

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

GW Pta Gen 179079

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



our future through science

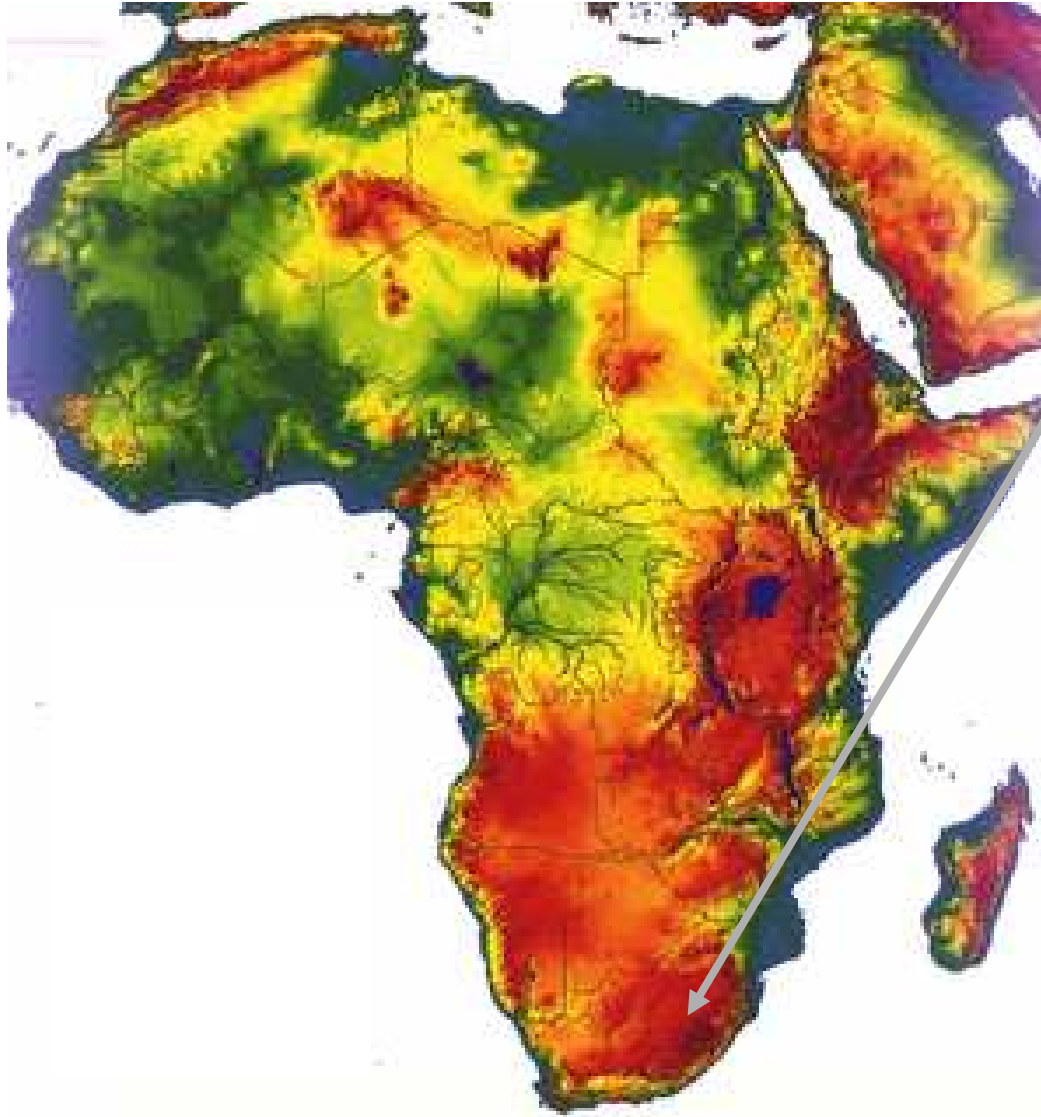
Acknowledgement

- 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

Researchers: 750



The CSIR Mandate

“The objectives 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 its research history is paper based but we also have several collections of 'samples' (eg wood, textiles, diatoms)

2007: What should we digitise & where should we start?

What could we establish about digitisation?

- Much advice and best/good practice guidelines – no need to stumble in the dark or reinvent the wheel
 - 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
- Experienced labour is required
- Equipment is expensive ... or not ... it depends on volume and type

Our process ...

- **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?
 - What digital formats would the CSIR be prepared to maintain?
- **Initiated a five step strategy**

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

eResearch Framework

Executive/ Research Managers

Development & innovation

Trial/experimental Resources/Tools and Activities	New initiatives
Free Resources	Use of blogs
Tools	Digitisation of content
Repositories	Creation of integration standards
Curation	Identification and evaluation of research tools
Managing data	Linking efforts to the SANReN
New IT solutions (eg HPC, clusters)	

Researcher Representatives

Accountability

Function

Activities



Transfer if suitable and when funding is available

Responsibility

Maintenance

Distributed to competent units/ departments in the CSIR

Service delivery

Access Interface
Portal/ Web Services Framework

Unique Resources/Tools	Shared Tools, Resources & Services
Commercial (eg SA Journal of Mining)	Commercial (eg SciencDirect)
Internal (eg Trees & Diatoms)	Research Space
Tools	TOdB)
IT Infrastructure	Tools (eg RefManager)
Necessary bandwidth	
Maintenance (including security and back-up)	

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 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 until our digital curation processes have been finalised

CSIR Research Space – our institutional repository

- CSIR actively supports OA but needs to keep client contracts in mind
- Using DSpace as the repository application
- Currently contains >2424 items (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
 - Experimented with sound and film ... files are very large
- Specifically encourage the use of 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

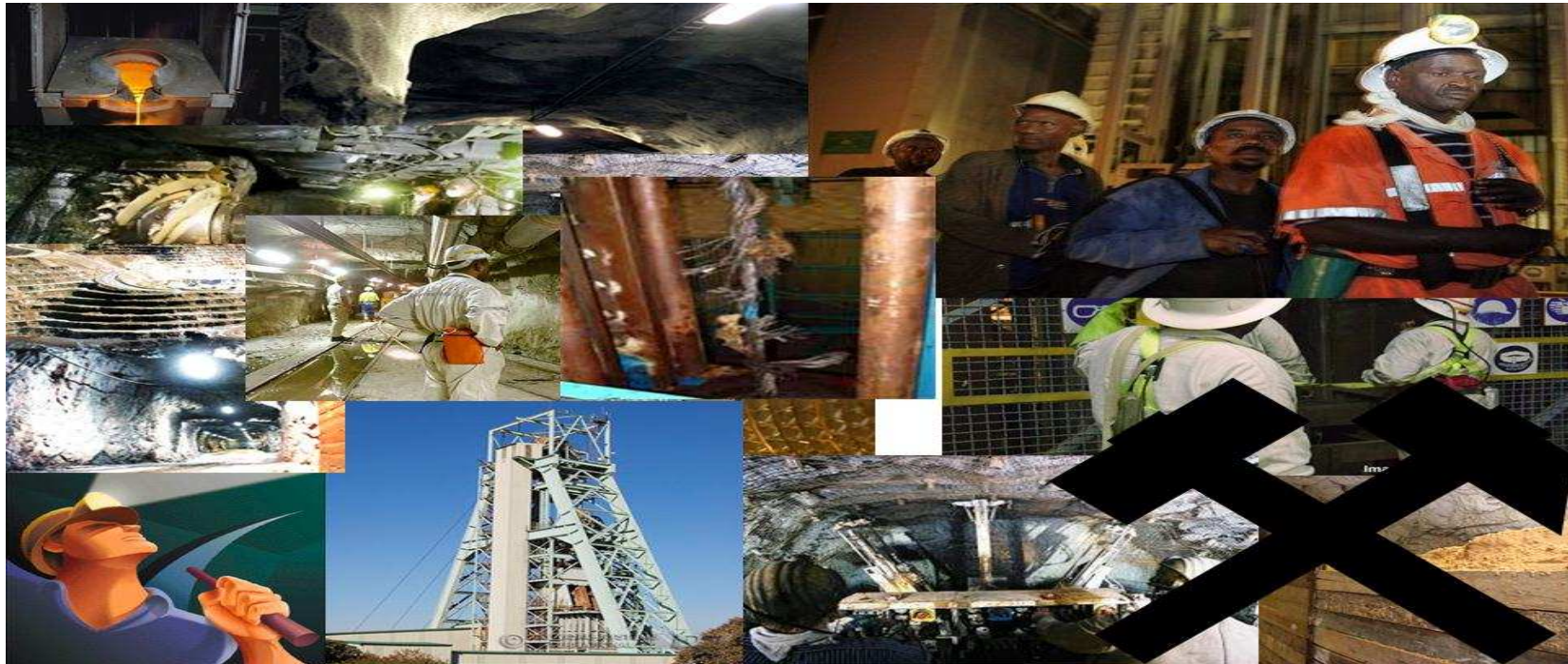
Estuaries of the Cape series – 42 Reports

Focus: State of all South African estuaries



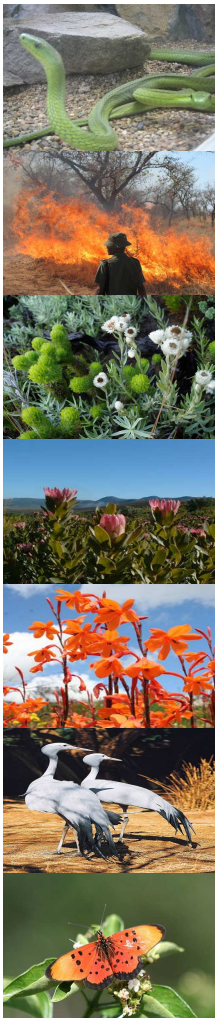
Future: digitisation and linking of aerial photographs

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



Future: digitisation of the larger report collection
>2 500 reports

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 & the Antarctic research programme
- Red data books for birds, butterflies, reptiles, fish, small animals
- Several reports on early water research and early climate change studies

Future: South African Wool & Textile Research Institute Collection & add CSIR diatom reports and articles

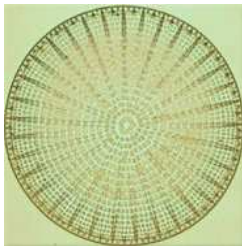
Usage ... should we continue?

- **2 Digitised documents reached the ‘top 20 items downloaded from the repository’ list** (April 2008 to March 2009)
- **For perspective**
 - #1 published in 2008, was downloaded 2388 times, and
 - # 20 published in 2008, 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, Jammie, E and Heyl, L. **1984**. Acceptability of the integral solar water heater by householders in the low income urban community. National Building Research Institute (downloaded **1080** times)
- **Lesson**
 - Perhaps start looking for themes to digitise

Digitising scientific collections

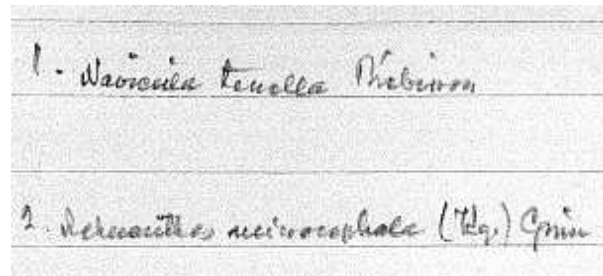
Diatoms

- Only one example discussed here – Southern African diatom collection
- Diatoms - micro algae occurring in marine, estuarine and freshwater habitats
- Collection started ~ 1950
- Properly curated but has been in disuse since ~1980
- Collection was inaccessible
- Renewed interest since ~ 2004 (why linked to linked to climate change)



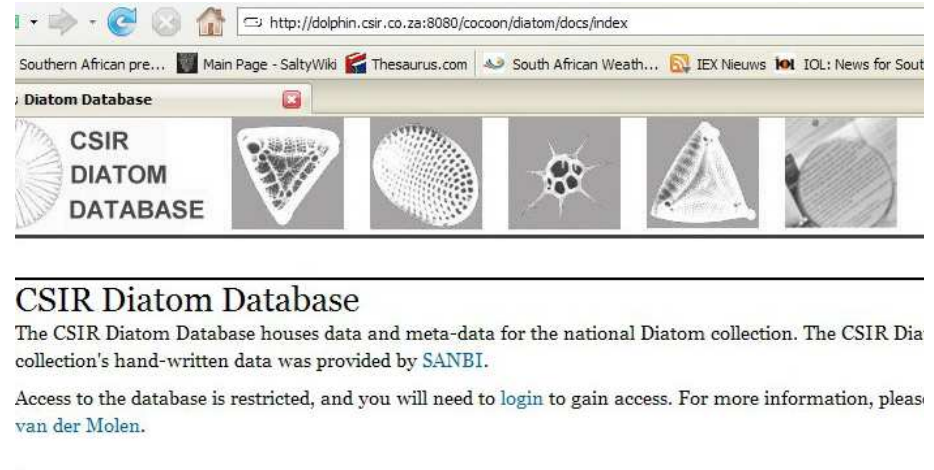
Diatom collection

- Sample bottles (8 000)
- Microscope slides (26 000)
- Analyses sheets
- Literature (books (350) & papers (15 000))
- Maps



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
- Physical collection has been transferred to the North West University – for further research, digitisation funded by SANBI – database now belongs to them
- (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

Next steps ...

- 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 ... seek funding where necessary
- 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 stimulate 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 libraries and archives.
- Minerva Working Group 6. 2004. Good practices handbook
- Page-Shipp, R.J., 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>

Contact details

Martie van Deventer

Portfolio Manager: Information Services

Council for Scientific and Industrial Research (CSIR)

South Africa

e-Mail: mvandeve@csir.co.za

Tel: +27 12 841 3278