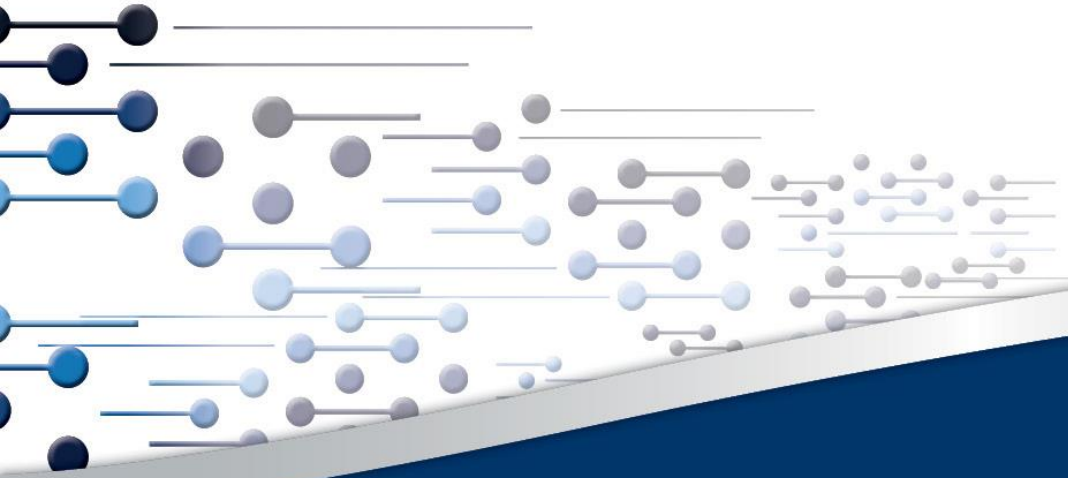
From Process to Practice: Establishing a Research Data Management Function in a Resources-Constrained Environment

Presented by: Adèle van der Merwe avdmerwe@csir.co.za
Co-authors: Martie van Deventer and Louise Patterton



NeDICC

CSIR
our future through science

Roadmap

1. More about the CSIR
2. RDM activities 2007 – 2014
3. Planned activities
4. NeDICC
5. Conclusion

CSIR
our future through science

NeDICC

The CSIR mandate

"The objects of the CSIR are, through **directed** and particularly **multi-disciplinary research** and **technological innovation,** to foster, in the national interest and in fields which in its opinion should receive preference**, industrial and scientific development**, either by itself or **in co-operation with principals from the private or public sectors,** and thereby to contribute to the **improvement of the quality of life** of the people of the Republic, and to perform any other functions that may be assigned to the CSIR by or under this Act."

(Scientific Research Council Act (Act 46 of 1988, amended by Act 71 of 1990)

- The CSIR is a schedule 3b entity: National Government Business Enterprise
- Governed by:
  - National Archives and Records Services of South Africa Act (Act 43 of 1996)
  - Spatial Infrastructure Act (Act 54 of 2003)
  - And many other

The CSIR at a glance

- The CSIR is a science council, classified as a national government business enterprise
- The CSIR's Executive Authority is the Minister of the Department of Science and Technology

69 years in 2014

2411 total staff

1 691 total in SET base

310 doctoral qualifications

~R2.15 bn total operating income

CSIR
our future through science

NeDICC

The mandate unpacked

Fulfilling the CSIR mandate

Inputs | Activities | Outputs | Outcomes | Impact

Impact on economy and society

- Industrial and scientific development
- Improved quality of life of the people of the Republic

RD&I outcomes

- Scientific and technological support to strategic and national initiatives and to industry
- Scientific and technological capabilities, *e.g.* materials science, photonics, robotics, modelling, ICT

RD&I outputs

- **Publications**
- **Reports**
- **Patents**
- **Technologies**
- **Data**

Records Management

RD&I activities

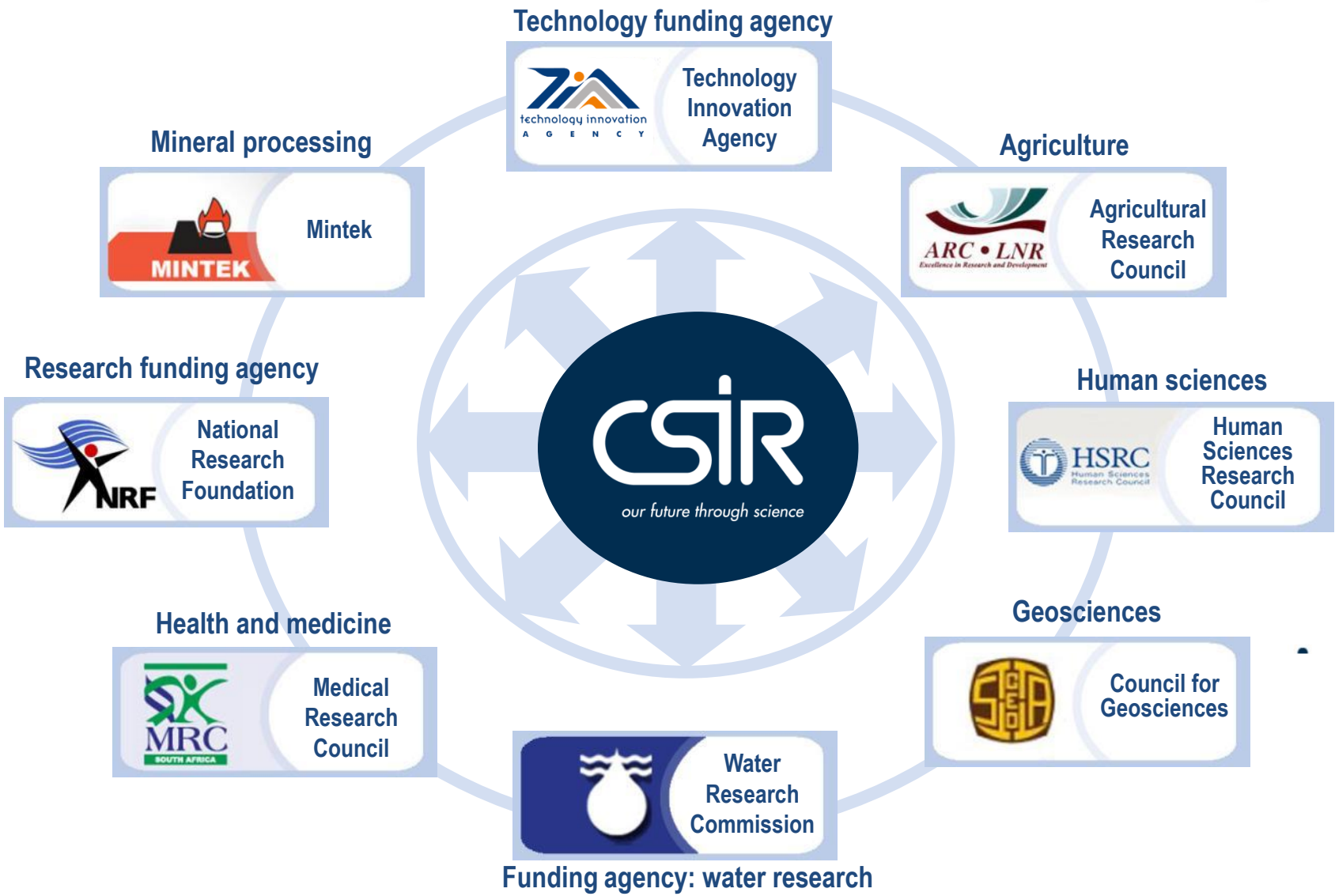
- Research
- Partnering
- RD&I management
- Technology transfer

Inputs

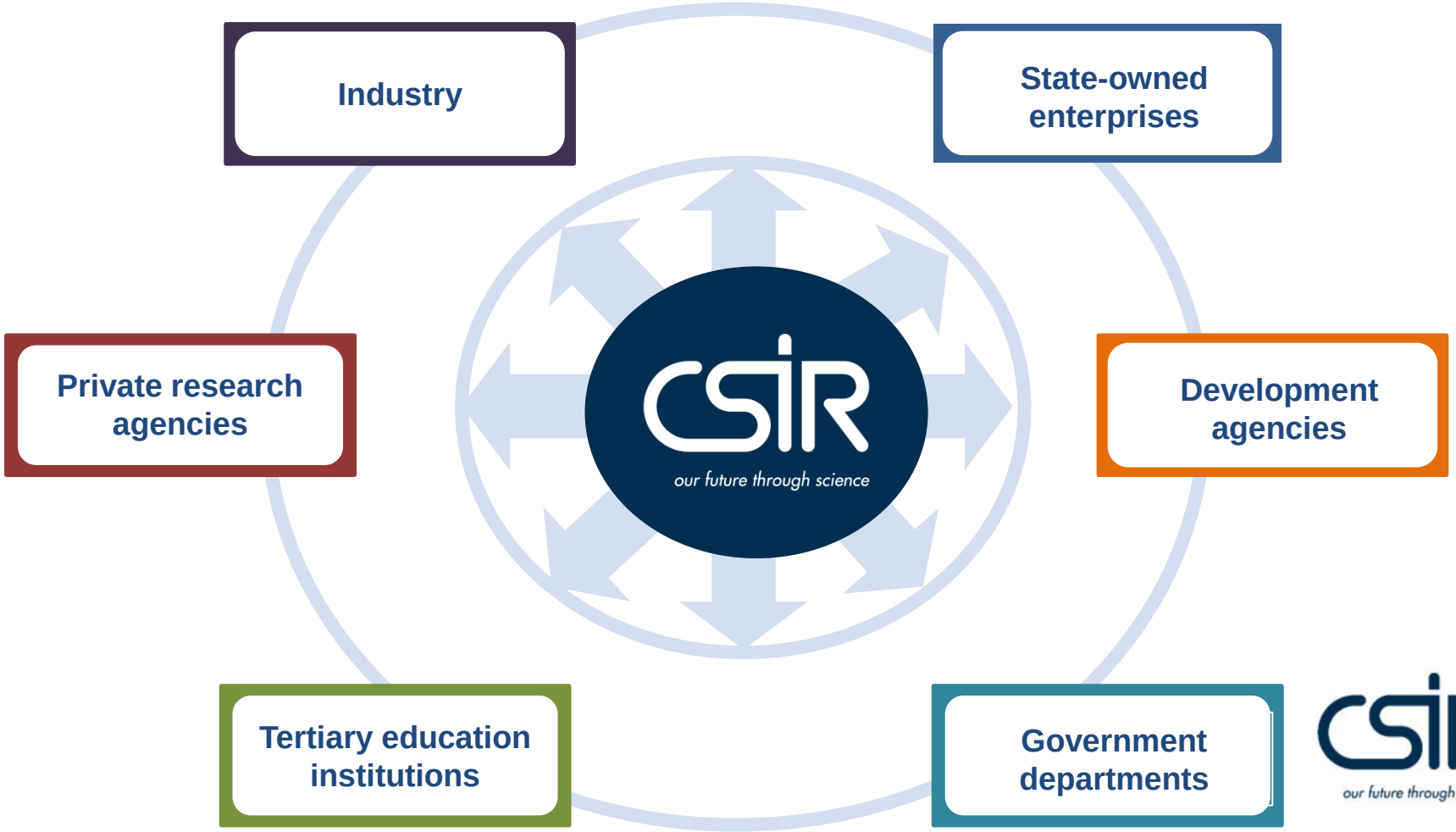
- Human capital
- Research facilities
- Financial resources
- Governance

RD&I: Research, Development and Innovation

The CSIR interacts with other public research institutions



The CSIR interacts with other stakeholders in the innovation system



CSIR
our future through science

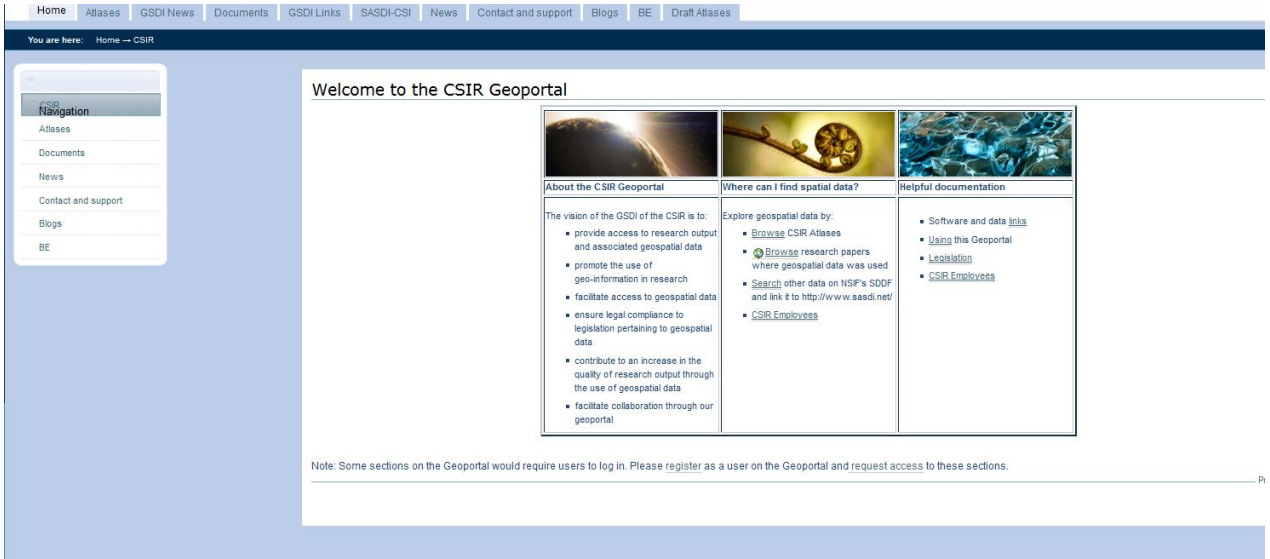
NeDICC

Status report: 2010

- Records management initiative
  - ✓ Focus on research records
  - ✓ Several natural sciences
  - ✓ 'Customised' approach required
  - ✓ File plan approved in 2010
    - ☺ Research data part of project file
    - ☺ Data sets defined as *Data that were accumulated in the process of due diligent research in accordance with a signed research contract*
    - ☹ No dedicated file server
    - ☹ No real understanding of the challenges

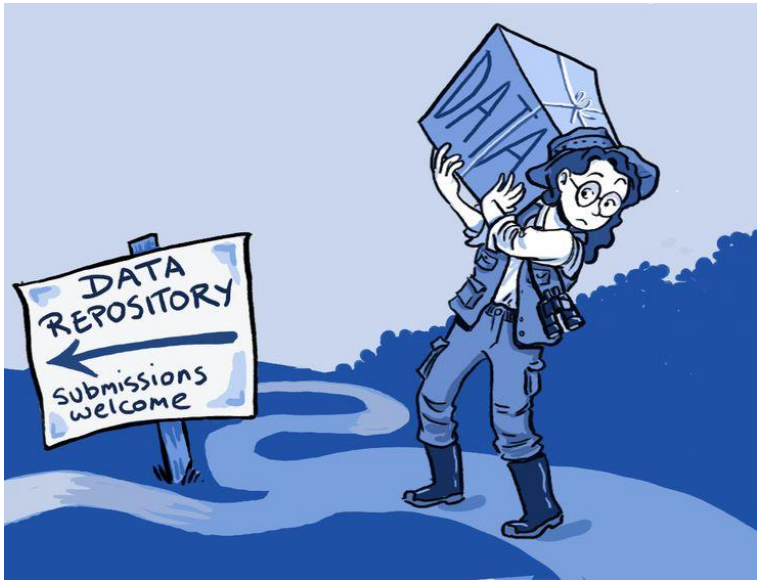
Status report: 2010 (Continue)

- COGIS (Cooperative Geographical Information System) Pilot project (http://gsdi.geoportal.csir.co.za/)
  - Provide access to research output and associated geospatial data
  - Promote the use of geo-information in research
  - Facilitate access to geospatial data
  - Ensure compliance to legislation pertaining to geospatial data
  - Contribute to an increase in the quality of research output
  - Facilitate collaboration
  - Inputs and insights gained from the Geosciences Council



Status report: 2012-2013

- We gained enormous insight and now understood the challenges of RDM **better**
- Full time HR resource: CSIR Data Librarian
  - **Responsibilities**
    - Provide support and guidance
    - Policies, procedures and guidelines
    - Training and support
    - Strategic plan
    - Mentoring
  - **Challenges**
    - Analysis and synthesis of complex concept and issues
    - Management of complex relationships
    - Meeting expectations



CSIR
our future through science

NeDICC

Status report: 2013-2014

- Challenge: understand the complexity of RDM within the CSIR context
- Identify and understand existing behaviour and (if any) good SOPs
- Approach was a survey-based investigation
  - Sample group
  - 23 open-ended questions
  - 36 Research Group Leaders
  - 9 Research units
  - Appointment based

RDM: familiarity with concept

know and apply RDM: 28%

have heard of it: 39%

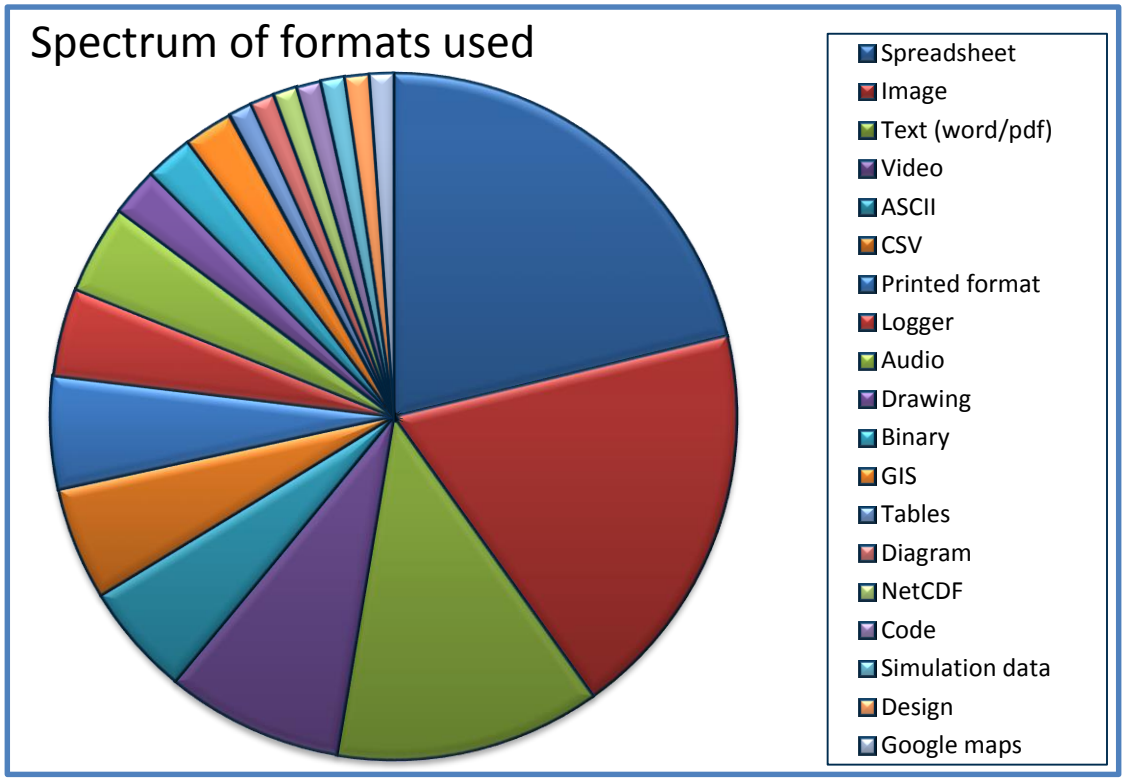
never used or heard of RDM: 33%

CSIR
our future through science

NeDICC

Research data formats

- Most popular formats:
  - Spreadsheet
  - Image
  - Text (word/pdf)
  - Video



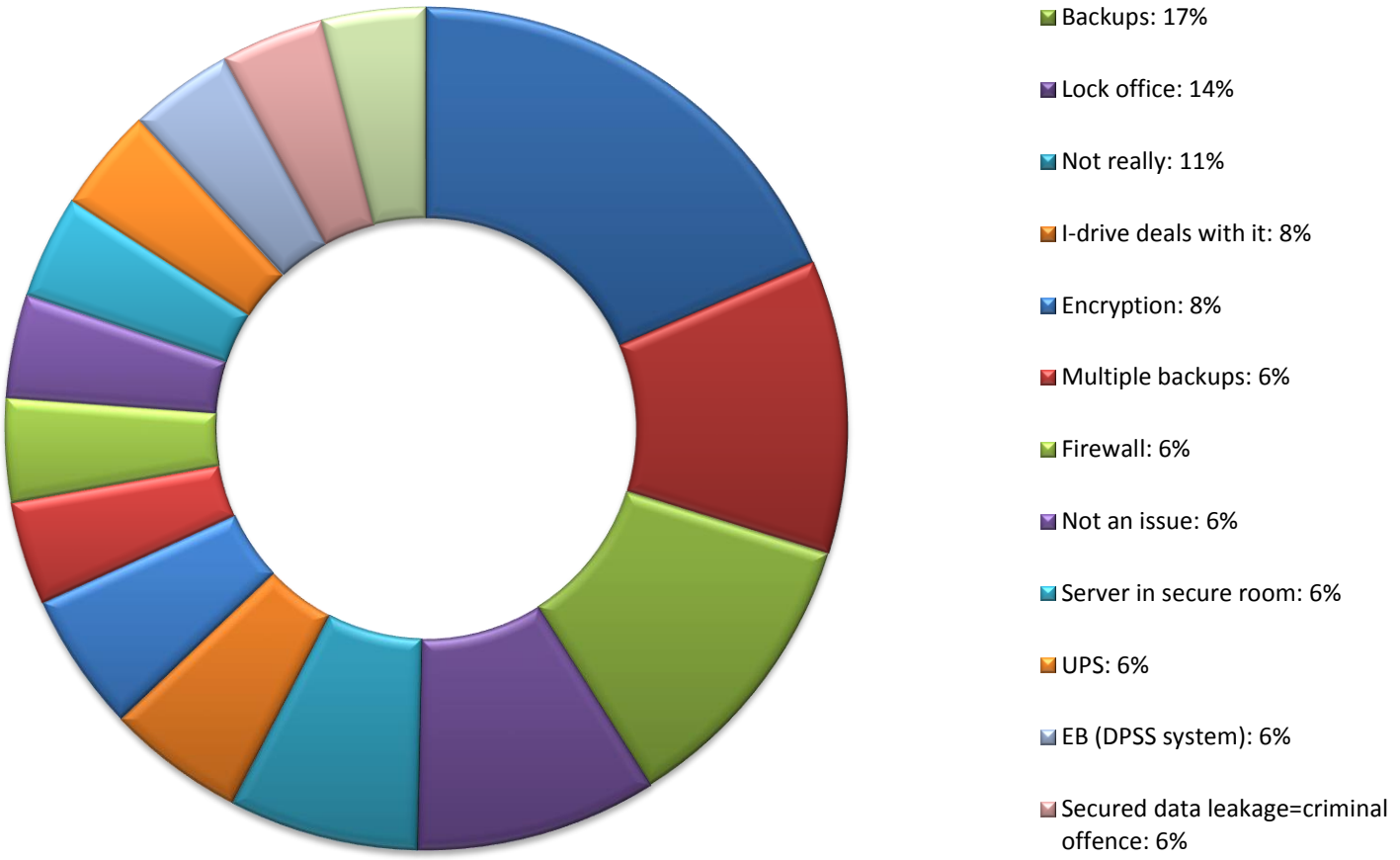
CSIR
our future through science

NeDICC

Where is your research data stored?

Data storage media	Prevalence
PC/laptop	61%
I-drive	47%
External hard drive	28%
Lab computers	11%
server	11%
EB*	8%
Project server	8%

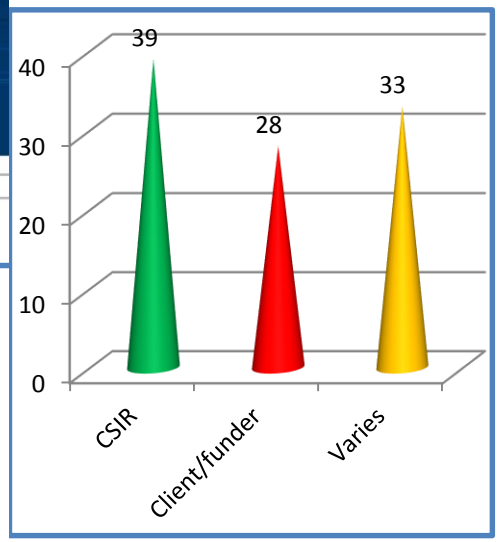
Data security

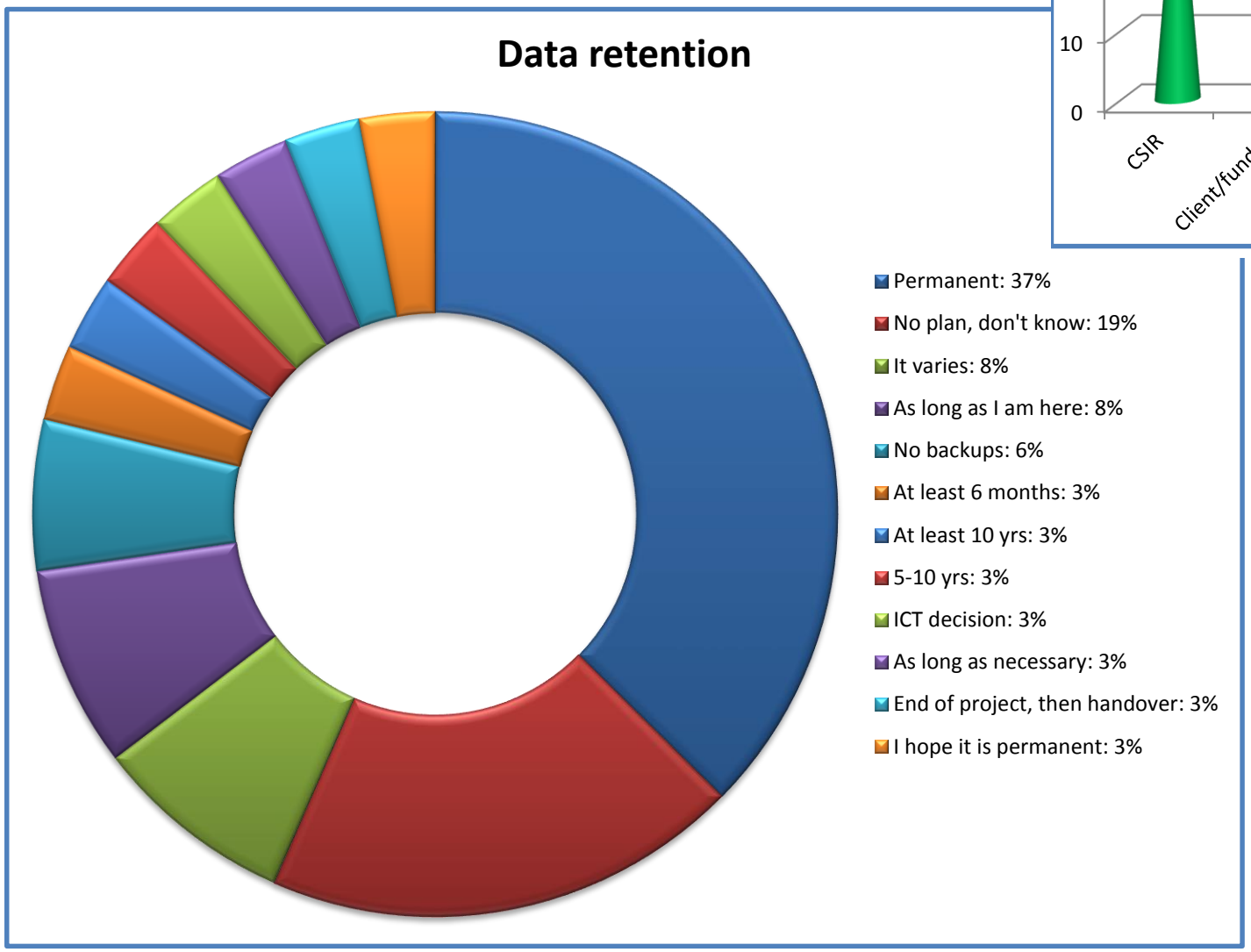


CSIR
our future through science

NeDICC

Data retention and ownership



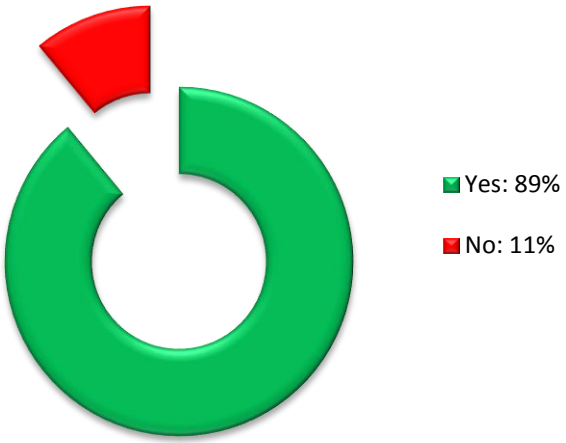


CSIR
our future through science

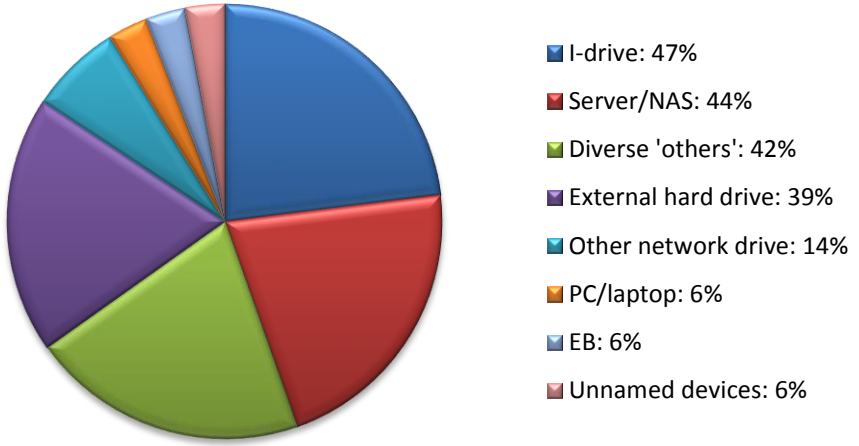
NeDICC

Do you backup your research data, how often and where?

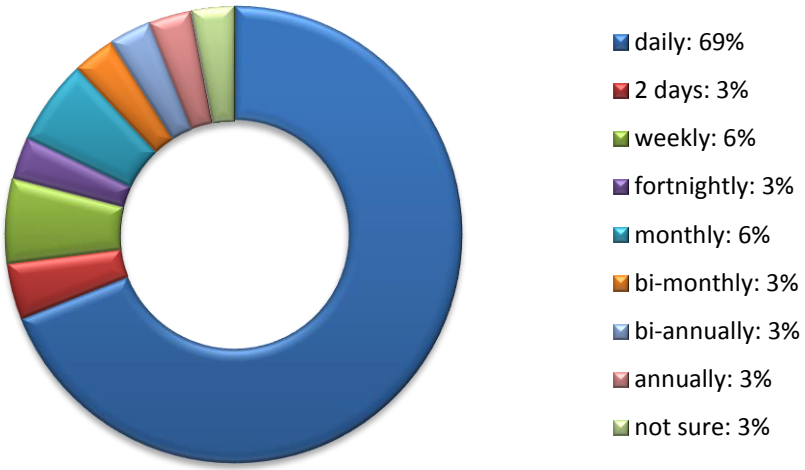
Backups



Backup media used



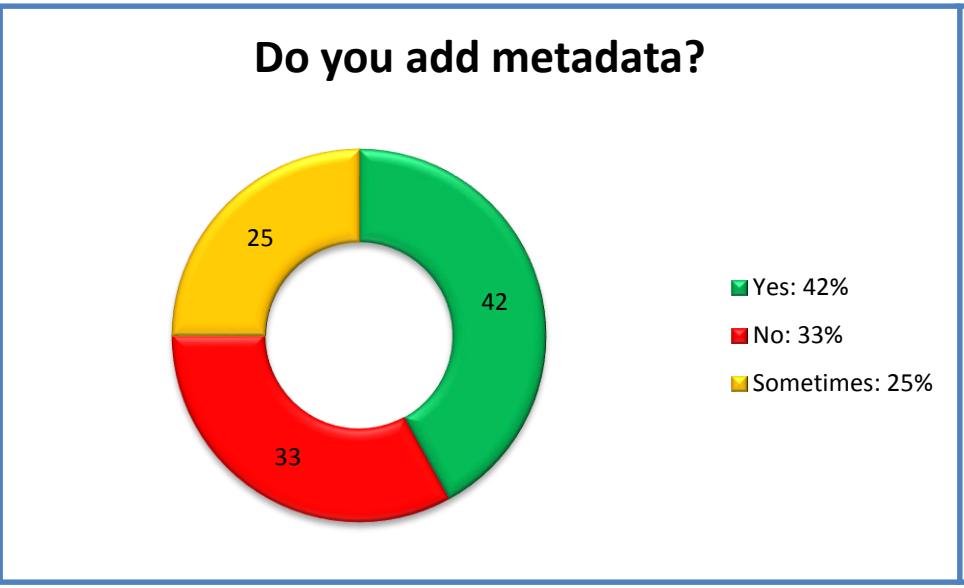
Back regularity

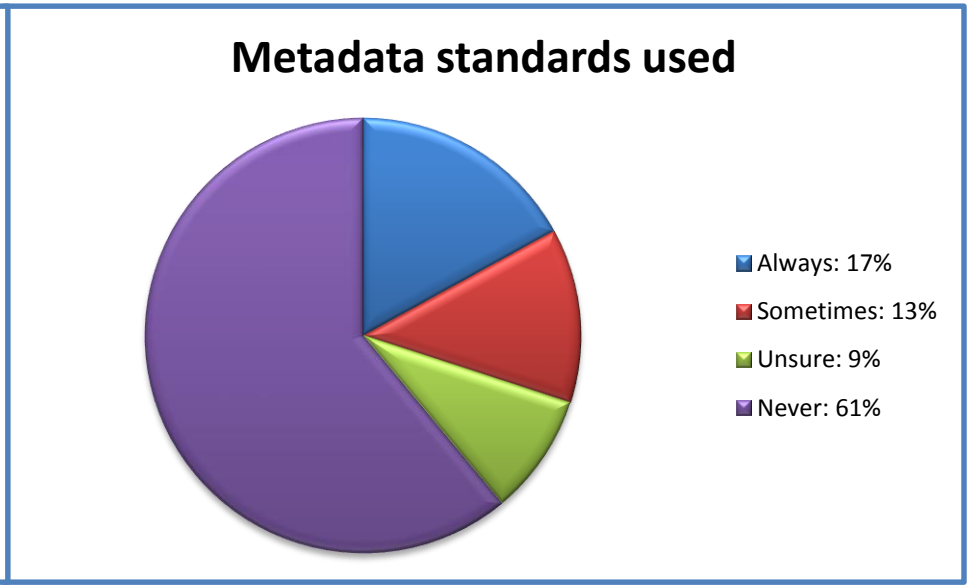


CSIR
our future through science

NeDICC

Do you add metadata?
Do you use metadata standards?

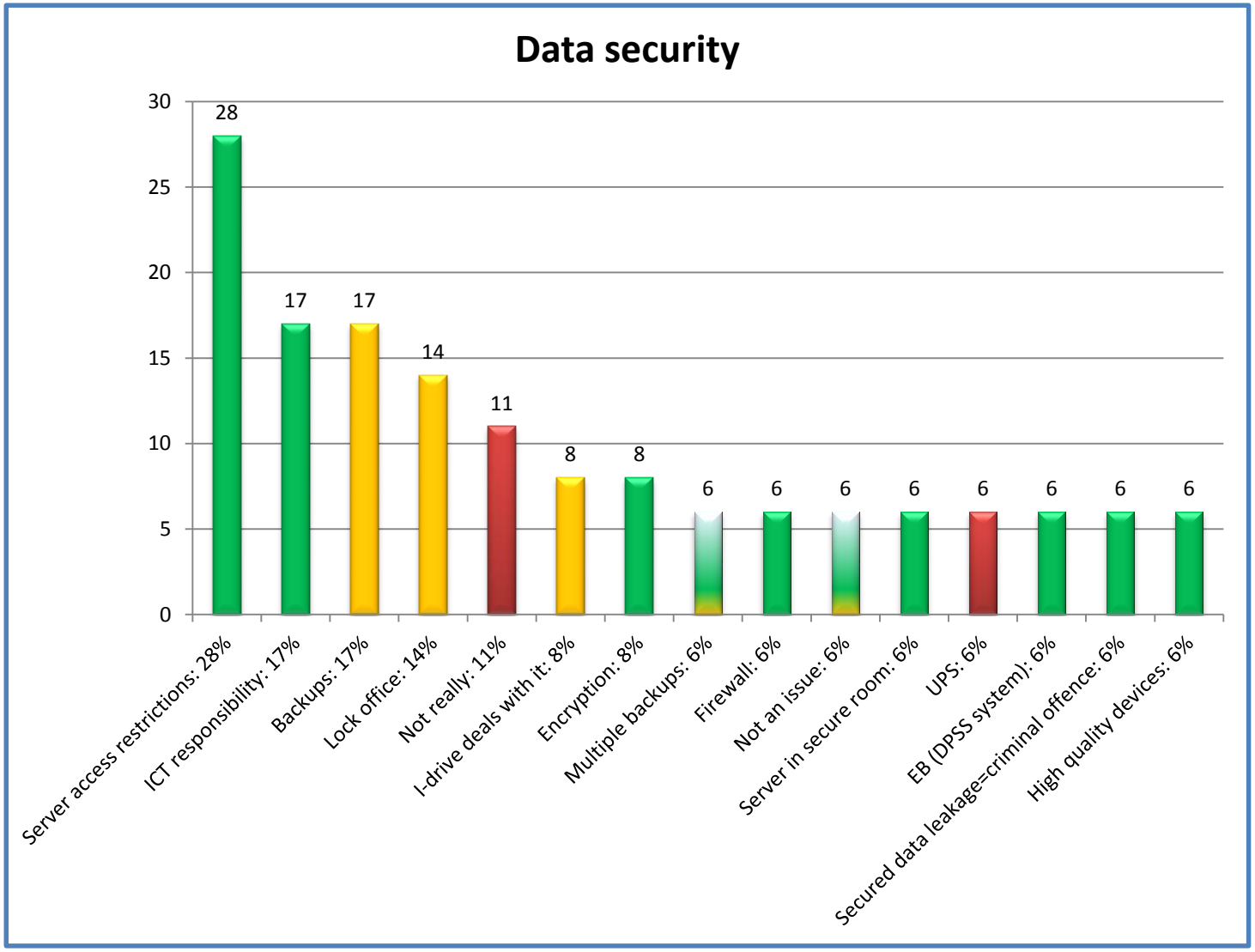




CSIR
our future through science

NeDICC

How do you secure your data?

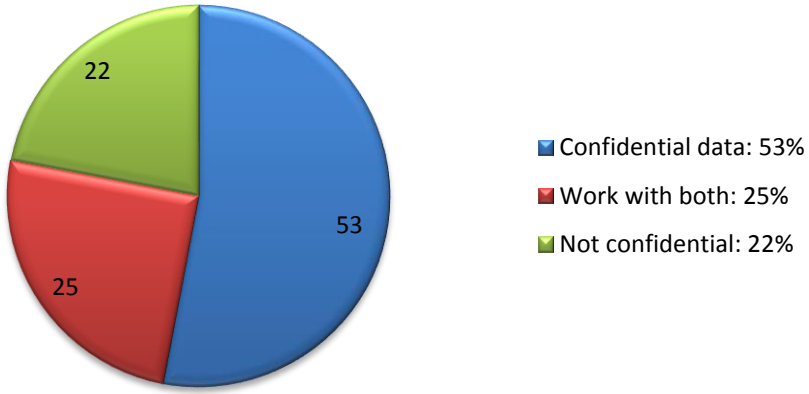


CSIR
our future through science

NeDICC

Is your data confidential in nature/can it be re-used?

Confidentiality of data



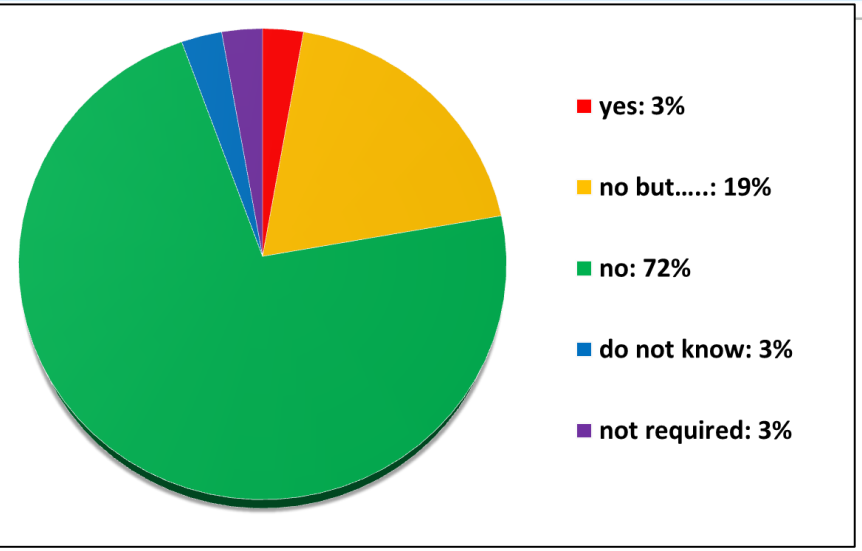
Potential for re-use

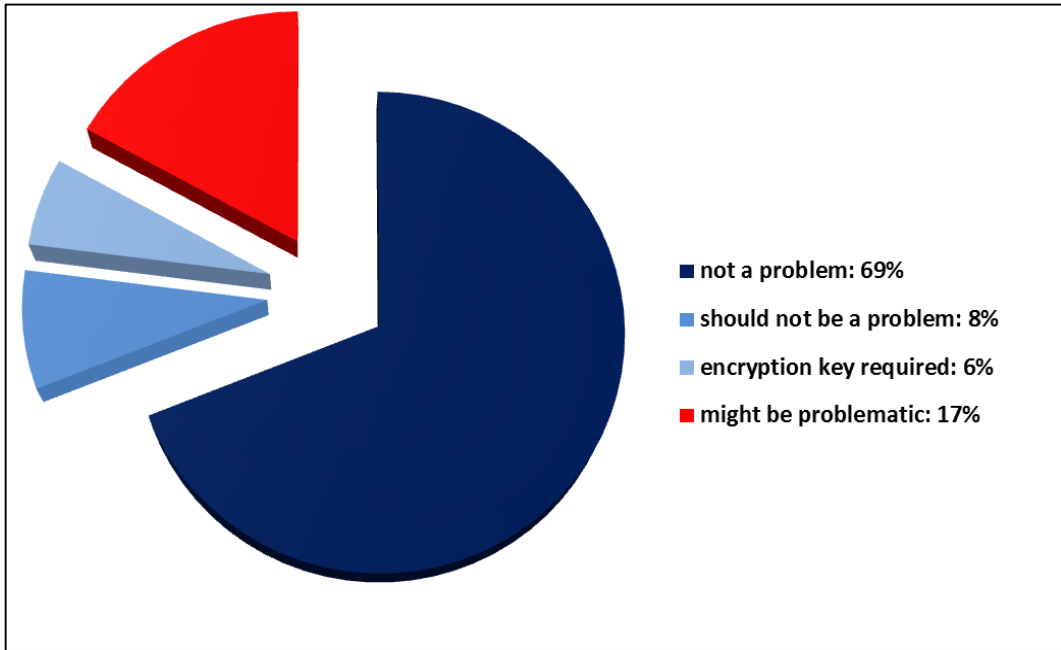
Potential audience

CSIR
our future through science

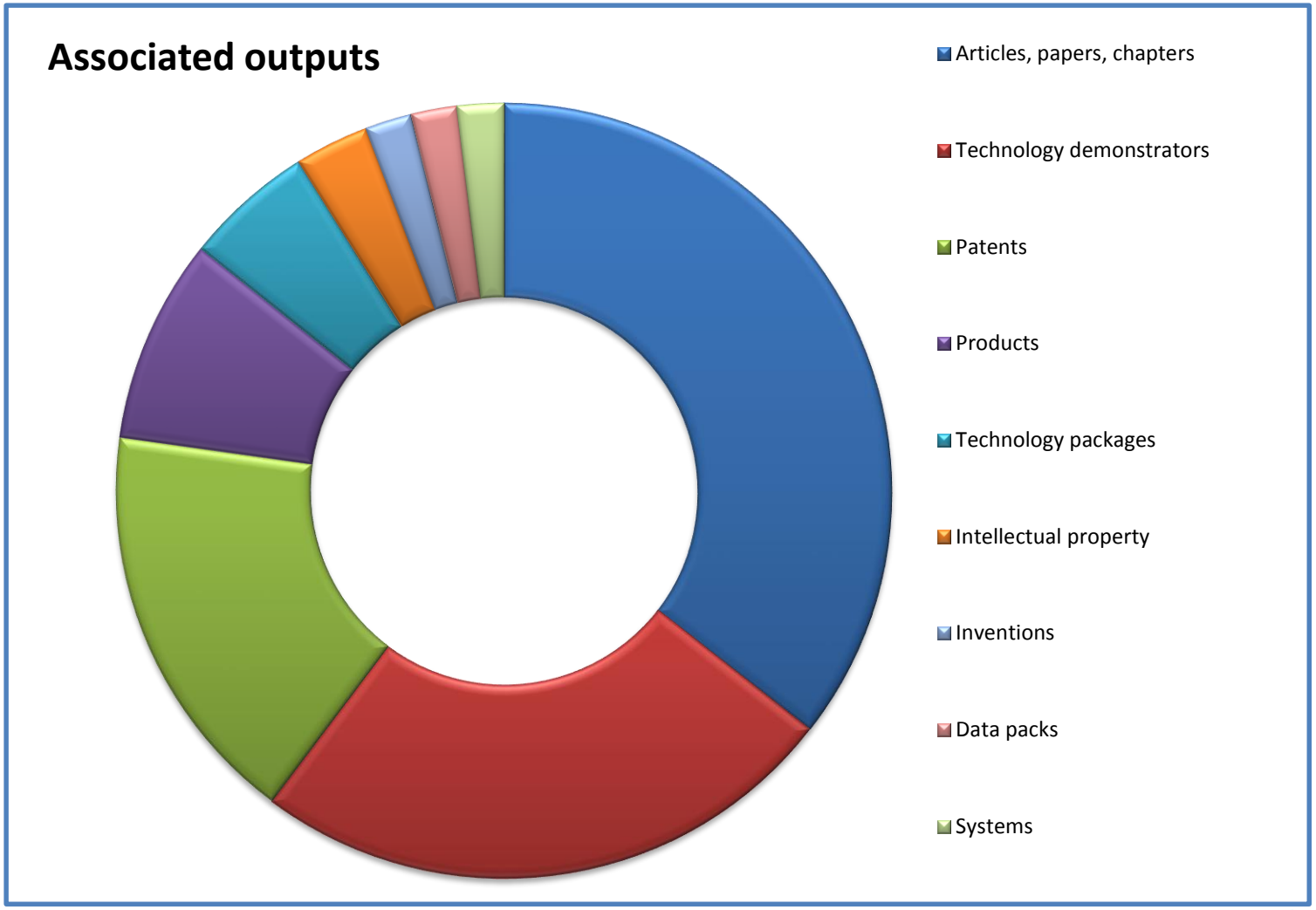
NeDICC

Do you have a disaster recovery plan?





What publications or discoveries result from your data?



CSIR
our future through science

NeDICC

Challenges/obstacles identified during the survey

- IT-related
- Financial
- Software
- RDM practices
- Data security issues
- Data sharing/confidentiality

CSIR
our future through science

NeDICC

CARDIO situation analysis

- Self analysis with inputs from the ICT unit
- Decided on priorities and act on recommendations
- Road to recovery:
  - Project registration ✓
  - Policy to be drafted ✓
  - Training materials ✗
  - Survey findings to be distributed ✓
  - RDM working group/project team ✓
  - Expand/improve existing services ✗
  - Trial project ✗

CARDIO
collaborative assessment of research data infrastructure and objectives

CSIR
our future through science

NeDICC

Next phase: 2015+

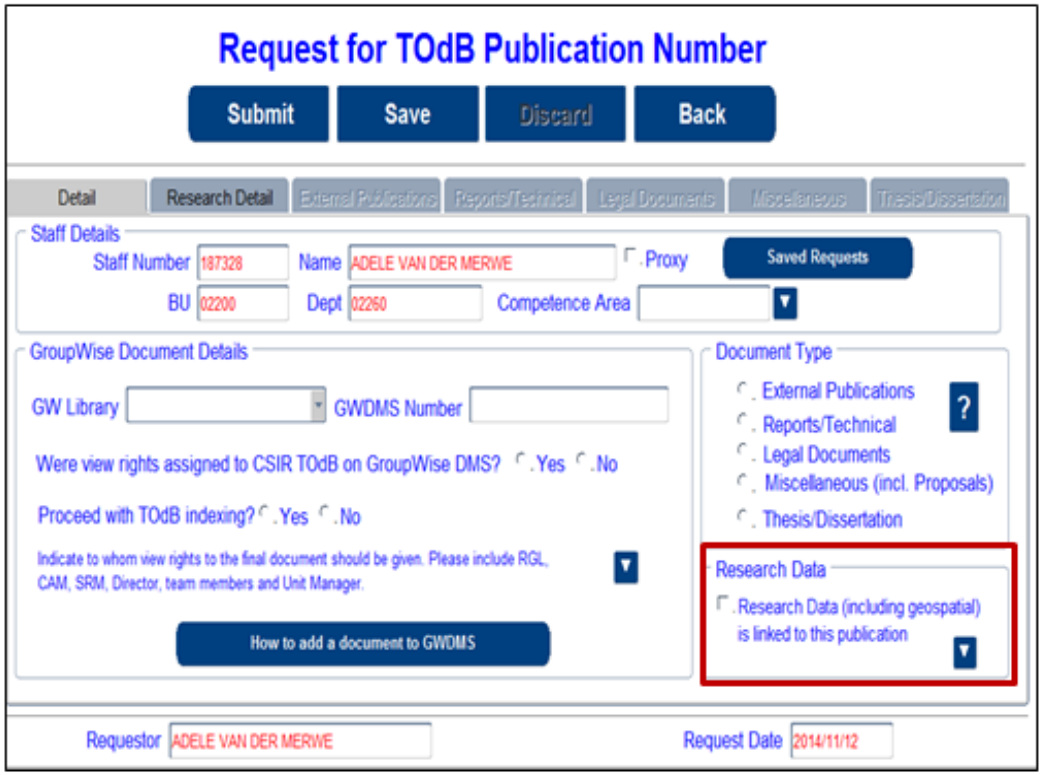
- Diverse range of answers
- Repository → unit-specific data archive
- Data specialist
- Designated cloud for data
- Storage away from building
- Archived/older data separately stored
- Collaboration space
- Guidance/standardised templates
- Compliance vs freedom
- Institutional commitment
- Training
- Marketing of RDM
- Researcher awareness
- Improving existing services

CSIR
our future through science

NeDICC

Workflow system

- **Purpose:** link datasets with research
  other outputs
- **Shortcomings:**
  - Not intuitive enough
  - Not comprehensive enough
  - Lack of awareness
  - Lack of compliance



Workflow engine modifications

Workflow system as a DMP tool:

- Format field:
- Specialised technology field: Identify location (link with format and technology)
  - File server, office, cloud, other
- If digital, URL/URI/DOI/???
- Security classification
  - Unclassified – open to public
  - Confidential – only open to research unit
  - Restricted – access requires permission from project manager
  - Secret – access requires permission from director or higher
- Retention/preservation period after end of contract
  - 5 years
  - 15 years
  - Permanently
- Ownership
  - CSIR
  - Client
  - Vendor
  - Other

CSIR
our future through science

NeDICC

Going forward

- Submit a RDM policy for Board approval
  - Clearly define research data and its role and value
  - Embed preservation as part of the EIM drive
    - Legal obligations
    - IP management
    - Risk management
      - Software and technology obsolescence
    - Trusted repository linking research output
- Build on our strengths:
  - Data as a record - part of KPI reporting
  - Workflow engine - Improve and embed the workflow system as a “DMP tool”
- Continue using CARDIO as benchmarking tool
- Implement all pilot project recommendation as far as possible
- Continue NeDICC involvement and relationship

Proverb: *If you want to go quickly, go alone. If you want to go far, go together*

NeDICC

Network of Data and Information Curation Communities

NeDICC Partners

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Technology funding agency

technology innovation AGENCY
Technology Innovation Agency

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Mineral processing

MINTEK
Mintek

Agriculture

ARC • LNR
Excellence in Research and Development
Agricultural Research Council

Research funding agency

NRF
National Research Foundation

CSIR
our future through science

Human sciences

HSRC
Human Sciences Research Council
Human Sciences Research Council

Health and medicine

MRC
SOUTH AFRICA
Medical Research Council

Geosciences

Council for Geosciences

Water Research Commission

Funding agency: water research

UNISA
university of south africa

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

NeDICC's role

- The provision of a forum
- Provide support and work towards solutions
- Expose the community to new developments and trends, provide opportunities to engage with a wider audience, as well as showcase work and initiatives.
- Develop the knowledge and skills of members.
- Promote awareness/best practices relating to digital preservation, dissemination and use of research outputs.
- Collaborate on projects in support of shared objectives

NeDICC's achievements

- Investigated the role of the funder
- Detailed look at the data management plans
- Experimented with Bag-It as preservation technology
- Received training in the Management of large data
- The integration of RDM with the Ethics process
- Detailed management of human sciences data – across the life-cycle
- Long term preservation activities
- The integration of RDM with the Records Management activities
- Workflows
- RDM within a VRE
- **Training librarians to do RDM (CPD programme)**
- Data citation
- Persistent identifiers
- Getting a grip on publications
- RDM situation analysis

We do this because …

- Our passion is:
  - Information, data, history, culture, organizational memory
- Our goal is:
  - Preservation by means of technology
  - Going forward in an innovative manner
- We need a better process
  - Asbestos poisoning litigation
    - Research records required
    - Research data required
  - 2014 SA Earthquake – Council for GeoSciences
    - Republic Observatory records dating back to 1901
    - Hand-written logs

CSIR
our future through science

NeDICC

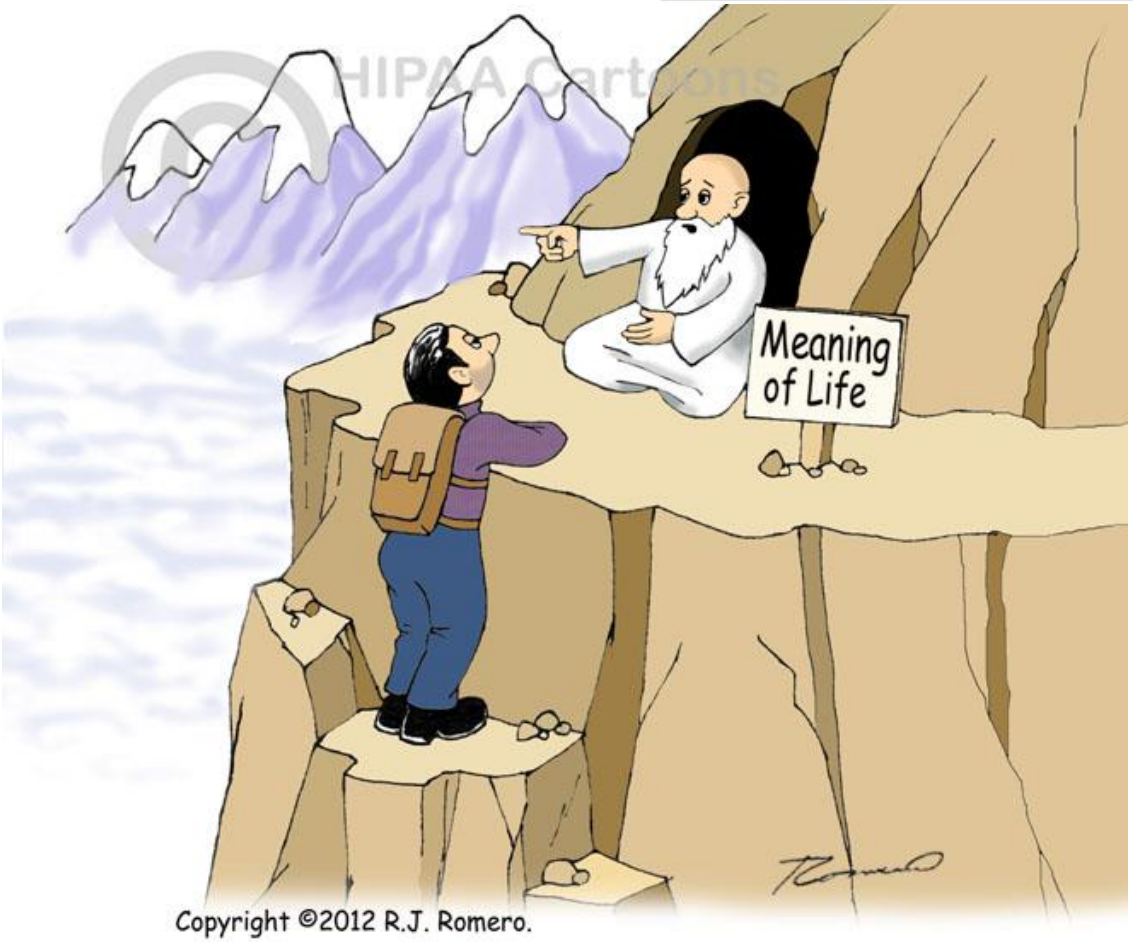
Conclusion

For the immediate future starting now:

- Develop and implement an awareness programme
- Develop and implement a change management programme
- Obtain the funding for and implement technologies
- Align and embed the RDM activities with the CSIR's Enterprise Information Management activities
- Obtain support, buy-in and enthusiasm for the drive from all our stakeholders.
- Continue with our involvement with the growing NeDICC community
- Embed a culture of continuous learning in order to ensure that the RDM drive remains sustainable and focussed.

NeDICC

Thank you and I will refer all questions to my co-authors



CSIR
our future through science

NeDICC