Vol 14 No 3 - September 2003

# SADCO

# SADSO

## SADCO's new web site

The URL of the SADCO web site (home page) was recently modified. Please remember that it is now

http://sadco.csir.co.za/

This also provides a link to the Inventory (see Article p 2).

Copies of the Newsletter are also placed on the web site.

Also in this Newsletter:

- Example of access to SADCO's Inventory p2 a few clicks away.
- Extraction speed of data improved considerably through hardware and software changes p 4 with more than 2Gb RAM and 6 processors, the hardware within which SADCO operates has received a considerable boost.

Southern African Data Centre for Oceanography
P O Box 320, Stellenbosch 7599
South Africa

Email: mgrundli@csir.co.za

Website: http://sadco.csir.co.za/

*SADCO is sponsored by ...*

Department of Environmental Affairs & Tourism
SA Navy
CSIR Environmentek
NRF (SA Universities)
Namibian Ministry for Fisheries & Marine Resources

ISSN 1017-2351

# Additions to the Inventory

It was reported in previous Newsletters that SADCO has added two significant sets to its Inventory:

- ***Current meters.*** These are deployments of most (all?) current meters around the southern African coast. The data originated from Marine and Coastal Management, the CSIR, and foreign research cruises. A total of 280 deployments, varying between weeks and months in duration, have been added to both the Inventory as well as the database.
- ***Automatic Weather Stations:*** a number of deployments were added, but they are located at 8 sites, mainly off the west coast.

**Example of access**

To obtain some insight into the Inventory of this data, access the SADCO Inventory at

http://morph.csir.co.za/sadco1/sadinv

or via the new SADCO home page

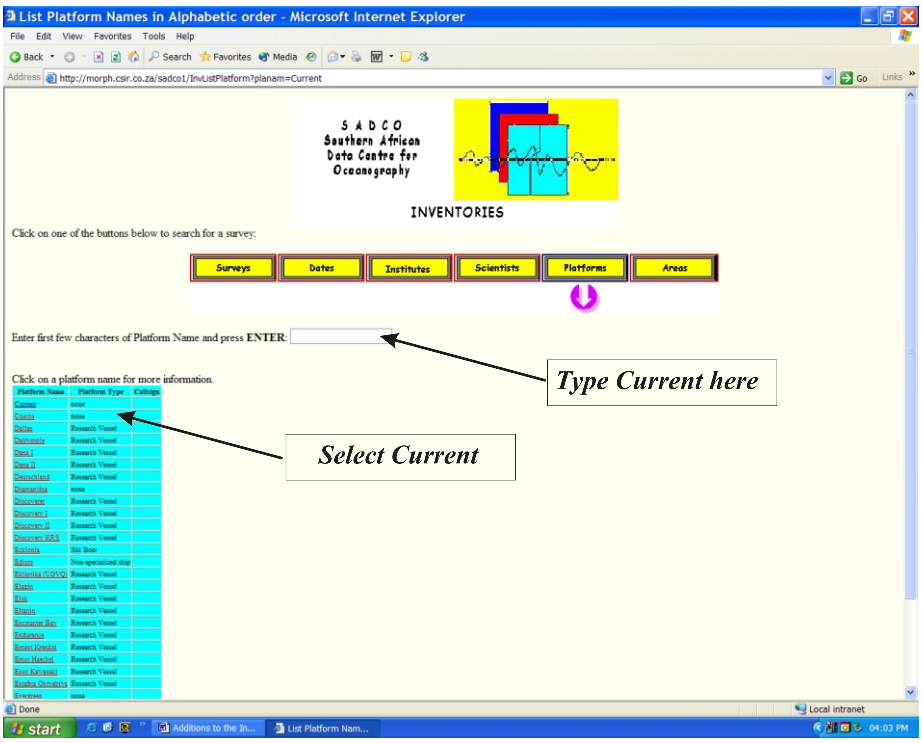
http://sadco.csir.co.za/

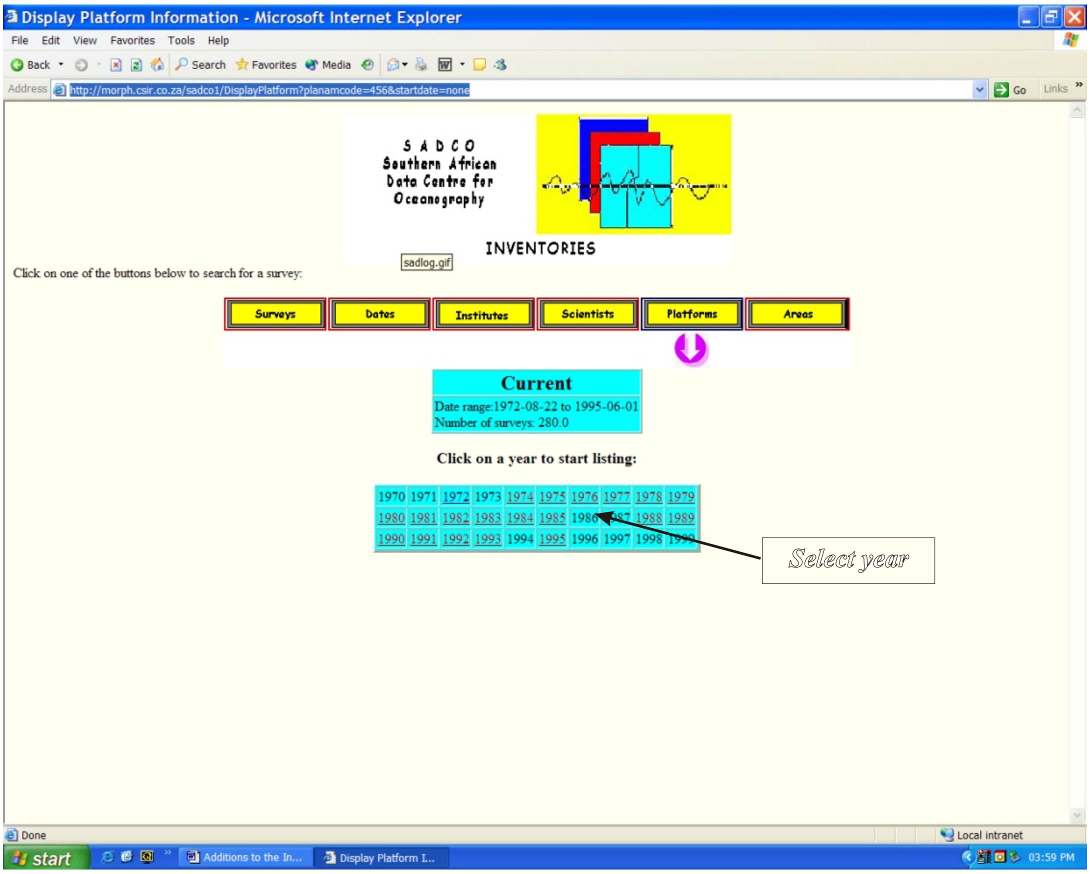
Select the *"Platform"* option, and type in (the first few letters of) *Current* or *Weather* into the drop-down box. Then select Currents, or Weather, from the menu.

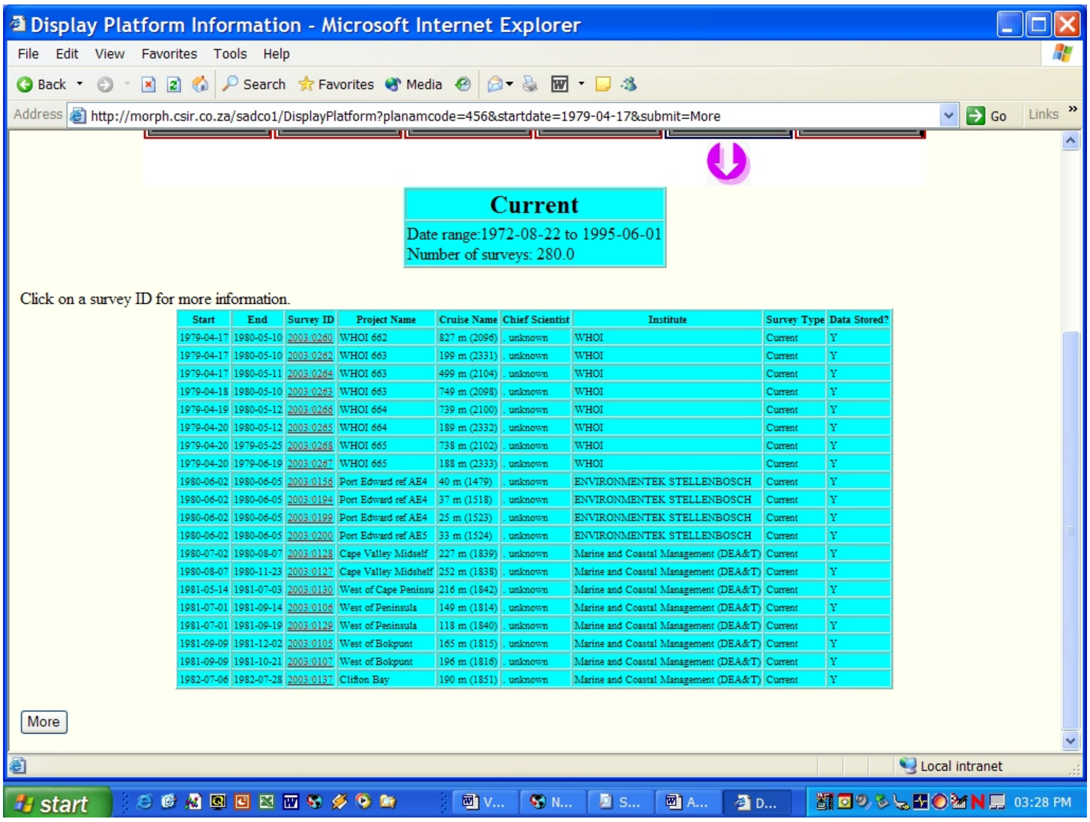
Select a year, and this will bring the inventory of all the current or weather deployments onto the screen.

*Note: We presently do not (yet) have a chart with the position indicated, but it is hoped that, at some stage, users will be able to see a screen chart, with the deployments positions indicated. This will allow a zoom-able method to determine the exact deployment required.*

It will be noticed that there are quite a number of Current meter deployments, but only few of Weather (in spite of what we had reported before on the number of Weather deployments). The reason for this is that weather stations are continued for some time, and although there are several intermediate redeployments, the data is stored under the same reference.






| Start | End | Survey ID | Project Name | Cruise Name | Chief Scientist | Institute | Survey Type | Data Stored? |
|---|---|---|---|---|---|---|---|---|
| 1979-04-17 | 1980-05-10 | 2003/0260 | WHOI 662 | 827 m (2096) | . unknown | WHOI | Current | Y |
| 1979-04-17 | 1980-05-10 | 2003/0262 | WHOI 663 | 199 m (2331) | . unknown | WHOI | Current | Y |
| 1979-04-17 | 1980-05-11 | 2003/0264 | WHOI 663 | 499 m (2104) | . unknown | WHOI | Current | Y |
| 1979-04-18 | 1980-05-10 | 2003/0263 | WHOI 663 | 749 m (2098) | . unknown | WHOI | Current | Y |
| 1979-04-19 | 1980-05-12 | 2003/0266 | WHOI 664 | 739 m (2100) | . unknown | WHOI | Current | Y |
| 1979-04-20 | 1980-05-12 | 2003/0265 | WHOI 664 | 189 m (2332) | . unknown | WHOI | Current | Y |
| 1979-04-20 | 1979-05-25 | 2003/0268 | WHOI 665 | 738 m (2102) | . unknown | WHOI | Current | Y |
| 1979-04-20 | 1979-06-19 | 2003/0267 | WHOI 665 | 188 m (2333) | . unknown | WHOI | Current | Y |
| 1980-06-02 | 1980-06-05 | 2003/0156 | Port Edward ref AE4 | 40 m (1479) | . unknown | ENVIRONMENTEK STELLENBOSCH | Current | Y |
| 1980-06-02 | 1980-06-05 | 2003/0194 | Port Edward ref AE4 | 37 m (1518) | . unknown | ENVIRONMENTEK STELLENBOSCH | Current | Y |
| 1980-06-02 | 1980-06-05 | 2003/0199 | Port Edward ref AE4 | 25 m (1523) | . unknown | ENVIRONMENTEK STELLENBOSCH | Current | Y |
| 1980-06-02 | 1980-06-05 | 2003/0200 | Port Edward ref AE5 | 33 m (1524) | . unknown | ENVIRONMENTEK STELLENBOSCH | Current | Y |
| 1980-07-02 | 1980-08-07 | 2003/0128 | Cape Valley Midself | 227 m (1839) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |
| 1980-08-07 | 1980-11-23 | 2003/0127 | Cape Valley Midshelf | 252 m (1838) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |
| 1981-05-14 | 1981-07-03 | 2003/0130 | West of Cape Peninsu | 216 m (1842) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |
| 1981-07-01 | 1981-09-14 | 2003/0106 | West of Peninsula | 149 m (1814) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |
| 1981-07-01 | 1981-09-19 | 2003/0129 | West of Peninsula | 118 m (1840) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |
| 1981-09-09 | 1981-12-02 | 2003/0105 | West of Bokpunt | 165 m (1815) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |
| 1981-09-09 | 1981-10-21 | 2003/0107 | West of Bokpunt | 196 m (1816) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |
| 1982-07-06 | 1982-07-28 | 2003/0137 | Clifton Bay | 190 m (1851) | . unknown | Marine and Coastal Management (DEA&T) | Current | Y |

# Some hardware and software improvements in SADCO

Almost 10 years ago, SADCO moved its databases from a large mainframe located in Pretoria (about 1500 km away from Stellenbosch, where SADCO is situated), to a PC in the same building.

This was quite a revolutionary and audacious move at the time. Moving databases onto PCs was unheard of. The impact on the staff working on the database was very beneficial: now they could control virtually all aspects of the hardware and software (software upgrades, hardware modifications, choice of database management system (DBMS), frequency and timing of backups and maintenance, etc). All this improved the efficiency by an order of magnitude, and more than made up for a reduction in the computing speed.

Later (around 2000) a choice was made to move to ORACLE, and proceed with a full web-enablement of the database.

The schematic arrangement of the flow of data is illustrated in the attached diagram (Configuration A in the enclosed figure). Here, all processes were dependant upon the ability of the Sun Workstation (MORPH), to handle the items. MORPH's memory (128 Mb) and its two processors, were stumbling blocks in the processing speed.

Because of the hardware constraints, and because of inherent slowness of the web procedures, access to the databases has been slowed down somewhat (compared to the previous, web-free system). This was not unexpected, but ways to speed up the process of extracting and manipulating the data were nevertheless sought.

## Improvements

The improvements were tested on the VOS (voluntary Observing Ships) database only. The speeding up process has thus far involved modifications on two fronts:

***(a) Hardware***

- The operations were split between MORPH (a Sun W/S), and FRED, a Sun workstation with 4 processors (see Configuration B).
- FRED was upgraded to 2 Gbyte of RAM, and MORPH to 600 Mbyte.

The result of these modifications was that the extraction process was accelerated, and that a typical extraction would be about twice as fast (i.e. take 50% of the time previously required). This is shown in the Table, by comparing the extraction times between Configuration A and Configuration B for various combinations of area and periods.

***(b) Software***

A submitted request for a given area and period is automatically modified into smaller "packets" (typically 1 x 1 deg and 1 year), to avoid memory overload during the data transfer process. Now that FRED has acquired larger memory, this risk was significantly reduced.

To improve the response time even further, Ursula therefore ran a number of trials where the size of the extraction "packets" was modified (e.g. making the spatial blocks smaller or larger, or the time "blocks" smaller or larger). Some trials were also run at times when the respective computers are under-utilised (e.g. after hours).

Six trials are shown in the Table (A - F), each having a different area/period combination ranging from a 5ºx5º (lat/long) block for one year, to a 30º x 30º for 30 years.

The Number of records extracted varied from about 2500 to about 360 000.

The "%" columns indicate the time relative to that of Configuration A.

The last two columns are the times taken after the area "packets" had been optimised, and after the **area and time** "packets" had been optimised.

The end result (Configuration B + area + time) has turned out to be very useful:

- The extraction process for relatively small areas and limited time coverage, was halved again (relative to Configuration B) and thus ultimately reduced to 23% of Configuration A.
- The improvement was most remarkable for large areas (e.g. 30 x 30 deg) and long periods (say, 30 years) that now take as little as 7% of the time for Configuration A. The latter, optimal improvement was achieved after hours.

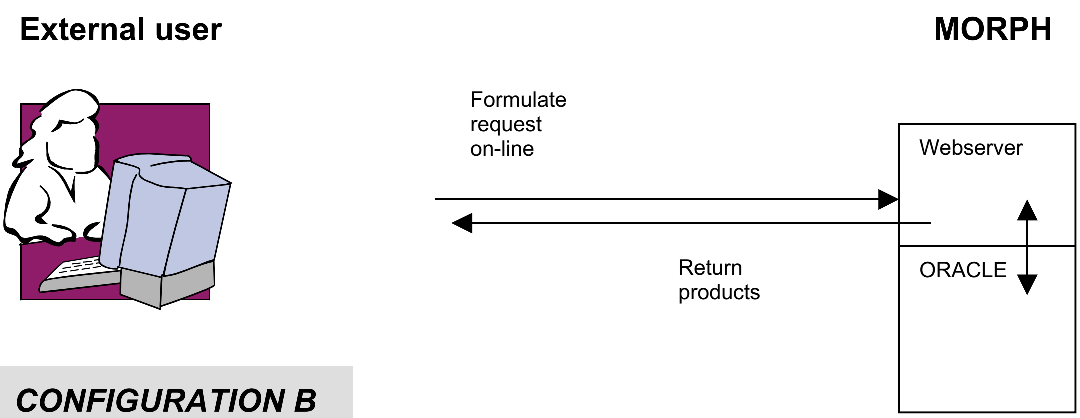
*Table: Extraction times for various space/time combinations, showing improvements due to hardware upgrade (Configuration B), and optimisation of the area and time extraction procedures, relative to the original configuration (Configuration A).*

| ID | Area (lat/long) | Years | Records | Extraction time | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Config A | | | | Config B | | | | Config B + area | | | | Config B +area+time | | | |
| | | | | h | m | s | % | h | m | s | % | h | m | s | % | h | m | s | % |
| A | 5X5 | 1 | 2566 | | 1 | 36 | **100%** | | | 47 | **49%** | | | 22 | **23%** | | | 22 | **23%** |
| B | 5X5 | 10 | 24252 | | 13 | 37 | **100%** | | 7 | 10 | **53%** | | 2 | 57 | **22%** | | 2 | 32 | **19%** |
| C | 10X10 | 10 | 47572 | | 28 | 25 | **100%** | | 16 | 17 | **57%** | | 5 | 35 | **20%** | | 4 | 18 | **15%** |
| D | 10X10 | 30 | 92029 | 1 | 6 | 4 | **100%** | | 44 | 47 | **68%** | | 12 | 49 | **19%** | | 8 | 58 | **14%** |
| E | 20X20 | 10 | 184527 | 3 | 2 | 59 | **100%** | | 58 | 50 | **32%** | | 20 | 53 | **11%** | | 16 | 19 | **9%** |
| F | 30x30 | 30 | 363627 | 7 | 23 | 39 | **100%** | 2 | 43 | 32 | **37%** | | 41 | 2 | **9%** | | 31 | 19 | **7%** |

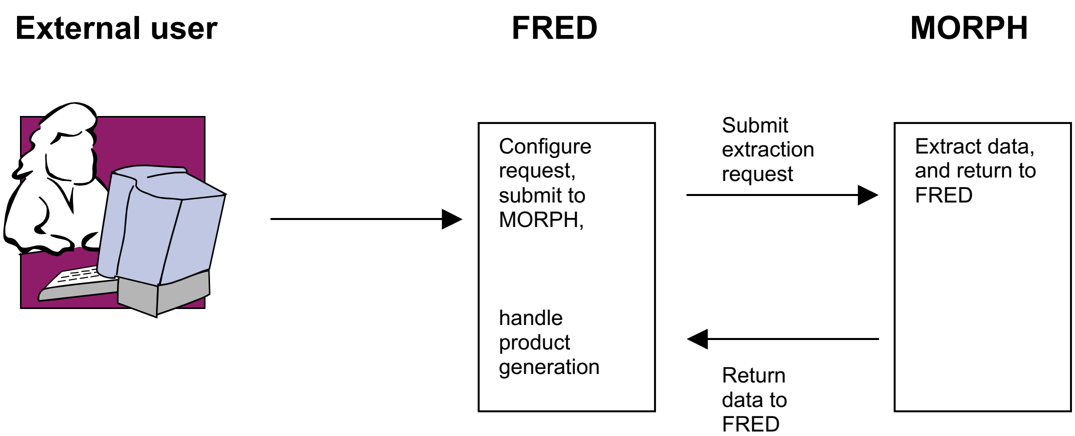
## Generation of products

The time required to generate products has not been tested (yet), and has never taken very long. Nevertheless, it is expected that the improvement will also be noticeable, considering that the products are now being generated on FRED, which has been given the 2 Gbyte of RAM and already has 4 processors.

### CONFIGURATION A




### CONFIGURATION B



# OCEANPortal - a gateway to information

In a previous Newsletter we referred readers to the *OceanPortal*, the website of the IOC/IODE. This portal is expanding continuously, and a host of new items have been added.

The following are useful items to consider:

**Data sources**

**Scientists**: A directory is maintained of marine scientists (obviously only those that have registered themselves). A call was made by SADCO for southern African marine scientists to enter their names on this directory.

**Vessels**

- The dimensions of a specific vessel (e.g. the Algoa, or Africana)?
- The schedule of the "international" vessels?

**Programmes**; There is a host of web pages of projects and programmes

- More information on the Agulhas Large Marine Ecosystem?

**Software**

- The potential temperature, or oxygen saturation, for a given set of parameters?

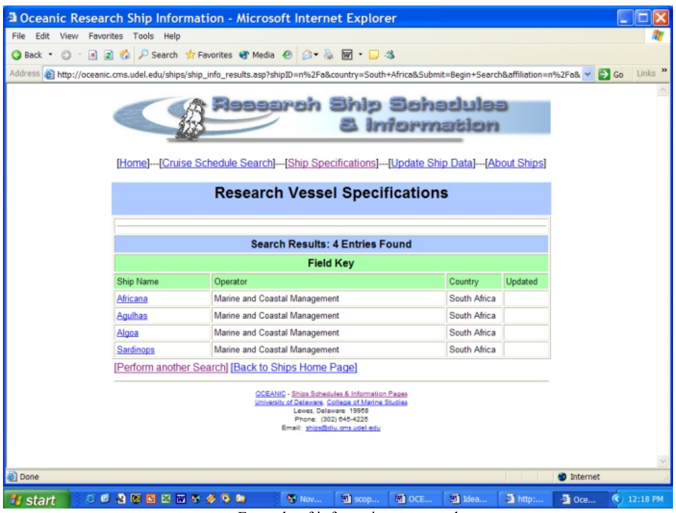
**Organisations**: International, national, non-governmental, etc

**Meetings**

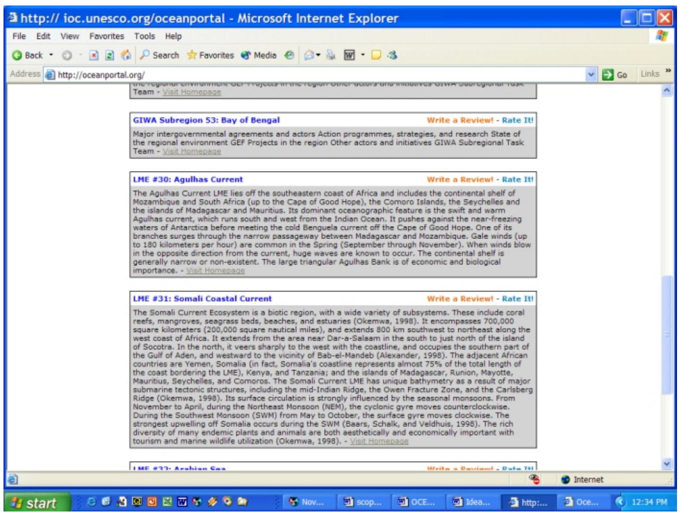
The URL of the site is:

***http://oceanportal.org***


The OceanPortal home page




Example of information on vessels

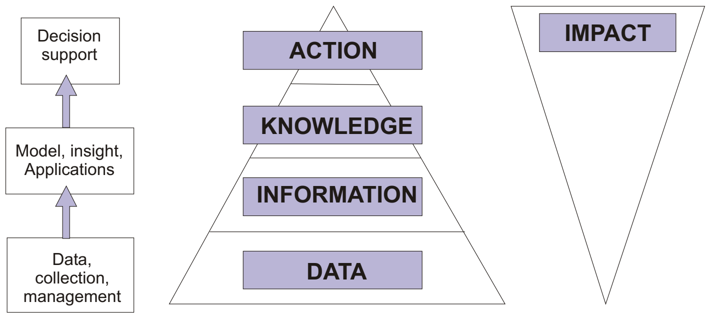



Link to the Agulhas Large Marine Ecosystem

# *What's all this fuss about historic (environmental) data? (or, Data Management 101) (Part II)*

*In the previous Newsletter it was indicated that:*

- *We need data to understand and live in harmony with our environment.*
- *We need large amounts of data, because only in this way can we generate knowledge, and only with sufficient knowledge can we make appropriate impacts (see diagram below).*
- *To look further into the future (i.e. predict) we often need to look further into the past (and spatially further afield).*




## *And the future?*

With time, man's desire to expand his understanding of nature has led to a growing urgency and effort in collecting environmental data. This process has been fuelled by increased technology, especially in the field of space oceanography. The expanding scope of satellite sensors, and their ability to transmit vast data sets on a 24-hour basis, is almost frightening, and the immense data sets that are generated must be continuously challenging the capabilities of storage and extraction media.

## *So, is our "Southern African" data getting more?*

This is obviously not easily quantifiable in a general sense, but should first be addressed for an individual parameter (e.g. number of temperature measurements), or instrument type (e.g. are more hydrographic stations being occupied per year than 10 or 20 years ago?). Let us have a rough check on some aspects.

**(a) Profile data**

In terms of physical data in the water column, the **advent of CTDs** in the late 1970's and early 1980's, with their fast temperature and salinity (conductivity) sensors, made a significant impact on the amount of data that could be obtained from a single cast. In addition, there are climate programmes that use ships-of-opportunity, where significant numbers of XBT's are deployed around our coast every year. The feeling is therefore that the amount of data has <u>increased</u>.

**(b) Current meters**

**Moored current meters** really started in the early 1970's, and the number of deployments seems to have decreased in the 1990's. But then the technology kicked into a higher gear with the advent of the ADCP, and a single instrument can record velocity simultaneously at a number of levels (or "bins"). The feeling is nevertheless that the number of current meter deployments have <u>decreased</u>.

**(c) Surface temperature**

While **coastal surf temperature** measurements used to be done by hand (often with a Crawford bucket), thereby limiting the amount of data to one or two observations per site per day, continuously recording temperature chips, cheap and easily deployable, record sea temperature around the clock, and this data seems to <u>increase</u>.

And then there is **satellite data**, the amounts of which are staggering and increasing every year as more satellites measure more parameters (not only temperature, but colour, topography, wind speed, waves, etc).

**(d) Ship-borne current velocity**

In the past, research vessels could record **current velocity on station** and only in proximity of the coast (because of coastal tracking issues), and with a large manpower complement, vessels like the *Africana* and *Algoa* can record vertical profiles on and off station, for the full time while at sea. So, spatial current velocity information has increased significantly.

**(e) Coastal stations**

The number of **automatic coastal (weather, wave) stations** seem to remain more or less the same.

**(f) Model data**

An aspect not often considered in conventional data management (such as SADCO) is model data. Models can provide vast amounts of data (e.g. temperature, waves, currents, etc) at places and times not normally covered by instrument deployments. These model outputs are approximating the accuracy of measurements, and can lead to equally valid conclusions. If this data would also be considered for data management, there would be a significant increase in the amounts of data.

***So, data seems to be increasing all the time. Why is mankind's drive for more scientific data so insatiable?***

I think the following could be some of the reasons:

- Scientific and sustainability questions are becoming so complex, that historic data, with its limited (initial) purpose, is unable to provide the (full) answer. Because of the fine line between various decision options, the new data has to be **more accurate**, **more reliable**, **more universal** than decades ago. This leads to larger, longer experiments, with high-frequency measurements and more global relevance and impact.
- It seems to have become cheaper to collect data using automated, low-manpower systems that collect data around the clock.
- There also seems to be the science equivalent of Parkinson's Law, for want of a better descriptor. The concept of "have vessel will travel", or "have instrument, will deploy", is fairly rife (I recall this was mentioned by Vere Shannon in a keynote address at one of the SAMSS symposia). Young researchers want to collect their "own" data, sometimes not even bothering to check whether suitable, freely available data already exists. In addition, funding streams urge scientists to apply for grants. The conclusion is that scientists will apparently be spurred to continue (and increase) data collection *ad infinitum*.

The final question concerns information about data (i.e. metadata): **Why don't we seem to know about the existence of historic (environmental) data** (in stead of finding out only much later)?

- In the haste to collect new data (more exciting), older data is often disregarded or even "swept under the carpet". I believe that young scientists should be exposed to existing data sets and their usefulness.
- The rapidity with which individual measurements are collected, also holds for cruises and experiments, i.e. they are executed so quickly that we lose track of whole data sets. So scientist often find it difficult to keep up with submitting collected data to a data storage system, or even have the time to make data available to fellow researchers. I believe data centres can provide well needed support to relieve the pressures off scientists, by keeping the data available for other researchers. This will also minimise duplication in data collection efforts.

## *Conclusion*

It seems that the need for data is growing, that data amounts are expanding. To keep the data viable the requirement for systematic data management will continue.



Layout & Design: Magdel van der Merwe
CSIR