

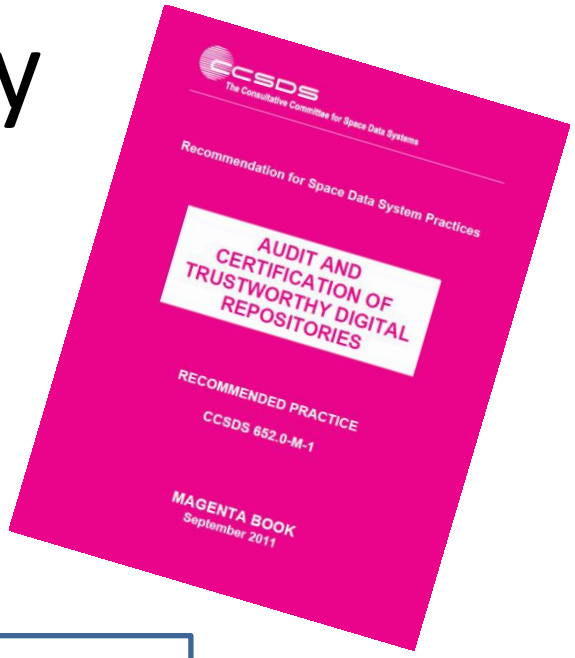
Digital Repository Standards





- **Open Archival Information System (OAIS) Reference Model**: ISO standard for an open archival system
- **Trustworthy Repositories Audit & Certification (TRAC)**: metric for validating an OAIS-compliant digital repository

TRAC Summary



TRAC ISO 16363:2012

Checklist covers

Organisational Structure

25 Measures

Administrative, staffing, financial and legal

Digital Object Management

45 Measures

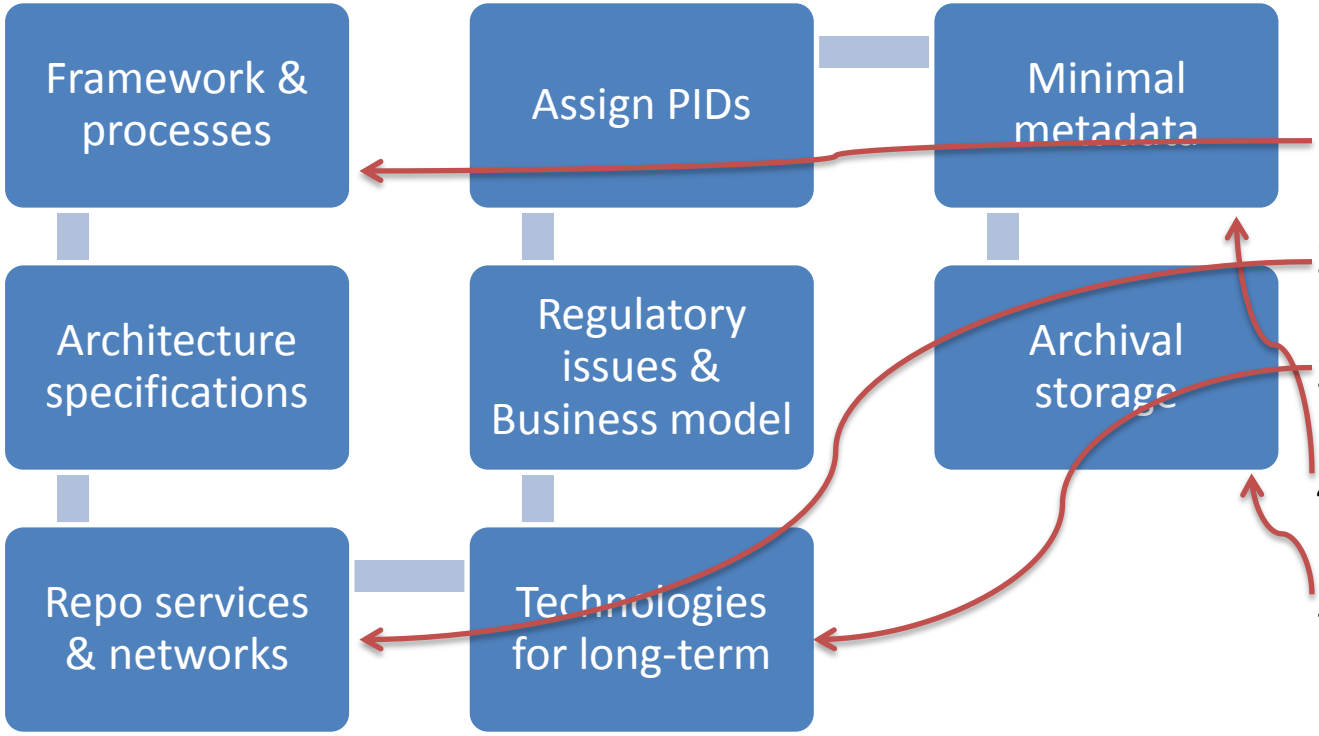
Handling of digital objects from ingest to access. Properties, etc

Infrastructure and Security

17 Measures

Technology used to handle ingested objects and Risk Management

Repo Certification Process



DIRISA Task list

Need stakeholder input!

1. Draft Certification Framework
2. Set up DOI system and DMP tool
3. Configure T1 storage technology & services
4. Develop procedure and template
5. Pilot partner with organisation/s

For organisations:

- What is the organisational strategy?
- How long is "Long-term"?
- Who are the Designated Community (Owners, Custodians, Users)?
- What are the salient features of the data?
- What business model for preservation management (periodic technology refresh, costs, etc)?
- How to scope Value & Significance?

Questions (1)

- So is DIRISA a “national data repository?”
  - Yes, Tier 1
  - No, supplementing the Tier 2/3 one at your institution
- What software are you going to use?
  - Lots available (Archivematica, Islandora,…) but we need coherent, standards-based overarching architecture
  - OpenStack first choice so far
- How long is it going to take?
  - Target: in place by Apr 2016
- What are the next steps as organisation?
  - Orientate management (Who are the DOMs?)
  - Assess archiving options (in-house, sub-contract, external)
  - Identify and package SIPs AIPs and DIPs

Questions (2)

- I need service XYZ. How can DIRISA assist or support?
  - See next question
- How can we work together?
- How much will it cost?

  Suggested business model
  - Tier 1: Free with caveats of Open Data (or embargo period) and DMP
  - Tier 2: Start-up support with caveats of institutional DMP, sustainability plan and federation
  - Tier 3 and beyond: Advisory support

.

THANK YOU

avahed@csir.co.za

Persistent Digital Identifiers

- Every object can be assigned a Digital Identifier
  - Persons: ID number
  - Computers connected to the Internet or Network: IP address or MAC address
- Usually for data / digital objects but even objects that are not digital!

Persistent	Digital	Identifier
• Lasts “forever” • Internationally recognised and part of a “network” of identifiers	• Machine readable • Usually sequence of alpha-num characters	• Unique • Always refer to same object (wherever that object may be)

PIDs and PDIs: What and Why?

- Fundamental importance to automate the "administration" of an object
- Manage access (linkage, reference) and "maintenance" by host/custodian of object (publishers, museums, libraries, data nodes, academia, government,…)

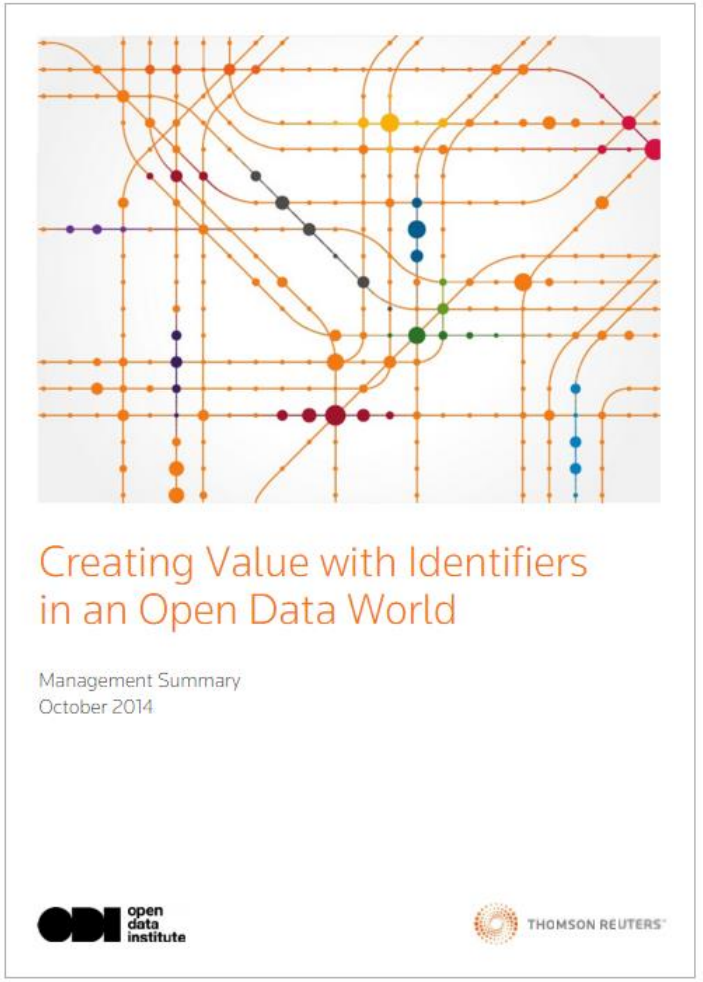
Persistent Identifier (PID): globally unique, long lasting reference to a digital object

Persistent Digital Identifier: machine **actionable** PID that they enable user to access a digital resource via **persistent** link.

Some persistent identifier systems

- Digital Object Identifiers (DOIs)
- Persistent Uniform Resource Locators (PURLs)
- Uniform Resource Names (URNs)

http://theodi.org/guides/data-identifiers-white-paper



PDI Landscape

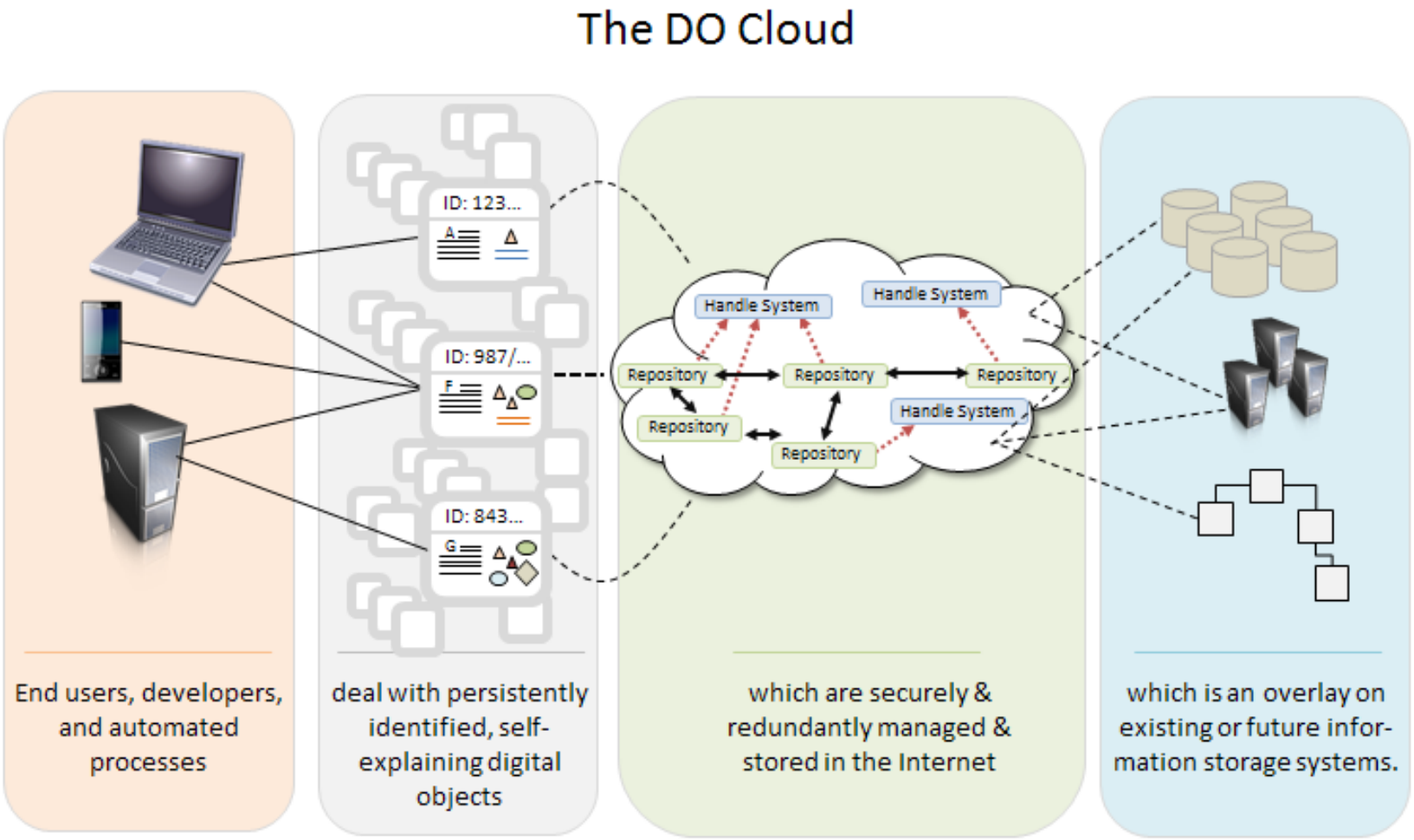
DONA manages **Multi-Primary Administrators** (operating **Global Handle Registry**)

MPAs

- Corporation for National Research Initiatives (CNRI)
- Coalition for Handle Services – China (ETIRI, CDI and CHC)
- Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)

DOI, under the IDF, manages **Registration Agencies**

Digital Object Architecture



From: https://www.doi.org/doi_handbook/5_Applications.html