


This is the story of what was done, and some lessons learnt as we bumped along on a very rough road.

In the process of our work in mapping vegetation, modelling to investigate the possible impacts of climate change on species, red list assessments, impact assessments, taxonomic research, invasive species research, contributing to the development of environmental policy, managing world-famous botanical gardens, developing human capital, and many other achievements, we had to mine data and develop ways to achieve our outcomes.



Banging our heads was worth it to help making it feasible for many users to travel a smooth data sharing road.

We work with and provide information for many colleagues and partners who can benefit from the lessons that we learned so that they do not need to repeat some of our mistakes.



A big focus of work at SANBI involves linking biodiversity assessment and monitoring with good biodiversity information management –this involves moving from having different toolboxes in diverse workshops and putting them into a single toolbox in a single workshop.

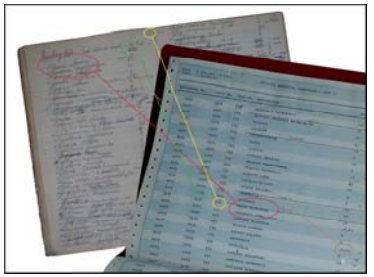
Well, that is a big dream that we have.

A good example of this experience is that over the last 25 years multiple data sources were combined.

Observation data from vegetation survey (plot data and Acocks samples), Protea Atlas, Millennium Seedbank, SAPIA (Southern African Plant Invaders Atlas), CREW (Custodians of Rare and Endangered Wildflowers) and other plant species distribution data were combined with herbarium specimen data in an MS Access database.

We wanted to integrate data to make it more powerful, correct errors because using data increases the likelihood of errors being detected.

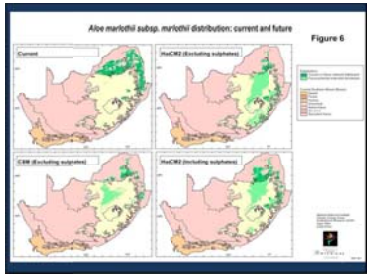




September 1993 was about when the start of transcribing Acocks field notes commenced.

There are 44 field note books, nearly 400 000 species occurrence records with locality to about 1 km precision, habitat and abundance given for many of the records.



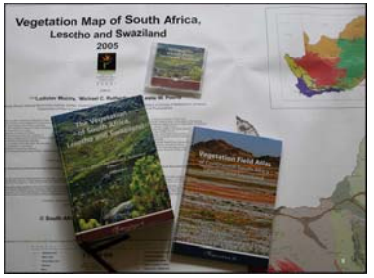



This opened the way to bioclimatic modelling for the South Africa Country Study for Climate Change report. Climate data at a scale of about 1.5 km, species localities at a similar precision.

Specimen data was at QDS, a 1:50 000 map sheet, approximately 18 km radius. Many specimens were georeferenced where possible to determine their coordinates to add to the data with precise localities.

Modelling the species climate envelope and how it might expand or shrink with climate change, predicting the survival or extinction of the species.






Enter Plot data as we looked at classifying the 450 vegetation types of South Africa, Lesotho and Swaziland.

Protea Atlas data, Millennium Seedbank, CREW (Custodians of Rare and Endangered Wildflowers), SAPIA (Southern African Plant Invaders Atlas) and other observation and specimen data were combined to look at the validity of claimed endemics for vegetation types and other products requiring species distribution data.

By combining data you have the advantage of discovering things that you have easily at your disposal to improve the quality of the data.

Work to capture the verbatim text without delay –do not wait to georeference or interpret.

Then use the larger dataset to facilitate interpretation and refining the data. Do batch corrections such as dates that are difficult to read, collector or place names, georeferencing by collector itinerary.




Plant species distribution data were served to the sector on a request by request basis – but that was still done by one operator and was useful to very few users.

The database with all of the combined data was useful for extracting the data for answering requests, and for species research, but only by a few users.

Tools from many toolboxes were combined into a single toolbox, but still used by few individuals.


Examples of distribution sources complementing specimen data.

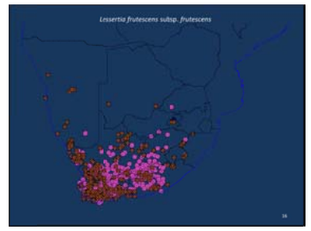

It is very important to have a feedback ID that will link back to the species record ID in the original data source. This will facilitate feedback and quality control.

CREW

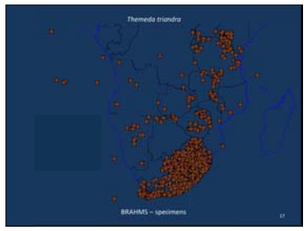


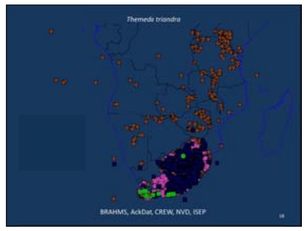



All of the data combined. Another example of combining data sources. Note the localities in the sea –low hanging fruit for data cleaning!



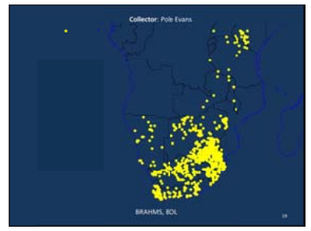


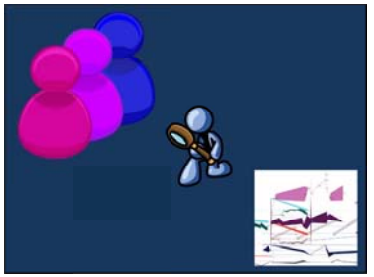




Observation or other datasets provide additional temporal and spatial data for monitoring, sometimes making it possible to compare rates of spread or decline of invasives or populations.



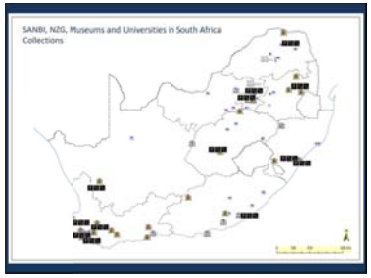



Records for a collector. This facilitates tracing where the collector was close to the time of collecting a specific record to improve accuracy in georeferencing, especially when there are multiple matches to a place name.



Recent work has been done to take all of these different data sources from this local database and put it into BODATSA (Botanical Database of Southern Africa) to make it accessible online in the global workshop. The data will be combined into National Biodiversity Information System (NBIS) where the user should be able to interrogate the data and extract what is required. The tools are being put into the global workshop where all users can use the tools themselves without the same amount of travelling to herbaria or museums, or loaning from institutions, and handling specimens or records that had previously been needed. Users anywhere in the world can use an internet connection to use the tools in the toolbox. But –there are still other tools that are in toolboxes in other workshops scattered around South Africa and the world.






We hope that all species distribution data from different institutions in southern Africa will in due course become available in this single online NBIS (National Biodiversity Information System) portal in the internet 'cloud'.

SANBI, NZG, Museums and Universities in South Africa. Living plants collections are not all integrated. There are also collections in SANParks, official and private nature reserves and conservancies, provincial nature conservation agencies, military bases, agricultural research stations, ESKOM, private consultancies, NGOs, and various other places scattered around the country.

In ecosystems there are producers and consumers. But in fact, each consumer also produces (even if only through being decomposed after death) and producers consume nutrients to survive and produce. Let each of us follow this model and contribute all that we can to our community and not be consumers and fail to produce.

'Together we can do more.' Yonah Seleti (BIMF/FBIP 2017)




Besides data scattered around South Africa, there are data for South Africa that are held in museums and herbaria in many parts of the world.

About four years ago, GBIF had data for South Africa in 30 different countries. South Africa also has data about specimens from 194 countries in its herbaria and museums.

It would be wonderful to have a one-stop-shop for all data, easily and efficiently extractable by any individual anywhere in the world with little need for extensive technological prowess. That is a big dream, and we are moving in that direction.

So –we have done great things –the challenge now lies in our doing even more and doing it better.


**Thank you to**

The taxpayers for making it possible for us to do our work.

Volunteers for transcribing and georeferencing specimens and records

Staff for their expertise and effort

Early biodiversity explorers for recording all that they did, often under difficult conditions.

You as the biodiversity explorers of the present and future

**Thanks for being here**

I enjoy sharing things –I hope you were not bored out of your wits….