Selective Digitisation of Information

The CSIR's strategy for a sustainable effort

Martie van Deventer

Portfolio Manager

Council for Scientific and Industrial Research, South Africa

1st International Conference on African Digital Libraries and Archives. Addis Ababa: 1-3 July 2009

our future through science

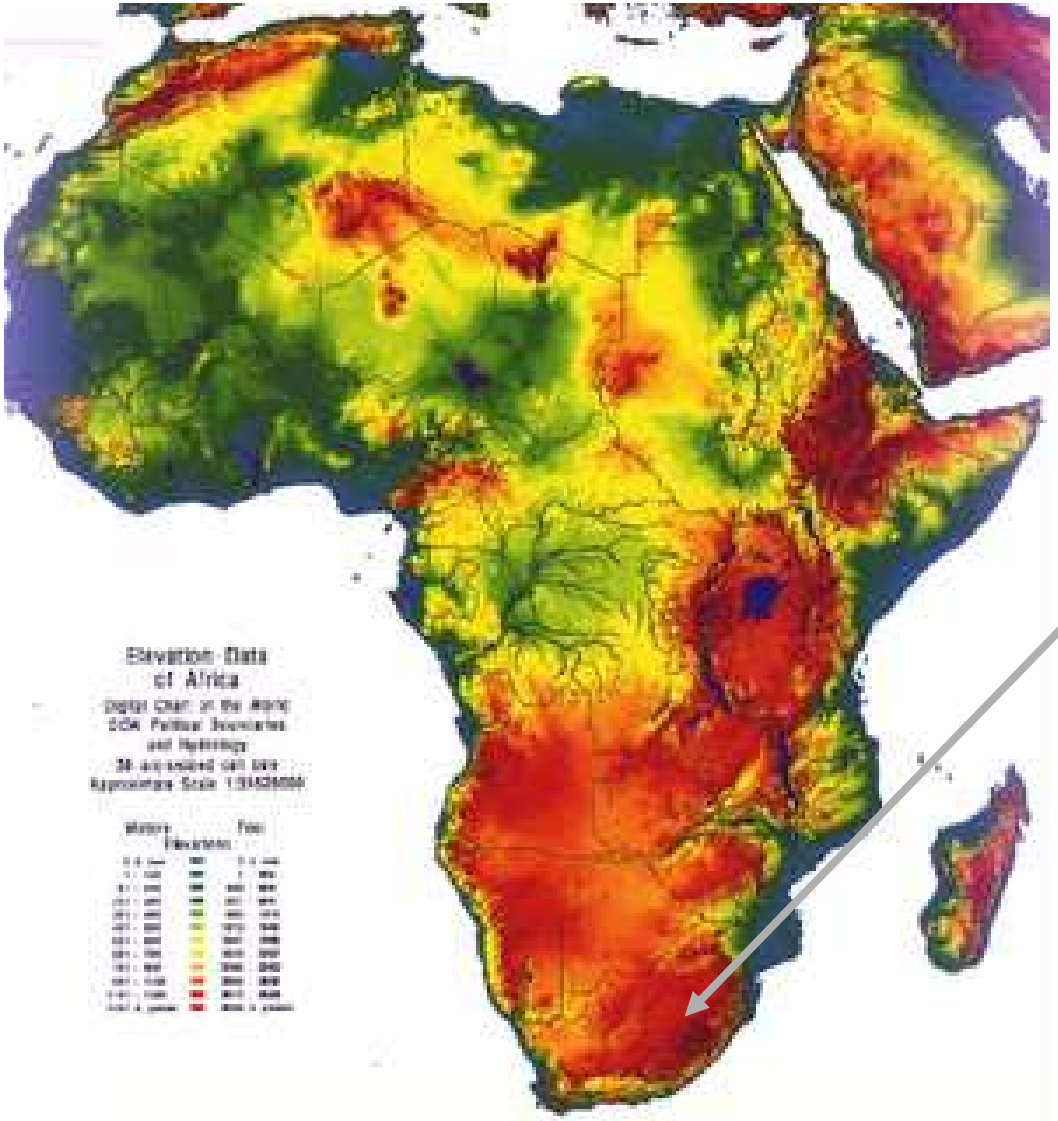
Acknowledgement

- George Harper – responsible for selecting and digitising the SIMRAC reports
- Pat Morant – responsible for the project planning and gaining funding for the digitisation of the ‘Estuaries of the Cape’ report collection
- Johan van der Molen – responsible for the planning and digitisation of the CSIR’s diatom collection
- Adele van der Merwe – responsible for managing our digitisation contracts
- Siphethile Muswelanto – responsible for maintaining and promoting our Institutional repository

Presentation Roadmap

- The Council for Scientific and Industrial Research of South Africa (CSIR)
- CSIR strategy for digitisation
  - Text
  - Special research collections, eg Diatoms
- Way forward

Where in Africa …?



Main site:

- Pretoria

Offices:

- Johannesburg,
- Stellenbosch,
- Durban and
- Port Elizabeth

Total staff: 2500

our future through science

The CSIR Mandate

"The objects of the CSIR are, through directed and particularly multidisciplinary 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 46 of 1988, amended by Act 71 of 1990)

CSIR Research and Development themes

Energy
Renewables

Natural environment
Wise use & safe future

Health
Affordable new treatments for disease. Better food

Built environment
Transport & Human Settlements

Defence & Security

Advanced Manufacturing

Mining
Extraction & Safety

ICT

Materials

Key Research Technologies

The CSIR has a long history … > 60 years of research

The CSIR's research output is therefore enormous … much of it's research history is paper based but we also have several collections of 'samples' (eg wood, textiles, diatoms)

What should we digitise & where should we start?

What did we know about digitisation?

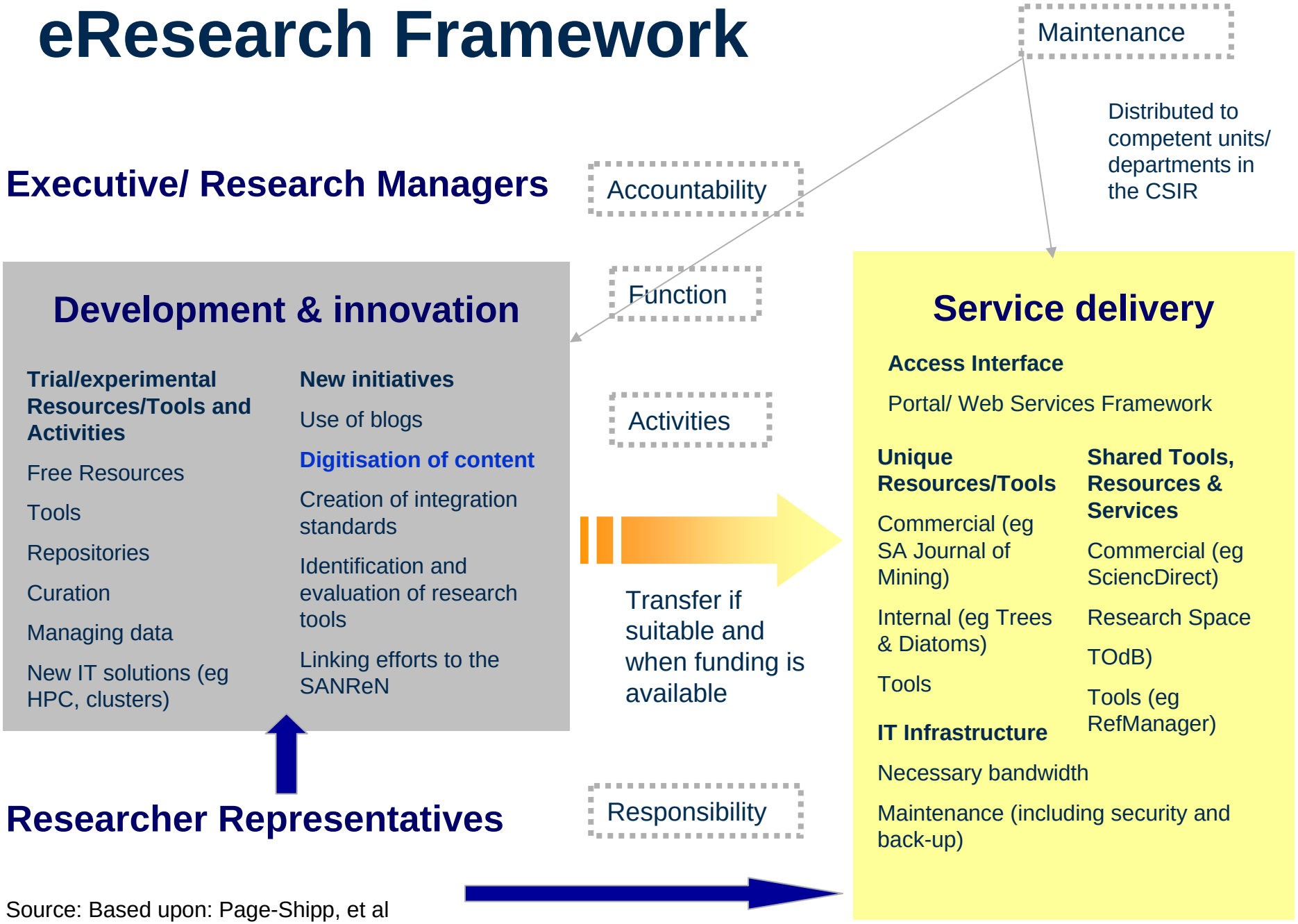
- Much advice and best/good practice guidelines – no need to stumble in the dark or reinvent the wheel
- Experienced labour is required
  - Each type (document, photograph, heritage object) has its own constraints and peculiarities – but the process to digitise any of these have been documented
  - There are many tips and tricks when it comes to file size in a bandwidth constraint environment
- Equipment is expensive … or not … it depends on volume and type

Questions we needed to answer …

- Is the CSIR able to afford the effort and resources required to place historic research output online?
- If not … what is sustainable to do?
- Sustainability: requires a look into the future
  - Once digitised you actually need to maintain two collections … could we afford to do so?

Step 1: We added digitisation to our eResearch guiding framework …

eResearch Framework



Source: Based upon: Page-Shipp, et al 2005

Step 2: Created a strategy to digitise on demand and to, as far as is possible, incorporate digitised items within our institutional repository collection …

Step 3: We make use of outsourcing agents to digitise sets of items

- no equipment to buy and no equipment maintenance
- could build staff expertise is in quality checking

Step 4: We make use of technical staff and inexpensive equipment to digitise individual small documents

Step 5: We maintain a secure master paper collection for digitised text

CSIR Research Space – our institutional repository

- Using DSpace as the repository application
- Contains 2424 items (but grows daily)
- Focus on adding born digital content
- … but we have also added >350 digitised items
  - Especially for adding content created before 1990
  - Digitisation has thus far mainly been text based documents
- Specifically encourage using the repository
  - It is harvested widely
  - Single source of CSIR output

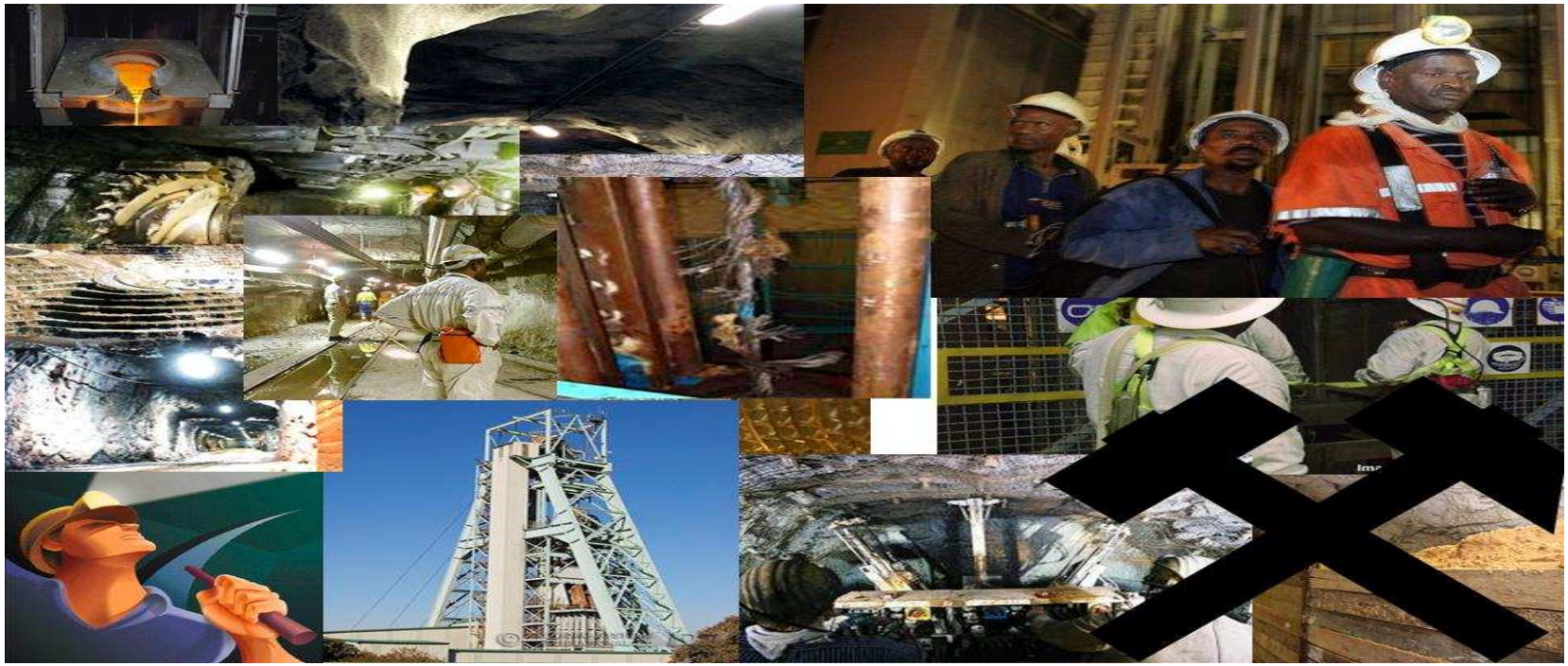
Digitising text

- On demand – any report, paper or article for which we hold copyright and which could be placed in open access
  - Format: only use .pdf
- Mainly 3 report series
- Also CSIR photographs – not open to the public
  - Format: .jpg
- Experimenting with video and sound files but not much success yet – files are too large

Estuaries of the Cape series – 42 Reports
Focus: State of all South African estuaries

Future: digitisation and linking of aerial photographs

Safety in Mines Research Advisory Council Reports – 229
Focus: Mine health & safety



Future: digitisation of the larger report collection

South African Scientific Research Programme Reports - 152

- Many are South African biodiversity related
- Several bibliographies (eg for Fynbos research)
- Rivers, Fire, Invasive species
- Marine ecologies
- Antarctic research programme
- Red data books for birds, butterflies, reptiles, fish, small animals
- Several on early water research and early climate change studies

Future: South African Wool & Textile Research Institute Collection

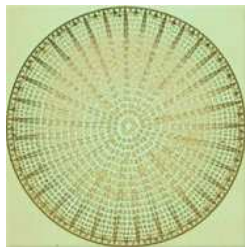
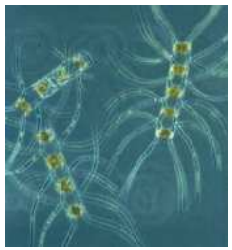
Digitised documents reached the top 20 list

(April 2008 to March 2009)

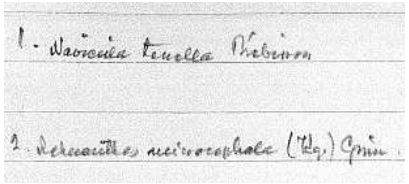
- **For perspective**
  - #1 downloaded 2388 times, and
  - # 20 was downloaded 414 times
- **For digitised content**
  - # **5** Whillier, A. **1953**. Utilization of solar energy in South Africa. Journal of South African Mechanical Engineers, vol. 2(9), pp 1-7 (downloaded **1230** times)
  - # **8** Basson, FA, Jammine, E and Heyl, L. **1984**. Acceptability of the integral solar water heater by householders in the low income urban community. National Building Research Institute (**1080**)
- **Lesson**
  - Perhaps start looking for themes to digitise

Digitising scientific collections

- Only one example discussed here – Southern African diatom collection
- Diatoms are found in water bodies in Southern Africa, mostly rivers
- Collection started ~ 1950
- Properly curated but was in disuse since ~1990
- Collection was inaccessible
- Renewed interest since ~ 2004 (linked to climate change)


Diatom collection



- Sample bottles (8 000)
- Microscope slides (20 000)
- Analyses sheets
- Literature (books (350) & papers (5 000))
- Maps
- Glass plates

Digitisation Phase 1

- Catalogue of bottle collection in database
- Catalogue of slide collection in database
- Digitised analysis sheets and link to slides (in part)
- Import literature reference data
- Digitised microscope slides (in part)
- Database is on-line ~ 2009
- Collection has been transferred to the North West University – for further research
- (Inter) National interest is growing

Digitisation Phase 2

- Digitise and link the remainder of analysis sheets
- Focus on the eResearch components
  - Extend functionality of database
    - Spatial data
    - Taxonomic data
  - Compile taxonomic standard list
  - Compile geo-reference data - will enable distribution maps
  - Link to River Database - will facilitate selection of reference sites for River Health Programme

Digitisation - going sustainably forward

- Will continue to actively support the contribution of African content to international harvesters
- Will continue to digitise on demand … thematically
- Will continue to outsource the digitisation effort
- Will encourage further scientific collection digitisation efforts
- Realise that an increase in digitisation requires dedicated project management capacity
- Realise that curation of content means we need to limit the digital formats we use … education role to play but we have to
  - Add photos
  - Add datasets
  - Add geospatial information
- Next phase - need to link and digitise research output artefact collections to new research focus areas

References

- Carignan,Y. et al., 2007. Best practice guidelines for digital collections at University of Maryland libraries
- Government of South Africa. 1990. Scientific Research Council Act 46 of 1988, amended by Act 71 of 1990
- IFLA Working Group. 2002. Guidelines for digitization projects, for collections and holdings in the public domain, particularly those held by librariesand archives.
- Minerva Working Group 6. 2004. Good practices handbook
- Page-Shipp, RJ., et al. 2005. eResearch support services: responding to a challenge facing the South African research and information communities. SA Journal of Information Management, December 2005, 7(4) http://www.sajim.co.za (Accessed 5 June 2009)
- Van der Molen, J. 2009. Digitisation of the South African diatom collection. Available http://hdl.handle.net/10204/3437
- Van der Molen, J. 2009. South African diatom collection: a future for historic data. Available http://hdl.handle.net/10204/3438