BIMF-FBIP FORUM REPORT 2017

Joint Forum
Biodiversity Information Management
& Foundational Biodiversity Information Programme

14–17 August 2017
Salt Rock Hotel and Beach Resort, Durban
## CONTENTS

**SANBI CEO’s message** .................................................................................................................. 4

DST opening address highlights: Dr Yonah Seleti .................................................................................. 5

Keynote address ................................................................................................................................... 6

**Data-science-policy interface** ........................................................................................................ 8

  - Bridging the data to policy chasm – Selwyn Willoughby (Refleqt) .............................................. 8
  - SDG’s and implications for SA – Wadzi Mandivenyi (DEA) ....................................................... 9
  - Mining microbial diversity – Gwynneth Matcher (RU) ............................................................... 9

**FBIP** ............................................................................................................................................... 12

  - FBIP funding approach & themes – Michelle Hamer (SANBI) ................................................... 12
  - Key elements for successful proposals and analyses of funded projects – Lita Pauw (SANBI) .... 13
  - SeaKeys: Lessons from the first large FBIP project – Wayne Florence (Iziko) ......................... 14
  - Biogaps: Overview of the project: challenges & lessons learnt - Theresa Sethusa (SANBI) ..... 16

**The natural science collections facility** ............................................................................................ 19

  - The Natural Science Collections Facility: Specimen Data Objectives – Michelle Hamer (SANBI) .... 19
  - Update of the digitizing of the Killick Herbarium: feedback on the vision, progress, and challenges – Boyd Escott (Ezemvelo KZN Wildlife) ......................................................................................... 20
  - The National Collections of Fungi: The portal for phytopathogenic fungi from SA – Adriana Jacobs-Venter (ARC) ................................................................................................................................. 21

**Use, Application & Impact of Biodiversity Data** ............................................................................. 24

  - Biodiversity informatics: Meeting sustainable development challenges for fisheries in the face of climate change in southern Africa – Fatima Parker-Allie .................................................. 24
  - Towards expanding the SA Rhizobium Culture Collection as a genetic resource and its application in sustainable agriculture – Ahmed Idris Hassen (ARC) ................................................................. 26
  - Bridging the gap: Making amphibian biodiversity data relatable in SA – Fortunate Phaka (UKZN) ... 27

**Assessments, Monitoring Frameworks and Indicator Development** ............................................... 29

  - The Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES): the importance and relevance of data in the African Regional Assessment – Anicia Malebajaoa Maoela .................. 29
  - Data needs for high-level Biodiversity indicators: Is there a gap? – Jeff Manuel (SANBI) ........... 32
  - A quarter of a century of taking databases from paper to cloud – Les Powrie (SANBI) ................. 33
  - South Africa’s first national status report on biological invasions – Heather Terrapon (SANBI) ...... 35

**Data sharing between Sanbi and partners & information systems and Architecture** ..................... 36

  - FBIP data sharing - Lita Pauw (SANBI) ..................................................................................... 36
  - SANBI’s data sharing agreement - key principles – Fatima Parker-Allie (SANBI) ......................... 36

**Information systems & architecture** ................................................................................................. 37

  - Update on the National Biodiversity Information System (NBIS) – Brenda Daly (SANBI) ............. 37

**Advances in data science & information management** .................................................................... 39

  - Power to the people: Citizen Science setting collections data free (Vince Smith, NHM London) .... 39
Virtual safaris mobilizing biodiversity information in African countries (Les Powrie, SANBI) ........... 40
Toward next generation biodiversity research (Morne du Plessis, NZG)........................................... 42
The NZG Biobank information & information systems: Analysis and improvement (Kim Labuschagne, NZG) ................................................................................................................................. 43
Turning the Animal Demography Unit databases inside out with modern application architectures (Ian Engelbrecht, University of Pretoria).............................................................................. 44
A system to automatically generate annotated checklists (Willem Coetzer, SAIAB) .................. 44
Validating records georeferenced for the BioGaps Project (Willem Coetzer, SAIAB) .................. 45
The Namibia Biodiversity Database – content, capabilities, and current status (John Irish) ........ 45
Existing data management systems for species data ............................................................................. 47
Building the biodiversity informatics africa agenda ............................................................................. 48
The SANBI Regional Engagement Strategy – Fatima Parker-Allie (SANBI) ................................. 48
The African Biodiversity Challenge: An incentive-driven approach to mobilizing biodiversity data - Matthew Child (SANBI) .............................................................................................................. 49
Wildlife biodiversity for sustainable socio-economic development - Kenneth Uiseb (Ministry of Environment and Tourism, Namibia) ........................................................................................................ 50
An online resource for the flora of the western Indian Ocean region – Benny Bytebier (UKZN) ....... 51
Microbes are biodiversity too: the African Soil Microbiology project – Don Cowan (UP) .............. 52
Closing................................................................................................................................................... 53
SUMMARY

The conference theme dealt broadly with natural science collections data management and efforts to build momentum in digitisation. There was also a focus on an enabling environment as well as how the Department of Environmental Affairs (DEA) is linking this work into international reporting (policy interventions). And how we as a community are linking into the rest of Africa. It was heartening to see much engagement in the social media space as well as networking happening. The conference was epitomised by “together we can do more” as mentioned by Dr Yonah Seleti. – Jeff Manuel (Director: Biodiversity Information and Planning)
SANBI is proud to host the 2nd Joint Biodiversity Information Management & Foundational Biodiversity Information Programme Forum. SANBI has played a key role in convening the Biodiversity Information Management community since 2007, in an endeavor to respond to its mandate and to harmonise biodiversity information sharing across institutions. SANBI has also led the DST-funded Foundational Biodiversity Information Programme since 2013. This programme allocates funds through the NRF for the generation and mobilisation of foundational biodiversity information to address key national needs for data and knowledge.

The hosting of the 2017 joint forum provides an opportunity to share information, advancements and opportunities in the area of Biodiversity Information Management. Through this platform, we want to bring about a holistic understanding of institutional needs, priorities and capacity relating to biodiversity data across the sector. An understanding of these needs between SANBI and its partners will support the effective generation, management and use of data and its uptake into science and policy frameworks, and will contribute to a sustainable development agenda.

Globally the value of data is increasingly being recognised. The DST has decided to specifically invest in big data in science, in other words the gathering and storing of large datasets for analysis. SANBI co-ordinates the FBIP and has recently been appointed to co-ordinate the Natural Science Collection Facility. Both of these will mobilise data and information that can be used in analyses to support research, policy development and decision-making.

SANBI also hosts the SANBI-GBIF Node and to date more than 20 million biodiversity records have been mobilised through our South African Partners connected to the SANBI-GBIF Node. Through our governmental mandates, we strive to ensure that data supports a strong research and developmental agenda, and in this matter demonstrates the data value chain. SANBI-GBIF is also instrumental in leading the biodiversity informatics agenda on the continent.

We also want to take this opportunity to highlight existing regional efforts, and explore opportunities to build the Biodiversity Informatics Africa agenda. This is an exciting time for our megadiverse continent and the Africa agenda.

We have an exciting agenda planned with a number of key sessions including:
- National Agenda’s in Support of a Data-Science-Policy Interface, the Regional Biodiversity Information Agenda, Foundational Biodiversity Information Programme, Advances in Data Science, Information Systems and Data Management.
- We also have a Postgraduate Forum planned on the 16th August, as well as a SANBI-GBIF Georeferencing training event and a workshop on FBIP proposal writing on the 17th August.
DST opening address highlights: Dr Yonah Seleti

- Through the FBIP, government has enabled the sustainable use and responsible custodianship of biological resources.
- The DST partnership with SANBI has grown and it is a ‘long-term commitment’.
- The output datasets are critical for ecosystem mapping and other high level assessments.
- The programme has funded 82 projects to the value of R29m.
- South Africa’s biodiversity data scientists were challenged to think of developing the field in terms of scope and scale.
Keynote address

SPECIMENS TO PIXELS TO ACTIONS: CREATING, MANAGING, AND EXPLOITING DIGITAL COLLECTIONS

DR VINCE SMITH – NATURAL HISTORY MUSEUM, LONDON

Abstract:

Digital technologies are having a profound impact on how we manage, access and use natural science collections. Increasing reliance on digital services has the potential to generate efficiencies that can transform institutions and create opportunities to integrate our work, within and across national boundaries. This transformation comes at a cost as organisations create new roles, systems and policies to support digital representations of our collections. In this presentation, I will provide an overview of how London’s Natural History Museum’s (NHM’s) Digital Collections Programme is effecting this change, and how we are internationalising this work through European and global programmes. I will highlight some of the challenges within our institution, and some of the resource implications as we shift the intellectual business model of the NHM from one focused exclusively on physical collections, to one that places increasing emphasis on their digital surrogates.

Presentation summary:

There is a shift to make the Natural History Museum ‘more digital’. The institution’s ambition is to digitize 20 million records in 5 years. According to Smith this will not be achieved, but pilot digitization projects have yielded valuable lessons.

For a butterfly project Smith’s team has managed to digitize ~800k specimens with several million still in the pipeline. In the process they have progressed from a manual digitization process (at a time/cost expense of 2min/£1 per specimen) to a faster method using ‘computer vision’, which uses several cameras to extract data from labels.

But it’s not just about digitization, says Smith. “It is more of a cultural change, to build digitisation into your organisation you need to think about prioritisation, outreach (opening it up across the community), team organisation, etc.”

Establishing standards

According to Smith a lot of standards and ‘ways of doing things’ needed to be built for digitization, data curation, and outcomes. The biggest challenge was moving from ‘closed by default’ (people had to come to the museum to look at data) to ‘open by default’, where most of the data is open-access.

NHM data portal

A steady stream of data is flowing out of the NHM portal – ~2.3 billion records have been downloaded. The portal is being sustained and there is a drive to align people, projects, and publications around it. At the same time, measuring engagement is also important – what does 2.3 billion downloads mean?
Smith highlighted the benefits of business intelligence systems to see where the data is going, and tracking digital collection use.

GBIF’s ‘traffic light’ service is being used to flag problems with the data. “This can’t be fixed very easily but we can warn users where there might be problems,” Smith said.

**Parting message**

Biodiversity data industry players often think of their collections as their domain, whether individual or institutional. Smith sees it differently.

“This is one big global collection and collectively if we can start seeing our collections not as the preserve of individual institutions, but one big global scientific infrastructure, then there is an enormous amount of research that we can do for the betterment of society.”

**Questions and Discussion**

**Question:** Are you linking into the phylogenetic/genetic/barcode information?

**Answer:** A little bit - post digitisation. We often mine the GENBANK/biodiversity heritage library to link it back up to our data. We do a little bit pre-digitisation. Not as integrated as we should be, and this is a critical deficit.

**Question:** There is confusion around ‘alphabet soup’ of institutions out there – what do you think of the value of GBIF – there are a lot of other players out there?

**Answer:** I’m a huge advocate of GBIF. If we didn’t have it, the first thing we’d do is reinvent it. In the northern hemisphere we struggle to get our government interested in GBIF. And yet, the third biggest user of GBIF are UK scientists! But GBIF is a fundamental part of global data sharing.

**Question:** I see a lot of the data comes from colonial past. What is your policy around data sharing of countries of origin?

**Answer:** We make all of our data globally open and accessible. We only hold back a tiny amount of our data. What’s more important is that the nation of origin is not as involved as it should be in the prioritisation of our collections digitisation. We need to do more work to create MoUs with our founding countries so we can take into account their needs.
DATA-SCIENCE-POLICY INTERFACE
UNDERSTANDING THE NATIONAL AGENDAS IN SUPPORT OF SUSTAINABLE DEVELOPMENT

Bridging the data to policy chasm – Selwyn Willoughby (Refleqt)

Abstract:
How do we make sure that policy decisions are based on sound evidence? How do we take our data and transform it into information that is used in an accountable, responsible and transparent manner? Also, how do we ensure that the data is used in the first place? These questions are increasingly being asked as the gulf between the data and policy extremes is difficult to navigate and fraught with many, often foreseen, hazards. Earlier this year the inaugural United Nations World Data (UNWDF) Forum took place in Cape Town, South Africa, to address how data can be used to achieve the objectives of the 2030 Agenda for Sustainable Development. This acknowledgement of the data to policy chasm resulted in the drafting of the Cape Town Global Action Plan for Sustainable Development Data. The plan not only recognises the challenges of the data to policy continuum, but also the underlying challenges of collecting, managing and disseminating data. The plan itself can be seen as creating an enabling environment and a pathway for data to be policy relevant. Biodiversity data is an integral part of the sustainable development data mix. The challenge is therefore to ensure that biodiversity data is accounted for and used in the national data accounting system to report against the progress of addressing the Sustainable Development Goals.

Presentation Summary
Willoughby attended the UN world data forum in Cape Town – the first time that a global platform for data was held. One of the most important messages that came out of the high level gathering, according to Willoughby, was that data must be evidence-based. “All decisions are becoming data driven,” he said.

“Capacity, capacity, capacity!” - Individual capacity (training the person) and institutional capacity (providing tools and resources) are vital. Willoughby said that the private sector is
willing ‘to come on board’ and that the mining industry was in possession of lots of biodiversity data. Willoughby’s presentation left delegates with several key questions:

- Do you have a single version of the truth?
- Who are the role-players? What are the contributions that people are making? I.e. developing infrastructure towards a purpose.

**SDG’s and implications for SA – Wadzi Mandivenyi (DEA)**

Mandivenyi recognized the gap between the policy and science domains. Out of this chasm the Sustainable Development Goals (SDGs) emerges as a management tool for the SA government. The formulation of the SDGs followed an integrated approach and represented an opportunity to ‘mainstream biodiversity’.

‘We are moving away from a system with broad goals towards a system driven by data,’ said Mandivenyi, who also listed the following requirements from data managers:

- Success in monitoring the SDGs will require a data revolution
- Harnessing new innovative sources of data and moving towards annual monitoring
- High quality data from all stakeholders
- Existing and new data systems will require continuous strengthening

We need to “domesticate” the SDGs –and align the National Development Plan with them. The link between biodiversity & poverty has been clearly articulated.

We need the data revolution, we need new high quality data. Weak data or not, we have to continue reporting. We need data managers to work out how we get data/indicators, how we clean it, and how we prioritize or weigh data.

**Mining microbial diversity – Gwynneth Matcher (RU)**

**Abstract:**

Ecosystems are a delicate and perfect balance between the biological community and their physical environment. Despite the fact that microorganisms are critical components of ecosystem functioning and are present in every ecological niche, microbes are an oft overlooked component of ecosystems. Metagenomics, which is the study of the genetic material of an entire population within a given environmental sample, has come to the fore in the study of microorganisms in recent years. This approach allows for rapid identification of diverse microbes and their functionality without limiting the study to a few isolated individuals. This powerful tool not only allows characterisation of microbiomes, but can also be applied to economically relevant research foci in the bioeconomy sector. This presentation will cover a brief background on microbes and their ecological and economic importance (e.g. natural products discovery, bioremediation, food industry, etc.) as well as the application of metagenomics for the mining of microbial diversity for the purposes of biodiversity surveys as well as bioeconomical applications.

**Summary:**

Microbes are awesome but the problem was working out what microbes actually are. Now genetic sequencing is used for the taxonomic hierarchy of microbes and bacteria. Scientists can sequence DNA from individual organisms, or they can do metagenomics – take a teaspoon of soil and sequence all the organisms included in it. You can then use an algorithm to cluster the genes together to group them into species. The microbes (especially in sea sponges) have huge applications – bioprospecting, and commercial potential. A huge amount of data is generated through sequencing microbes.
QUESTIONS AND DISCUSSION

Question to Selwyn Willoughby: One of the slides referred to a single authoritative version of the truth, considering all our data providers like GBIF, I’m beginning to question the concept. Do you have any reflections on this?

Selwyn Willoughby response:

- The issue revolves around when an issue is heard in court – what/whose information do you use?
- You have multiple ‘access points’, but you need a single version of authoritative data – who will deliver this data?
- For example if SANBI gives an opinion on an EIA or information, and at the same time DEA has individuals delivering information. There is a strong case for throwing out the decision because there is no authority to provide the information.
- Multiple versions is possible, but a final authority is needed to sign off and carry the process forward.

Tanya Abrahamse input:

- The ‘true/not true’ issue is ongoing – we need a schematic which states who are the contributors to information.
- [We need to consider] a Bill around indigenous knowledge. How do we view data driven from a Western perspective and data driven from another perspective? There has to be convergence at some point.
- On the topic of data use by policymakers, it is important to consider that they use [our] data in various other contexts, not only biodiversity.
- For example, we have been dealing with a mining license issue and we had comprehensive images of water, biodiversity, threatened species and redlists. But we came to the realization that we have no information about the people – how many people there are? What jobs do they do? Etc.
- If we are going to operate in the ‘data space’ we need to provide more ‘palatable’ data covering a 360° spectrum.
- We must include socioeconomic data into our biodiversity data…we can’t just talk about the physical environment and species anymore.

Question to Selwyn Willoughby: What is next for institutions in this space? Does the City of Cape Town have a model that the rest of SA could learn from?

Selwyn Willoughby response:

- StatsSA has an important role to play – how the data gets delivered through StatsSA.
- The issue of data is going to become more complicated because SA is signing the Cyber Security Bill into law.
- This will add many more layers of responsibility at organizational level. SANBI can be classified as ‘research infrastructure’ which makes it part of the ‘classified data’ sphere.

Question to Wadzi Mandivenyi: On the issue of data not being clear to policy-makers – why is it that the policy-makers are people that are not inclined to the subject?
Wadzi Mandivenyi response:

- All the policy-makers are scientists by training. They are up-to-date on what scientists are working on. It’s about keeping the interface aligned.
- Policy-makers practice a different kind of science. They are practitioners of science as opposed to generators of science.

Question to Gwynneth Matcher: What is the state of knowledge around the links between soil microorganisms and ecosystem function, crop production etc.? [It does not get as much attention as the charismatic megafauna and birds…]

Gwynneth Matcher response:

- We need more information – we know very little.
- At Rhodes University there is one research group…looking at the fungi associated with plants in the rhizosphere that help fix nitrogen and help the plants grow. They’ve actually commercialized a product based on the latter for use in the SA farming community.
- Prof Don Cowan is planning on doing a proper survey of the soil microbes and that is a longtime coming, because particularly for bacteria there is a massive knowledge gap. There are quite a number of several small studies, but nothing comprehensive.
FBIP & LESSONS LEARNT FROM LARGE FBIP INTEGRATED PROJECTS

FBIP funding approach & themes – Michelle Hamer (SANBI)

Abstract:

The Foundational Biodiversity Information Programme (FBIP) was established by the Department of Science & Technology (DST) in 2013 in recognition that sustainable use and management of South Africa’s biodiversity require a solid knowledge base and access to relevant information. However, a number of challenges have limited the use of the data and knowledge generated through research for decision-making. The mandate of the FBIP therefore is to fill the large gaps in our knowledge by means of a strategic approach and to unblock the value chain for foundational biodiversity data generation. The FBIP provides grants for the generation of knowledge related to documenting South Africa’s biodiversity, mobilisation of species occurrence or distribution data, generation of DNA barcode data that will allow identification of biological material, and compilation of descriptive information on species. The National Research Foundation (NRF) manages the proposal review and grant allocation process through a competitive process. SANBI is responsible for the implementation of the programme. The grants must result in the release of data to the FBIP/SANBI for archiving, integration, management and dissemination. This presentation will explain the FBIP funding approach and themes, outline the key elements for successful proposals and give an analysis of funded projects.

Presentation summary:

The FBIP seeks to address the issue of how information that is generated by researchers, is used by decision-makers. A further issue is that data generated through research projects or by digitizing collections remains with researchers or within institutions leading to a massive loss of data or inaccessible data. This results in repetitive data collection and slow progress in knowledge generation.

“The FBIP’s aim is to generate, manage, and disseminate appropriate foundational biodiversity information as the basis for research which can catalyze the bio-economy [and for decision-making] to promote human wellbeing.”

Funding research and data mobilization are the major areas of work. The FBIP tries to align the needs and that which is generated so projects must align with national priorities/strategies – it is government money. A total of R38m has been spent with the following outputs:

- 520 000 specimen / occurrence records (what species, where and when) – covering 9 000 localities and 22000 species (20% of SA’s species)
- 3 700 species and 8 000 specimens DNA barcoded
- Descriptions of 20 + microbe communities from soil, roots
- 2 400 species pages compiled
- 600 species revised, >20 new species discovered
- 59 postgraduate students provided with bursaries
- Estimated 170 papers published (71 published to date)
Examples of small projects

- Microbial diversity in soil associated with rooibos and honeybush plants
- DNA barcoding of earthworms in the vermiculture industry
- Survey of frogs in KZN and inclusion of frogs in ecotourism activities
- Impact of climate change on soil microbes
- One baobab species in SA or two?
- Database of commercial mushroom pathogens
- Barcoding of all freshwater fish and frog species in SA.
- Plant survey of the Stormberg.

Key elements for successful proposals and analyses of funded projects – Lita Pauw (SANBI)


- Proposals received: 292
- Successful: 94

Large Grants: 2013-2017
SeaKeys: Lessons from the first large FBIP project – Wayne Florence (Iziko)

Abstract:

The state of biodiversity informatics in South Africa lags behind global best practice with respect to adoption of protocols, workflows and practices based on international standards. The SeaKeys project aimed to mobilise and generate fit-for-purpose marine biodiversity information for uptake through the coordinated digitisation and dissemination of foundational biodiversity information. Additional aims included stimulation of taxonomic research, capacity building and adoption of new workflows and practices. A total of 26 national species checklists were generated, of which 17 are new for South Africa. Newly digitised occurrence records totalled 143 011, and 260 species pages were compiled. Some project successes are: new genera and species, bioprospecting discoveries, use of data in sensitive area mapping, collation and mapping of monitoring efforts and application in policy advice. Genetic barcoding efforts need improvement, but mutually beneficial collaborations with molecular researchers can help. In this talk we demonstrate that SeaKeys has been an effective vehicle for testing collaborative desire to adopt standards that render the data interoperable for online dissemination, and the utilisation of automated data cleaning and enrichment methodologies for improving data fitness-for-purpose. Generally, however, many institutions grappled with standards and agreement on the taxonomic backbone to support integrated biodiversity databases, national species checklists and collation of very large distribution data sets for applied research.
Presentation summary:

The marine biodiversity community was very disconnected and dysfunctional before the inception of the SeaKeys project. SeaKeys established itself as a coordinated multi-institutional partnership and comprises a diverse group of scientists, decision-makers, and the general public. The project aimed to unlock marine biodiversity knowledge through the coordinated mobilisation, capture, and dissemination of data [limited] and focused new research.

Key objectives and results:

- Develop new collaborations to improve the coordination, management, and dissemination of marine biodiversity data, information, and knowledge.
  - 13 co-investigators and collaboration with other large projects like the African Coelocanth Ecosystem Project (ACEP)
  - Development of 90 new links and collaborations
- Unlock priority marine biodiversity records through the collation, verification or validation, and digitisation of existing or historical and uncaptured ‘contemporary data’ (including use of citizen science initiatives).
  - 150 000 new records unlocked
  - Iziko used the opportunity to curate ‘backlogs’, 30 000 specimens using automated methods.
- Undertake genetic, species, and ecosystem research to support systematics, ecosystem description and evidence-based policy and management advice for priority taxa and ecosystems.
  - Coral taxonomy and barcoding feeding into monitoring of climate change
  - Historical occurrence data used to establish historical baselines and measure changes in fish assemblages.
  - Use of collections-based data to do mapping of sensitive taxa feeding into marine protected area design
  - Revision of national marine ecosystem classification and maps for marine spatial planning including environmental assessments and decision-making.
  - The next NBA will benefit from unlocked marine invertebrate data.
- SeaKeys has changed how the community operates - greater input into the NBA...
- Ensure that marine biodiversity data and knowledge is disseminated and applied through coordinated online platforms and applications to management and policy.
  - Deliverables were delayed due to delays in FBIP platform and templates.
- Pilot innovative capacity building knowledge generation and information flow approaches to increase, improve and apply knowledge in marine resources and ecosystem management.
  - 26 postgraduate students
  - 22 workshops and training courses were presented.
  - International taxonomists were brought in to provide taxonomic training
  - Targets were exceeded through the production of 36 checklists equating to 7500 species.
  - 350 online species pages produced
  - 42 peer reviewed papers, two book chapters, several popular articles, and a temporary museum exhibition titled ‘SeaKeys’.

Challenges & Lessons

- Coordination – we needed a dedicated project coordinator for the project.
Biogaps: Overview of the project: challenges & lessons learnt - Theresa Sethusa (SANBI)

“Biogaps: Filling biodiversity information gaps to support development decision making in the Karoo”

Abstract:

The South African National Biodiversity Institute (SANBI) is responsible for the biodiversity component of the Shale Gas Development (SDG) Strategic Environmental Assessment (SEA). Biodiversity data gaps for the Karoo were identified at the outset of this SEA and the need for rapid biodiversity recordal and assessment as part of the project. During the process of accumulating and assessing existing biodiversity data for the SEA, a more nuanced understanding of the information gaps and biases was developed for Karoo plants and animals. SANBI led a consortium of institutions in securing funding from the National Research Foundation’s (NRF) Foundational Biodiversity Information Programme (FBIP) for a three-year project entitled ‘BioGaps: Filling biodiversity information gaps to support development decision making in the Karoo’ (commonly referred to as the Karoo BioGaps Project). This project is aimed at comprehensively surveying the area to ensure increased geographic and taxonomic coverage that will allow for the mapping of species ranges, identification of important habitats, and classification of wetlands and rivers. The project involves a consortium of researchers and institutions, provides research opportunities for 11 young scientists, and is training 8 students towards post graduate qualifications. It also pilots novel approaches to engaging and developing citizen scientists. The project will mobilise 200 000 new primary occurrence records, which will inform species occupancy and habitat richness models, and which, along with 300 Red List assessments of species of conservation concern, will be served to decision makers via SANBI’s Land Use Decision Support (LUDS) tool. We report on lessons learnt to date and mitigating actions taken to ensure the smooth running and success of the project.

Presentation summary:

The project was initiated because we realized how little information we had to inform environmentally responsible development for shale gas exploration in the Karoo. Together with about 40 institutions SANBI pulled together all existing data and came up with a tally of ~2000 plant records and slightly under 1000 animal records. This is an indication of how little we know about the Karoo. In addition the data was biased towards areas having road infrastructure and tourist attractions. Hence the project set out to fill these knowledge gaps.

- Delays in data management and delivery. The FBIP templates were delayed.
- Disagreement between managers and taxonomists on ‘what constitutes a methodology for a checklist’.
- WoRMS limitations – not all taxonomic literature was captured into the system.
- Challenges in adopting occurrence record templates.
- Challenges in data quality relating to people not understanding that data quality revolves around taxonomy.
- Overestimation of targets for occurrence records.
- Project was difficult to scope.
- Unexpected expenses and procurement.
- The ‘taxonomic impediment’ is a real thing!
- HR challenges were manifold (sick leave, resignations etc.)
The main aims of the project were, firstly, to mobilise foundational biodiversity data to put us in a better position to inform decision making and secondly to systematically survey the entire Karoo.”

Outputs

- Mobilise 200 000 new species occurrence records
- Research on species distribution patterns
- 11 young scientists will take part in the project
- 300 Red List assessments to be completed for priority Karoo endemics
- Add 1 340 animal and 300 plant barcodes to IBOL, with corresponding vouchers deposited at partner museums & herbaria
- Compile 800 species pages
- Accurate distribution records of priority species and associated land use guidelines fed to decision makers at the international, national, provincial and local levels

**QUESTIONS AND DISCUSSION**

**Question to Michelle Hamer**: Are ecosystem level projects not being considered under FBIP funding?

**Michelle Hamer response**: We have considered funding projects related to vegetation mapping. The reason we did not fund it was because we didn't think that they had thought through the management side of things. Applicants will have to think about how to carry out an ecosystem project that still delivers on FBIP targets, for example species pages, occurrence records, DNA barcodes. So, we can fund ecosystem projects, but it cannot just be an ecosystem mapping project. In addition, it will still have to deliver on national strategic priorities.

**Question to Michelle Hamer**: It is clear that the entire system is gauged towards the macroeukaryotes. I mean you talk about species pages. Now, I know you’ve got some little projects on the microbial fraction but that is in fact the community that is
least well understood across this nation and everywhere else across the world. It does not seem to be a very good structure for projects of that sort, nor very good integration of that type of data. You cannot do a species page, nor would you want to for a microbial diversity study. I would doubt the fact that microbial ecology in this country is somewhere in the 18th century or equivalent in fact we are way behind the world here. And so across the biodiversity structure of this country there is a complete lack of understanding in microbial ecology, the lower order organisms.

Michelle Hamer response:

So again, it comes back to what are our deliverables at the moment. So specimen records, barcodes etc. For the microbes projects we've been funding… we will need to know where that microbe comes from and when was it collected. The sequence will need to be deposited in GenBank. So, you can still deliver on the barcode data and on the occurrence data requirements of FBIP. And it's getting the micro people to understand that primary data - a specimen or sample code linked to the identity, locality and date of collection - can be generated for microbes. So, what we need to do is bridge that gap. And then, how do we use that microbe data? You might want to use the community for bioprospecting, for inoculating agricultural fields.

Question to the Forum: Quite a few people are using iSpot. So it would be interesting to get people's experiences about that. Are there any ideas around the long term future of iSpot?

Jeff Manuel response:

SANBI is concerned about the sustainability of iSpot [as a whole]. The platform is currently being redeveloped and we need to see [first] technically where that goes and whether that matches with our aims. The SA community uses iSpot as a citizen science platform in a very different way to [what] the UK community is using it - primarily for education. Within that obviously is capacity as well. Where you maintain that platform going forward. And we’re on the verge of making a final choice because we’re in the process now of developing our own biodiversity information system. Sort of locking in our technology for [the different components] of that system. And citizen science will remain a key part of that. But we need to decide whether we will stay with iSpot or move to a different platform or develop our own platform. And look at the costs and risks involved with those.

ACTIONS:

- iSpot viability assessment
  - Platform choice
  - Costs
  - Risks
The Natural Science Collections Facility: Specimen Data Objectives – Michelle Hamer (SANBI)

Abstract:

The Natural Science Collections Facility is a national research infrastructure project selected by the Department of Science & Technology for implementation in 2017. This is a distributed network of institutions, which together hold over 30 million natural science specimens collected over more than 100 years. The NSCF will focus on making the collections and associated data accessible for research and decision-making that addresses the needs of society. An assessment of collections carried out in 2009 and 2010 highlighted the need for common software for data and specimen management, improved and common standards for data and for increased access to data. The need for online images of priority (e.g. type) specimens is also becoming urgent as the risks and restrictions associated with sending material out on loans increase. Over the next three years the focus will be on upgrading vertebrate, macrofungi, priority plant and Karoo fossil specimen data, and imaging selected type specimens and making these accessible online.

Over 80 different institutions across SA hold natural science collections. The NSCF aims to create a network of institutions which will work in a coordinated way to tackle challenges. The initiative forms part of the SA government’s Research Infrastructure (RIs) Roadmap. Key requirements are that the facility must be used for research and data must be made openly accessible.

Collections facilities in SA face the challenge of transforming a highly fragmented landscape, with many staffing and resourcing challenges dating back many years…

Central coordinating hub model: The Facility is envisioned as a network of institutions working in a coordinated way, thus contributing to common objectives.

Current situation
Challenges with software
- Data quality
- Procedures, standards, and guidelines lacking
- Capacity a major constraint
- No standards or systems for imaging of specimens

Plans for the next 3 years
- Common standards, procedures and workflows for data
- Migration to standard software (BRAHMS, Specify)
- Upgrading of data, georeferencing, cleaning
- Imaging of type specimens for the same groups – virtual collection
- Scanning of accession registers, field notes
- Capacity development – staff at institutions and contract staff appointed

Update of the digitizing of the Killick Herbarium: feedback on the vision, progress, and challenges – Boyd Escott (Ezemvelo KZN Wildlife)

**Goal: To digitize the Killick Herbarium**

**Vision: The herbarium needs to be accessible**

Abstract:

This Killick Herbarium Plant Specimen Digitising Project represents a collaborative initiative between SANBI (through the Foundational Biodiversity Information Programme) and Ezemvelo KZN Wildlife (EKZNW), which serves to bolster the resources available to enable Ezemvelo to digitally capture, verify and share (via BRAHMS) the c. 33 000 vouchers in the Killick Herbarium currently housed at EKZNW’s head office at Queen Elizabeth Park, Pietermaritzburg. The work is being carried out with the kind technical assistance of the BEWS Herbarium staff based at the University of KwaZulu-Natal, Pietermaritzburg campus. I will discuss the current progress to date, some of the challenges experienced, and more importantly, some of the lessons learnt, as well as the way forward for the project as a whole.

Presentation summary:

The herbarium has a very large collection of over 33 000 voucher specimens. New vouchers are being uncovered on a regular basis – the collection is growing even though no active collecting is being carried out. The project funded by the FBIP has three phases:

1. Barcode and photograph vouchers
2. Record verification
3. Extract and upload into BRAHMS
The National Collections of Fungi: The portal for phytopathogenic fungi from SA – Adriana Jacobs-Venter (ARC)

Abstract:
The Mycology unit of the Biosystematics Programme ARC-Plant Protection Research serves as the custodian of South Africa’s National Collections of Fungi (NCF). The NCF comprises two major collections as well as several smaller collections. The live culture collection (PPRI) houses 23,000 specimens and is affiliated with the World Federation of Culture Collections. The herbarium collection (PREM) traces its origin back 111 years and currently accommodates more than 61,000 specimens, including ca. 3,000 type specimens. These specimens represent not only South African, but African fungal diversity. The mobilisation of collection data associated with PREM specimen holdings and literature, as well as the generation of DNA
barcodes for all species represented in the PPRI collection, has led to the establishment of a portal for South African phytopathogenic fungi on the MycoBank website: www/mycobank.com/.

Presentation summary:

The National Collection of Fungi (Aka PREM) holds 61,285 specimens with 12,841 species and ~2300 type specimens.

“We have a very rich history in that we moved from agriculture to mycotoxins to veterinary sciences, back to agriculture, and fungal biodiversity. And we have all the collection data and the specimens representing the [different] focus areas in SA history.”

5-6 years ago the facility decided to digitise its collection. Initially the collection was not openly accessible so we decided to digitise the physical specimens in the collections to ensure that that entities are connected to entries.

Milestones:

- Successfully digitised the PREM collection
- Revision of all protocols
- List of PREM types
- Literature database and scanned literature
- PPRI living collection

Current collaborations/projects

- NRF Thutuka: Grassland
- FBIP: DNA Barcoding and taxonomic revision
- NRF SA China Bilateral: Macro fungi
- Oppenheimer and Son: Fusarium Survey
- University of the Free State: Macro fungi
- University of Johannesburg: Fusarium and Karoo Surveys
- Stellenbosch University: Fynbos
- University of Pretoria: Succulent Karoo
- The Royal Botanic gardens and Domain trust, Sydney: Fusarium Survey
- SANBI: Karoo Survey
- External:
  - Medicinal Plant Survey: North West University
  - Rooibos and Honeybush Surveys: Stellenbosch University
  - And Fusarium survey: Free State University

Objectives of the SA-China bilateral project:

- Fungal soil diversity
  - Kunming Institute of Botany, Chinese Academy of Sciences, China
  - Limpopo Survey
- Agricultural Research Council
- University of Stellenbosch
- Free State University
  - 12 genera:
    - 5 Provinces
    - 4 Crops
- Capacity building:
  - 2 PhD student (Stellenbosch University)
  - 2 MSc (University of Johannesburg; Stellenbosch University)
  - 1 BSc Hons (University of Johannesburg)
  - 3 BTech students (Tshwane University of Technology; Cape University of Technology)

DNA barcoding

- One gene region per isolate
- ITS
- Other gene regions
- Ca. 400 voucher specimens have been accessioned in PREM
- Biodiversity:
  - Fungi isolated from mesemb seed from Namaqua National Park
  - Fungal endophytes isolated from Alzoeaeae in the Succulent Karoo biodiversity hotspot during the dry and flowering seasons
  - Fusarium grassland biome project
    - Indigenous grasses
    - Soil
  - Fungi isolated from the Karst ecosystem
- Agriculture:
  - Alternaria and Fusarium isolates from vegetable crops
- Collections based:
  - Isolates not identified to species level
  - Fusarium, Trichoderma, Aspergillus and Penicillium

- Soil and phytopathogenic fungi:
  - Capturing of NCF data
  - Clean and verification of NCF data
  - Crous 2002 list
  - Add DNA Barcodes
  - Add new accession
  - Add reports from literature 2002-2014
  - Geo-referencing historical localities
  - Taxonomic changes: One fungus one name
  - Isolates not reported in literature but accessioned in NCF
QUESTIONS AND DISCUSSION

Question to Adriana Jacobs-Venter: What is your estimate of the proportion of the total South African fungal diversity that you've got in your collection?

Adriana Jacobs-Venter:

For each [plant] species we expect to have 10 unique fungal species. That's a very conservative estimate.

Question to Boyd Escott: How are you linking the database records back to the original specimen? Do you use barcodes or what kind of identifiers do you have and how do you make sure those links are persistent?

Boyd Escott:

Easy answer, yes, we do have unique identifiers. They're on the voucher as well, but we're also adding barcodes to [BRAHMS] as well. So we're building both in.

Obviously we don't [go into] detail with regards to the databasing but one of the other things that we're busy doing now [with the database] is we're also making a function available to have a scanned reference copy of the original data as well. So if we've got the voucher, we're checking that. But if it comes off a field form, that's also getting scanned and linked to the record. You can see the original document that it comes off.

Question to Michelle Hamer: One of the problems that we have in the institutions is that we lack capacity in terms of specialist experts to deal with digitisation, very basic tasks. And a lot of these things have already been established. Protocols have been established. There is no need to really invent the wheel. Are there any ideas on how we actually improve the capacity situation institutionally? Because that's where we need the most intervention.

Michelle Hamer:

One of the reasons that we've got the working group is so that people can bring existing documents that they've got and we certainly don't want to start from scratch again. Let us look at what other people have developed and try to harmonise across institutions. In terms of the capacity: it's the biggest constraint that we've got in these institutions - there's just not enough people. But in government at the moment, they are not hiring masses of people. Anybody in government will tell you they are trying to decrease the wage bill, and it's a big risk to just keep employing more and more people, so I think that's our big challenge (i.e. how do we get the capacity without employing lots of people). So, whether its contracts, whether its citizen scientists, whether it's through capacity development. SAIAB is a good example where they've got a small core staff, but they've got masses of postdocs and postgrads that contribute to doing the work. So we have to think about doing things differently and think about doing things cleverly because the reality is that we're not creating another 100 posts that we need at
these institutions - not going to happen. So, that’s why we’re going to have working groups for the clever people to put their heads together and solve this collectively.

Jeff Manuel:

Michelle has addressed the issue of building more capacity and we know we’ve got very few people in institutions. In most cases data management [especially] is a shared responsibility. A person’s third job might be data management. Another problem that we need to deal is that this field [globally] is evolving quite rapidly and SANBI [does] invest in getting our people onto some of the platforms. TDWG (Taxonomic Data Working Group) is the most important one where these discussions happen...I don’t think anybody else in SA really attends. So we need to get our people more hooked into the global community where these discussions are happening. The research data alliance as well. All these platforms where standards get set, practices get discussed. So that we can come back with that knowledge and then invest more in peer learning. At the moment, through the part of the FBIP grant dedicated to capacity development that we receive, we host a couple of training sessions a year etc. But that’s not nearly enough. And we’re limited by the amount of training that SANBI can give by the fact that the people giving the training are doing that at the tail end of three other jobs. But we do need to find ways despite capacity constraints while remaining at the forefront and continuing to build our skills etc.

**ACTIONS:**

- South African biodiversity data role-players need to get involved with global platforms such as the Taxonomic Data Working Group (TDWG)

**USE, APPLICATION & IMPACT OF BIODIVERSITY DATA**

**CHALLENGES & OPPORTUNITIES**

Biodiversity informatics: Meeting sustainable development challenges for fisheries in the face of climate change in southern Africa – Fatima Parker-Allie

**Abstract:**

Target 19 of the 2020 Convention on Biological Diversity indicates that, by 2020, knowledge, the science base and technologies relating to biodiversity, its values, functioning, status and trends, as well as the consequences of its loss, will be improved, widely shared and transferred, and applied. This goal can realistically be achieved only through broad implementation of biodiversity informatics, which in South Africa, as in other parts of the world, is a young and dynamic field of science. Here I present an overview of biodiversity informatics and discuss ways that SANBI–GBIF aims to assist South Africa in reaching its target, through both human capital development and informatics platforms. SANBI presently supports a knowledge-management platform, and is a major publisher of biodiversity data, making millions of data records available to the global community. Because it is important that the data have direct relevance to science and key policy issues related to biodiversity, I illustrate ideas via examples of how digital accessible knowledge (DAK) can be relevant to current research questions. In particular, I assess impacts of climate change on the distribution, diversity, and species richness of fish faunas off the coast of South Africa, exploring implications for fisheries and food security. I will use ecological niche models to characterise current, past and likely future geographic ranges, thereby permitting identification of range contractions and expansions under different climatic scenarios. This project will involve the entire life cycle of the data to improve fitness-for-use of data, for analysis, interpretation and assessment. This work is novel, as the biodiversity informatics techniques used have not been largely applied to the marine environment in South Africa and few studies have been done globally.
Presentation summary:

- Economic growth relies on science and technology – transition to a knowledge based economy to meet development objectives.
- Economic growth = innovation = human capital (underpinned by technical advancements)
- Is citation intensity a proxy for development progress?
- Information as a strategic asset – value foundational data.
- Global Biodiversity Information Outlook (GBIO) is a global vision for biodiversity informatics.
- The Global Biodiversity Information Facility (GBIF) also tries to establish strategic partnerships between other international organisations.
- We engage in mobilisation activities ad hoc but we want to build stronger and more consistent relationships.
- Biodiversity Information Management (BIM) is an emerging and dynamic field but capacity and retention of staff remains challenging.
- SANBI is looking at developing a Centre for Biodiversity Information Management, with a focus on capacity development and training/curricula (Honours)
- Niche models used to link data-science interface.
- SANBI is working with universities like Sol Plaatje and UCT.
- Management and improvement of data: downloaded SA fish occurrence data from GBIF, assessed fitness for use – and looked at data cleaning (“information” loss) techniques.
- Modelled commercially valuable species in response to climate change
- These sorts of analyses can really help inform SDGs. Need to ensure priority datasets are mobilised to support such analyses.
Towards expanding the SA Rhizobium Culture Collection as a genetic resource and its application in sustainable agriculture – Ahmed Idris Hassen (ARC)

Abstract:

The South African Rhizobium Culture Collection (SARCC) is one of South Africa’s public good assets hosted by the ARC-Plant Protection Research Institute. The collection hosts a wide range of strains of Rhizobia that are of paramount importance in the inoculation of legumes. The cultures are preserved in lyophilised form as well as frozen at ultralow temperatures, and are regularly maintained and checked for purity and viability. Demand on culture collections to supply authenticated, reliable biological material and associated information is on the rise. One of the strategies to ensure optimal microbial resource utilisation involves proper maintenance and identification, as well as database management systems that enhance the development of culture collections. There was a considerable gap in the past in terms of data management and utilisation of the rhizobium cultures due to lack of technical expertise in molecular identification techniques as well as in modern microbial database management and utilisation. Recently an initiative has been started towards expanding the SARCC as a genetic resource by identifying the strains up to species levels and by establishing a new catalogue system, which links the collection with the Global Catalogue of Microorganisms (GCM).

Presentation summary:

Work focuses on economically important legumes.

- Soybean in SA doesn’t have a specific type of bacteria, which was imported to enhance productivity.
- The Rhizobium cultures are regularly maintained (1500 strands)
- ARC an integral part of legume industry – supply stocks to commercial and subsistence farmers.
- Steps:
  - Import strain (from multiple sources)
  - Analysis of root nodules to check purity
  - Grow and concentrate bacteria
  - Taxonomic identification
- Many cultures are 40 years old, kept in freezers
- Databases however are very primitive and much is kept in analog form, difficult to read (e.g. bad hand writing), so digitizing is difficult. Digitised records have gaps etc. – shows importance of having a data management system in place from the beginning
- Going to link to World Federation of Culture Collections/Global Catalogue of Microorganisms (GBIF of culture world!). To date, none registered from SA. Important, because cultures can be exchanged and info is freely available.
- Applications:
  - Sustainable agriculture – enhance nitrogen fixation.
  - Rhizobium inoculants (e.g. soybean).
  - Research: e.g. Rooibos inoculated by a wide range of rhizobium, which allows isolation and commercialization of effective strains to increase productivity.
- Conclusions:
  - In house data management systems key for biodiversity collections, which lays foundation to improve science and contribute to research
  - Important clarification, all strains must in some way contribute to food security. If it doesn’t inoculate any crops, they throw it away.
Abstract:

Amphibians are vital to ecosystem wellbeing, and incorporating them in conservation planning is of utmost importance. However, amphibians remain largely understudied or misunderstood. Their conservation, as with other conservation initiatives, is not relatable to people and thus often ignored. This study pilots methods of increasing knowledge and understanding of amphibians along with the importance of their conservation in the South African context. Local communities and the tourism industry are impacted the most by the outcomes of conservation initiatives, thus their buy-in is vital to the success of such initiatives. Making amphibian biodiversity data accessible and appealing to these stakeholders lays the foundation for increased success of conservation initiatives. The study aims to do so through presenting frogs as a tourist attraction and investigating frog-related indigenous knowledge. Outcomes of this pilot are in the form of an amphibian diversity workshops and a bilingual book for Zululand, and integration of frogs into the list of tourist attractions at Ndumo Game Reserve. Lessons learnt from this study will inform the design of a new, large-scale project focused on making amphibian biodiversity relatable to non-scientists across South Africa.
Presentation summary:

- Other side of the coin – data to stimulate culture change (non-commercial).
- Focus on ecotourism and community aspects.
- Two new frog species discovered in Zululand.
- Project a pilot to test whether understanding biodiversity leads to improved success of conservation initiatives and access to bioeconomy – planned PhD to focus on national scale.
- Workshops with local stakeholders to package data and information – treated each community as a different ‘publication’ with their own ‘author guidelines’.
- Tourists and safari guides are fast using frogs as attractions - unlikely to become sole attractions, but can be complementary.
- Uses app to help guides verify ID to increase tourism impact.
- Bi-lingual frog book framed according to how community members wanted to learn about frogs.

Bridging gaps: When frogs ‘bring rain’

The Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES): the importance and relevance of data in the African Regional Assessment – Anicia Malebajoa Maoela

Abstract:

The Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services (IPBES) is geared towards strengthening the capacity and knowledge foundations of the science-policy interface in relation to biodiversity and ecosystem services at both global and regional level. IPBES identified the improvement of access to data, information and knowledge as key toward achieving its mandate. In this regard the platform established partnerships with data mines such as GBIF and IUCN, to enable access to data sharing infrastructure and to facilitate the interpretation of biodiversity data in the assessment of status and trends of biodiversity and nature’s value to people. Despite numerous efforts by IPBES to improve on access to information there are still challenges. To be discussed in this talk are the progress and challenges of addressing (i) capacity building interventions and (ii) inclusion of research institutes in Africa to partner solidly with IPBES on promotion of shared interests. Such common interests include, but are not limited to, data standards and exchange formats, and provision of free-to-use tools and information resources on data mobilisation, publication, access and use.

Presentation summary:

• An independent intergovernmental body, established by governments in 2012 – currently 126 Members.
• In response to requests from governments, IPBES provides policymakers with objective scientific assessments about the state of knowledge regarding:
  - The planet’s biodiversity, ecosystems and their contributions to people
  - Tools & methods to protect and sustainably use these vital natural assets
• Provides options for responses based on the best-available science.
• IPBES’s Mission:
  - To strengthen knowledge foundations for better policy through science, for the conservation and sustainable use of biodiversity, long-term human well-being and sustainable development.
• IPBES does for biodiversity what the Intergovernmental Panel on Climate Change (IPCC) does for climate change

What does IPBES do?

1) Expert Assessments (synthesis & critical evaluation of available knowledge)
   a. On specific themes (e.g. “Pollinators, Pollination & Food Production”).
   b. On methodological issues (e.g. “Scenarios and Models”) at both the regional and global levels (e.g. “Global Assessment on Biodiversity and Ecosystem Services”).
   c. Scientific credibility ensured by IPCC-inspired processes: selection of experts by experts, based on CVs and thorough peer-review (thousands of comments, addressed individually, posted online).

2) Policy support
   a. Identifying policy-relevant tools and methodologies.
   b. Facilitating their use & catalyzing their future development.
3) Building capacity & knowledge
   a. Identifying & meeting priority capacity, knowledge & data needs of IPBES Members, experts & stakeholders.

4) Communications & Outreach
   a. Ensuring the widest reach and impact of IPBES’s work

Why does IPBES matter?
- Biodiversity & nature’s contributions to people underpin almost every aspect of human development.
- Also key to the success of the Sustainable Development Goals (SDGs) and Aichi Biodiversity Targets.
- Yet biodiversity is being depleted & degraded faster now than at any point in human history.
- Help meet African Union (AU) Agenda 2063 and implement initiatives e.g. useful to:
  - AU and its agencies, African Ministerial Conference on the Environment (AMCEN), African Development Bank (AfDB)
  - Sub regional political and economic groups (South African Development Community (SADC), Economic Community of West African States (ECOWAS) etc., COMIFAC, and the Great Green Wall Initiative etc.) and regional banks.
  - Other stakeholders (private sector, international organizations, NGOs, media and each one of us).

Challenges
- Voluntary nature
- Availability of expertise and limited experience
- Strategic partnerships
- Financial resources
- Capacity building
- Access to information still a challenge to some authors
- Lack of experience in conducting ecosystem assessments
- Communication: Wide scope, number of key messages, languages and accessibility

Knowledge, information, and data needed?
1. Data/information sources
   - Published data/information: scientific and grey literature, and indigenous and local knowledge
   - Existing data/information held by international (global, regional, sub regional) and national institutions/bodies
2. Strategic partnership and initiatives
   - Providers of data, metadata and projections
   - Users including at science – policy interface
   - Supporters of capacities
3. Operational structure
   - Authors: interact with each other, with similar groups undertaking global, thematic and methodological assessments in order to ensure conceptual and methodological coherence
4. Capacity building
   - Technical support unit (part of the secretariat)
   - Experience
   - Young fellows
Outcome: We’re hoping to determine how threatened our ecosystems are...

The guide on diverse conceptualisations of values considers five different categories of values, namely biophysical, socio-cultural, health, economic, and holistic values. It also provides a 6-step approach to assessing diverse values within IPBES assessments.
Data needs for high-level Biodiversity indicators: Is there a gap? – Jeff Manuel (SANBI)

Abstract:

The South African National Biodiversity Institute (SANBI) is mandated to monitor and report on the status of biodiversity in South Africa. SANBI meets part of this mandate through the National Biodiversity Assessment (NBA), which reports on the status and trends of biodiversity and ecosystems on a 5–7 year cycle. The NBA focusses specifically on high-level indicators such as threat status and protection level for species and ecosystems. SANBI and DEA have started a process of aligning the biodiversity indicators used in national and international reporting processes linked to the NBA, South Africa’s Environmental Outlook (SAEO), the Convention on Biological Diversity (CBD) (including the Aichi targets), the Intergovernmental Panel on Biodiversity and Ecosystem Services (IPBES) and the Sustainable Development Goals (SDGs). But what are the data needs linked to these indicators? Is there a data gap? Using the Sustainable Development Goal (SDG) Indicators and South Africa’s National Biodiversity Assessment we illustrate the kind of information required to populate these high-level indicators. We highlight the importance of working collectively to ensure that appropriate national datasets are used in these national and international monitoring processes, and address some of the risks in using standard global datasets (often recommended).

Presentation summary:

• Biodiversity monitoring: systematic focused observation and measurement of changes.

• Benefits of monitoring:
  - discovery
  - guide policy
  - adaptive management

• SANBI provides a science based policy to government.

• NBA focuses on high level indicators for monitoring, whereas SANBI focuses on state & pressure indicators.

• Two very powerful National Biodiversity Assessment indicator(s)
  - Threat status
  - Protection level

• SANBI tracks protection level of ecosystems

• Over 60% of wetland ecosystems are threatened in South Africa.

• Developing indicators are not simple; an ecosystem threat level assessment is needed.
- over 100 years of bio science  
- a lot of repeatability; therefore we need a system for repeat assessments  

• What is required for the threat status?  
  - We need a good understanding of biodiversity for threat status  
  - We need biogeographic information to build a map of ecosystem types  
  - We need to develop a classification system.  
  - Imperative: good scientific basis to set targets.

• In South Africa we have sufficient data, however there are still gaps.

• SA’s terrestrial vegetation is stable.

• Freshwater ecosystems are hoped to be stabilized in 2018 by the NBA.

• Unlike climate change, we do not have variables to work with.

• National Biodiversity Information System (NBIS): is an important foundational element.

• It’s a challenge to improve monitoring framework.

A quarter of a century of taking databases from paper to cloud – Les Powrie (SANBI)

Abstract:

This is the story of what was done, and some lessons learnt as we bumped our heads on a very rough road. Banging our heads was worth it to help making it feasible for many users to travel a smooth data sharing road. A big focus of work at SANBI involves linking biodiversity assessment and monitoring with good biodiversity information management – this involves taking data from different toolboxes and workshops and putting them into a single toolbox in a single workshop. A good example of this experience is that over the last 25 years multiple data sources were combined. Observation data from vegetation surveys (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. 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. Recent work has been done to take it from this local database and put it into BODATSA (Botanical Database of Southern Africa) to make it accessible online in the global workshop. 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’.

Presentation summary:

• What was done, and some lessons learnt…

• In the process of our work done 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, data was mined and therefore we had to develop ways to achieve our outcomes.
The hope is 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’.

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.

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.

Example; 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. Aim: Wanted to integrate data to make it more powerful, correct errors because using data increases the likelihood of errors being detected.

In September 1993 transcription of 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; 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.

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. Also, same collector and specimen number.

Work to capture the verbatim text without delay; 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.

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.
South Africa’s first national status report on biological invasions – Heather Terrapon (SANBI)

Abstract:

This presentation provides a preliminary assessment of the status of biological invasions in South Africa. In terms of regulations under the National Environmental Management: Biodiversity Act, the South African National Biodiversity Institute is required to produce a report on the status of invasive species, and on the effectiveness of control interventions and regulations, every three years. The team responsible for compiling the report has developed a set of indicators that assess the main aspects of invasions, as well as the effectiveness of control measures and of the regulations. One of the aims of this talk is to obtain feedback on the preliminary findings. Preliminary data on the number of alien species in the country, and their status in terms of distribution, dominance and impact will be presented, as well as the degree to which particular areas are invaded. An assessment of the effectiveness of control is difficult, as there are almost no adequate management plans that indicate the intended goals of control measures in particular areas, and no adequate monitoring and assessment of outcomes. While there are some localised successes, there is also a great deal of evidence that control measures have been ineffective in many areas. The regulations have been in place for less than three years, and it is probably premature to expect that their effectiveness could be assessed at this early stage. The status report on biological invasions has to be repeated every three years, so it will in future offer the opportunity to track invasions over time.

Presentation summary:

The purpose of the status report is to take information from research and monitoring, assess the status and then inform policy and management. The report will allow for risks and challenges to be identified and addressed. Hence more informed decisions will be made. SANBI needs to submit a report on the status of listed invasive species within 3 years, and every 3 years thereafter.

- The report considers the status of alien species, status of invaded areas, effectiveness of control measures and the effectiveness of regulations.
- Process for report: find advisory group; develop a suite of indicators; a draft is then released.
- 21 indicators for pathways; species; areas; control effectiveness.
- 4 high level indicators
- The 3 pathway indicators:
  - Number of pathways
  - Prominence
  - Introduction rates.
- Indicators for alien species:
  - Number and status of alien species; high confidence for plants/birds; low for the rest; 119 present but not naturalised, 307 naturalised but not invasive, 720 invasive.
  - Extent of alien species; high for plants/birds; low for the rest
  - Abundance of alien species; low confidence?
  - Impact of alien species; Low
- 2102 alien species regulated (559), or recorded in the wild, in South Africa.
- Increase between 2000 and 2016 by 172, from 601 to 773 species.
- Marked increases in distribution of nine species.
- Spread of species under biological control has slowed down or stopped.
- No detectable impact of control on other species.
- Alien species with massive impacts; pine, wattles, Prosopsis, Smallmouth bass, Mediterranean mussel.
- Indicators for invaded areas
  - Alien species richness;
  - Relative alien species richness
  - Relative alien species abundance
  - Combined impact of alien species
• What we learnt?
  o Western Cape is the most invaded province (28% of area invaded to some degree)
  o Mpumalanga (16%)
  o Northern Cape (14%)
  o KwaZulu-Natal (9%)
• ...more than 20 years out of date; changed considerably since; but no effective monitoring of this.
• What are we doing about this? Inputs; outputs; outcomes.
• Only one entry point monitored by DEA at present
• Indicators of effectiveness of control measures: Pathways
  o Separate measures by DAFF for agricultural pests could also reduce the rate of arrival of new environmental problems.
  o Insufficient data to link the impact of these measures to the outcomes
• Indicators of effectiveness of control measures: Species
  9 historical eradication programs
    o Three succeeded (cats on Marion Island, two terrestrial invertebrates)
    o Six failed
    o More eradication expected over the coming decade.
• Area control measures
  o Inputs: ~ R1.5 billion a year
  o Outputs: 2.8 million ha cleared, 8 million ha followed up (over 20 years)
  o Outcomes: Localised successes, but continued growth in populations of most species at a national scale.
• Contributing factors to success:
  o Ongoing direction from a diverse project steering committee
  o Rapid response team, regular monitoring
  o Focus on areas of low infestation, very flexible management approach.

DATA SHARING BETWEEN SANBI AND PARTNERS & INFORMATION SYSTEMS AND ARCHITECTURE

DATA SHARING

FBIP data sharing - Lita Pauw (SANBI)

• Granting processes managed by NRF through an expert panel.
• FBIP conditions – data must be made publically accessible. i.e. third-party agreements are required. However, ownership of the data will be attributed to the data provider.
• One of the key requirements of the FBIP, is to make the funded data accessible via the internet, including numerous online international portals. Two restrictions are possible:
  o Temporary embargoes for publishing
  o Sensitive species

SANBI’s data sharing agreement - key principles – Fatima Parker-Allie (SANBI)

• Biodiversity information is the key resource/asset.
• Our researchers are increasingly using and citing GBIF data – value proposition.
• BIMF plays a key role in harmonising the sharing of data.
Open Access discussions date back to 2000 – publicly funded institutions legally bound to making data publically accessible.

We used to have different data sharing agreements in different divisions – uncoordinated. Plan is to have one agreement per institution and to streamline administration and increase transparency. Only the annexes in the agreement will be updated.

The agreement provides for 3 (three) categories of non-exclusive use of the Data shared by the Data Provider to SANBI including

- Use of the Data by SANBI,
- Hosting/serving the Data through SANBI platforms and portals (primarily the Biodiversity Advisor) and
- Hosting through SANBI and allowing SANBI to publish the Data to other portals

A number of considerations will also be addressed including:

- Access to sensitive data will be restricted
- Licensing and attribution
- Metadata template will also be provided to data provider to complete

INFORMATION SYSTEMS & ARCHITECTURE

Update on the National Biodiversity Information System (NBIS) – Brenda Daly (SANBI)

Abstract:

This presentation provides a progress report on the proposed data sharing architecture for a National Biodiversity Information System (NBIS) to be developed by the South African National Biodiversity Institute (SANBI). NBIS includes the following key components (content types): i) systematics (taxonomy and nomenclature), ii) biogeographic (specimen and observations), iii) ecosystems, iv) descriptive data (resources metadata), v) molecular, vi) literature and vii) multimedia. The project has started promoting a shift from tactically based information systems, aimed at delivering products for individual project initiatives, to strategic systems that promote the building of capacity within biodiversity organisations and networks. This will encourage data to be managed more effectively within SANBI, but also encourage data to be shared by the community providing integrated products and services that are needed to address complex environmental issues.

Presentation summary:

The NBIS discussion in 2016 dealt with the prospects of developing an integrated architecture so that other institutions can feed in their data.

- Different datasets sit on different systems (Specify, BRAHMS, IPT etc.) – we want all these systems integrated.
- Not necessarily a big database, but an indexing system that pulls up everything about the species from all the various sources.
- It is very difficult to find the correct information at present because it is scattered across various websites – we want to merge them to have one user interface
- This will be done by TOGAF 9 Enterprise Architecture.
- We are at the point of how we’re going to get this done.
This plan will help reduce operating costs across all partners. The plan is essential to set community standards for all partners.

- Transfer, sharing and citation.
- Unique identifiers – this will help with versioning control.
- Licensing of data and applications
- It is important that we bring together data hosting portals, they must be sustainable. I.e. rather have a few portals with dedicated funding rather than many ephemeral or vulnerable ones.
- Animal Demography Unit and KZN-Ezemvelo are data partner test cases.

QUESTIONS & DISCUSSION

SANBI’s data sharing agreement – key principles (Fatima Parker-Allie)

1) **Boyd Escott question**: Feedback on errors – How do we communicate with data providers (DPs) about errors found by SANBI
   a. Jeff Manuel answers: this is where the 3rd SANBI category comes in – allows SANBI to transform data. But it still stays the data provider’s property. We will not overwrite fields but create new fields that allows options on how best to display data.
   b. Wayne Florence comment: It is an issue of quality. We want to update and clean the data ourselves, and the alternative is that we host the data ourselves and you extract the data from us – we work in real time with the data. This will require infrastructure being based at institutions or regionally.
      i. Jeff Manual response: This is possible. Datasets are sent back to data providers to clean, but this situation covers the case where DPs are incapable of cleaning data themselves. We are trying to put in place infrastructure to allow DPs to publish data themselves as long as data quality is managed in a coordinated way.
   c. Fatima Parker-Allie comment: When we initially started the SANBI-GBIF process we had a decentralized system. We were hoping that the system proposed by Wayne would happen, but there were capacity constraints around that. In an ideal world it would be great.

2) **Boyd Escott question**: Are there processes in place to document how corrections are done, so one can go back to the original source?
   a. Jeff Manuel answer: We do need to improve how we document the data cleaning that we do perform. We’re trying to increase transparency across the board.

3) **Selwyn Willoughby question**: How are you dealing with commercial use of data?
   a. Jeff Manuel response: Two licensing agreements – but preferred one is for non-commercial use. Default is going to be open with attribution. Commercial use is a grey area. We prefer to deal with it not as a licensing issue but a category of use issue.

Update on the National Biodiversity Information System (Brenda Daly)

1) Wayne Florence question: How can we help you?
   a. BD answers: We are going to get user (consumer) feedback to instill trust

2) Vince Smith question: what is the extent to which we have built the partner information systems, as they are dynamic and shifting? DiSSCo (Distributed Systems of Scientific Collections) looks conceptually very similar. You might want to share diagrams and plans.
   a. We are looking to build the parts for the other institutions too, such as national license for BRAHMS.

3) Les Powrie question: is there a chance that the plant ‘stuff’ will move out of BRAHMS?
   a. We need comparative systems to ensure any one system meets business requirements. We’ll just need to back up systems. But suitability of systems change (e.g. iSpot). We will need to constantly evaluate the suitability of various platforms and systems.
4) Jeff Manuel comment: Our practices align with global standards. Only constraint is being able to implement and track every data contributor. We are going to get the data-sharing agreements out to institutions so it can go through legal departments.

**ACTIONS:**

- Explore the possibility of data update and cleaning infrastructure being based at ‘data provider’ institutions.
- Develop processes whereby data providers can document how corrections are being done, with the option of accessing the original source.
- Develop consumer feedback tool for the NBIS
- Share ideas (diagrams and plans) with DiSSCo, for NBIS optimization.

**ADVANCES IN DATA SCIENCE & INFORMATION MANAGEMENT**

**DATA ANALYTICS, BIG DATA, DATA MINING & DATA VISUALISATION**

**Power to the people: Citizen Science setting collections data free**

*(Vince Smith, NHM London)*

**Abstract:**

Crowdsourcing data from the world’s natural science collections is a complex task with many challenges to face and opportunities to be seized. In this presentation, we reflect on our pilot activities in crowdsourcing the transcription of specimen labels from our digitised collections at the Natural History Museum London (NHM). We will share what we’ve learned about the behind-the-scenes elements of crowdsourcing – the parts the ‘crowd’ doesn’t see – e.g. flow of data between systems. For example, at the NHM London there are four systems involved (the Crowdsourcing Platform, Content Management System, Media Asset Management system and our Data Portal) and many policies and processes (data embargoes, our ‘visiteering’ programme, our marketing efforts to make unfamiliar groups more appealing, data quality procedures and efforts to attract funding). We also share our vision for the future of crowdsourcing and highlight emerging technologies that could open up new avenues of citizen science activity. We will close with a number of challenges and opportunities for the future that come from...
digitising collections of this volume and complexity, and engaging broad and diverse publics with these internationally important collections.

Presentation summary:

There is much interest to participate in citizen science projects at present, within the UK this is reflected in BBC programming and BBC news coverage. People choose to participate for various reasons and there are often combinations of factors.

Initial motivations to participate can be highly variable and available studies show that these are often quite different to the motivations of the project developer. There is an acceptance that we are unlikely to ever truly understand the motivations of all participants – but understanding dominant motivations of sample of participants can be of great help in supporting their involvement and developing the project to maximise this enthusiasm. Studies to date look at subsamples of participants and have aimed to extract dominant trends (e.g. reasons for initial involvement and motivations for on-going participation).

An interest in /affinity for project and its goals is a common thread throughout most projects.

There have been a number of recurring components that make citizen science and crowdsourcing projects successful:

- An active community, supported by the crowdsourcer that can provide quick feedback and acknowledgement – being able to flag problems or ask questions is crucial.
- Evidence that data are being used – seeing it become published in different media (news articles, popular science and traditional peer-reviewed papers)
- Social interactions and support from other users is important – feeling part of a community is a very strong motivator and means the community is more likely to remain active and happy for longer. Many users report becoming better at recognising and interpreting the handwriting of particular people (on labels) or becoming better at detecting localities.
- Super-contributors are a recurring feature – just look at the H@H top 20 users – less than 1% contributed more than 30%! A successful project should recognise and work with their super contributors.
- Gamification in general works for some of the community but not others!

There is a lot more to crowdsourcing than transcription alone. A project should be personally and socially relevant if it wants to succeed – it grows with importance relative to the size of the task. One of the missions of natural history museums is to work towards having a society that is more engaged with science and current societal challenges. Crowdsourcing is a way to achieve this mission.

By its nature, crowdsourcing is openly participatory and allows the public and media to understand our research narratives through contribution and discussion but the motivations of participants are often different to those who create projects.

Successful projects should meet the needs of multiple users as a single approach or a one-size-fits-all approach is unlikely to work. For example, not all users are motivated by gamification and leader boards while some have little interest in the community aspect.

The overall key to success is to work directly with your crowdsourcing participants to ensure you understand them and meet changes that will occur as the field and the community matures. Larger projects are likely to require dedicated community managers.

Virtual safaris mobilizing biodiversity information in African countries
(LES POWRIE, SANBI)

Abstract:

A seed was planted during 2012 that has led to the mobilisation of many species distribution records and powerful georeferencing capability in Africa. The birth of SAFARIS (Southern African Friends and Researchers Indexing Specimens) was announced during a meeting with delegates from several African countries later that year and it now has a cousin called BioGaps (Karoo BioGaps Project). Together, these projects are releasing information from historical sources such as field
notes and specimen labels to become available in NBIS (National Biodiversity Information System). There have already been about 53,000 plant species and 1,300 animal distribution records made available. The data can be used in research, monitoring, policy, and any other conceivable application. Some of the value derived from SAFARIS has been correcting information in BODATSA (Botanical Database of Southern Africa) for herbarium specimen label transcriptions with typographical errors, such as a year transcribed as 1838 instead of 1938. Georeferencing of the combined data is also far more accurate and effective with more data giving better context to individual records. Many are enjoying travelling around our regions with early explorers as they transcribe their records making the information available for various uses.

Presentation summary:
A seed was planted during 2012 that has led to the mobilisation of many species distribution records and powerful georeferencing capability in Africa. I have had varying degrees of input over many years with indexing birth, marriage and death records from microfilmed images of government and church records from mainly South Africa and the UK, and more recently web-based indexing. I, in turn, have benefitted from the indexing efforts of unknown volunteers who indexed records of some of my own family line – Dryden and Mason.

It occurred to me that we could do something similar with biodiversity records by indexing images of herbarium specimens and field notes of early botanical explorers in South Africa. My thinking went beyond that to museum records and other early collectors. I started to look at what could be done and found that mine was not the first mind to think of it! There were some other projects on the go. I approached our biodiversity information leads at SANBI and SABIF and they gave the go-ahead to proceed.

Shortly after that there was a meeting with delegates from several African countries, and from other continents including Australia, held at SANBI dealing with biodiversity information. This idea was still growing in my mind. The birth of SAFARIS (Southern African Friends and Researchers Indexing Specimens) came as I enthusiastically announced the concept at this meeting and SAFARIS now has a cousin called BioGaps (Karoo BioGaps Project).

Together, these projects are releasing information from historical sources such as field notes and specimen labels to become available in NBIS (National Biodiversity Information System). There have already been about 53,000 plant species and 1,300 animal distribution records made available on SAFARIS (on ALA – Atlas of Living Australia) and Transcribe.

The data can be used in research, monitoring, policy, and any other conceivable application. Instead of borrowing specimens or visiting herbaria or museums to view specimens, accessing high resolution images often meets the needs of researchers, unless physical access is required such as for chemical analysis of material from, or microscopic study of a specimen. Specimens are handled less because the images are viewed by most users because digital images with text associated are searchable and discoverable online enabling more users to be able to view them.

Besides the taxonomic use of specimens, the species occurrence data are made available in databases and have been invaluable for species distribution mapping, niche modelling, checking endemism of species, Red List assessments, impact assessment, and many other uses.

Examples of the value derived from SAFARIS has been correcting information in BODATSA (Botanical Database of Southern Africa) for herbarium specimen label transcriptions with typographical errors, include a year transcribed as 1838, in the century before the collector was born, instead of 1938, a collector name not recorded in full, and selection of the wrong named locality.

Duplicates of the same collection in different herbaria with different Determination status and species names attached. The more people interrogating and using the data, the more likelihood there is of errors being detected.

Georeferencing of the combined data is also far more accurate and effective with more data giving better context to individual records.

Much better results in a fraction of the time.

Example: There were 182 specimens collected by IB Pole Evans over a period of 11 days during October 1928. These specimens were dispersed in their respective species folders in three wings of the National Herbarium.

Before: Not georeferenced = 58%

Georeferenced: 42%. Done one at a time, would have taken about 40-50 hours spread over 35 years, probably about 70 different people, resulting in 37% specimens being mapped to a precision of about 18 km, 5% with errors ranging from 320 km to 650 km.

Not georeferenced: 3% with no locality, 14% with only regional description (province or district). 3% vs 58%

After: Georeferenced: 83%.
Georeferencing using the itinerary

- took less than 3 hours as opposed to 40-50 hours
- 70% 1-2 km precision. Compared to 48% to 18 km precision.
- 12% 2.5-5 km

Deductive work

It can also be fun and challenging working with the historical records. By using field notes we were able to improve a false georeferencing of Cashel Hotel to the only Cashel in the place names database. Instead of a QDS (18 km) in Kalahari, the locality was georeferenced to near Hartebeestpoort Dam with a precision of 5 km. Hopefully this can be improved still more.

Many volunteer indexers are enjoying travelling around our regions with early explorers as they transcribe these historical records making the information available for various uses, liberating the information from paper and elevating it to the Cloud.

Toward next generation biodiversity research (Morne du Plessis, NZG)

Abstract:

The National Zoological Gardens of South Africa is a major role-player in the field of wildlife conservation genetics and as such also recognises its role in contributing to biodiversity information and research. As part of its drive to advance research through next generation technologies, the organisation has acquired an Ion Torrent S5 sequencer, which enables the generation of biological ‘big data’, along with computer servers for the bioinformatics analysis aspects of this data. In line with these acquisitions we are proposing to utilise the technology, in conjunction with the diverse skills set of the research team, to evaluate the feasibility of conducting next generation biodiversity research. The strategy would focus on using environmental sampling in conjunction with next generation sequencing to assess the biodiversity environments. The NZG will focus on optimising a number of methods as techniques for biodiversity assessments, which include: (a) metabarcoding; (b) metagenome sequencing; (c) microbiome analyses and (d) metatranscriptomics. The strengths and weaknesses of these strategies will initially be evaluated to establish which are effective and relevant in terms of contributing to the Foundational Biodiversity Information Program. The selected strategies will then further be optimised for implementation in future assessment strategies.

Presentation summary:

Merging Next Generation Sequencing (NGS) + Environmental DNA (eDNA) + Bioinformatics = Next Generation Biodiversity assessment

Techniques – Variations of metagenomics approaches

- Microbiome analysis – 16S
- Barcoding - Animals – COI
- Plants – rbcl, matk trnH-psbA, ITS
- Shotgun sequencing – direct environmental sequencing
- Transcriptome analysis

eDNA – Direct from environment (soil, plant matter, animal matter, water)

Indirect from environ – Fecal matter – host and what they consume

Indirect parasites or insects that feed on other animals (e.g. bloodfeed analysis)
The NZG Biobank information & information systems: Analysis and improvement (Kim Labuschagne, NZG)

Abstract:

An analysis of the workflow and information management procedures and systems of the National Zoological Gardens of South Africa (NZG) Biobank was conducted and it was found that there was a lack of integration and interoperability across the board. It was therefore decided to develop and implement a new Biobank Information Management System (BIMS) based on the Specify platform. The first phase of the project was to improve the Biobank information and information system. Within the Biobank alone, data management procedures were historically based on four different files (Microsoft Access or Microsoft Excel). To develop the new system, data from these files were first cleansed and then migrated to the Specify database. Being hierarchical, the Specify taxon tree and storage tree are particularly effective in controlling data quality. We optimised the Specify user-forms for data capture and querying, and developed Specify labels with barcodes for vials.

Presentation summary:
The NZG Biobank converted their data into SPECIFY – imported and QCed. They have created custom queries and optimised user forms for input/output. They are integrating their various programs (some complete), and Specify has worked well for them.

Q: How did you get your data from Access & Excel into Specify?
A: Quite a laborious process, basically by hooking up Access into the Specify tables. But the workbench in Specify is not at all a technical process and is quite easy to use, and it does lots of verification automatically.

Turning the Animal Demography Unit databases inside out with modern application architectures (Ian Engelbrecht, University of Pretoria)

Abstract:

Database systems are traditionally designed to be inward facing, serving the needs of the organisations that create and use them and nothing further. Software applications communicate directly and seamlessly with the databases that support them, and security and data integrity concerns are built directly into those applications. This scenario is changing though, as organisations increasingly need to provide outward-facing applications for use by a wide range of external stakeholders. Biodiversity databases are no exception, and there is growing interest in using modern application architectures, especially web APIs (application programming interfaces), to provide data to a wide variety of users seamlessly and securely. An API is extra software that can be thought of as a broker between a database and external users of that database, providing a standard set of data outputs, while controlling access to and security of the data. Web APIs take the form of HTTP URLs that can be called from a web browser or a variety of other software programs, and which return data in JSON or XML format. APIs can also include the option to add data to a database, allowing for the possibility that independent parties can build and deploy their own products that make use of databases. This talk will present an overview of the APIs that have been built or are under construction for the Animal Demography Unit databases including the Virtual Museum and the South African Bird Atlas Project. The functionality of the APIs will be described and demonstrated, and lessons learnt in their development will be shared.

Presentation summary:

This talk gives an overview of how an Application Programming Interface (API) was developed for the Animal Demography Unit (ADU). There are large quantities of data (birds, reptiles, invertebrates – all of the MAP projects – e.g. MammalMAP, OdonataMAP, etc.). But how do you reach out and get hold of this data? Before you had to send a data request to the ADU, but now the data can be accessed directly via the API. There are several endpoints, authenticated and non-authenticated, including endpoints for summary statistics.

A system to automatically generate annotated checklists (Willem Coetzer, SAIAB)

Abstract:

Species checklists are needed by a broad range of workers in the field of biodiversity science, from researchers to government officials. The tradition of taxonomy and systematics is to publish an annotated checklist in a journal, for which the author usually receives benefits in addition to a growing publication record (e.g. increased funding). The lack of such recognition has been a stumbling block to data-sharing initiatives. A prototype information system named the Catalogue of Afrotropical Bees (CAB) is described. The CAB automates the generation, from a database, of a Darwin Core-compliant annotated checklist. While the checklist can be updated easily when needed, the author’s benefit from traditional publication need not be jeopardised due to the advent of the Biodiversity Data Journal, which will publish a scientific article associated with a dataset, the link to which is also included. The technology of the World Wide Web has improved dramatically since the days when checklists were published by scientists themselves, using simple HTML (essentially no different from paper publications). Biodiversity information standards, such as the Darwin Core, have allowed scientists to publish richer metadata directly from a database, resulting in better quality and consistency and reaching a wider base of users. The CAB uses the Specify database.
of species, citation and article records, and includes occurrences. The data are aggregated, transformed and standardised by database queries, and the final product, a Darwin Core Archive, is published to the GBIF Data Portal using an IPT server. The CAB is a typical annotated checklist that contains important nomenclatural citations (e.g. species descriptions, new combinations and synonymisations) as well as other interesting or useful information (e.g. biological or ecological information such as hostplant relationships). It is hoped that the CAB will be developed into a robust, industry-standard software tool for ongoing, wider use.

Publishing standardized, enriched biodiversity information

- Biodiversity information MUST be standardized i.e. use particular, accepted vocabularies of terms, mostly originating from TDWG
- Darwin Core is one of the most significant and widely used, ratified standards for sharing biodiversity data

Validating records georeferenced for the BioGaps Project (Willem Coetzer, SAIAB)

Abstract:

Georeferenced occurrence records of all South African biodiversity were required by the BioGaps Project to assess the potential effect of shale gas extraction on biodiversity in the Karoo. Records without coordinates were georeferenced by a team of workers at SANBI, and the completed records were sent back to the suppliers to be incorporated in the original databases. An exercise to assess the accuracy of the 4,683 georeferenced occurrences of Apoidea was undertaken. Overall the assessment of the accuracy of assigned coordinates was good. Between 10% and 15% of locality records were either incorrect or too imprecise and probably should not have been georeferenced. Observations that were made during the assessment included the fact that farms with common names (which could refer to many farms all over the country) had been singled out and georeferenced. The maximum uncertainty radius associated with the most imprecise estimates seemed to be too small. All records received from the georeferencing team need to be checked for accuracy before they are incorporated into a database. This is a potential bottle-neck in future georeferencing projects.

Presentation summary:

Main point: It’s important to validate the output of georeferencing before we feed back into the database. Biodiversity information must be standardised, and we should be concerned with data quality.

The Namibia Biodiversity Database – content, capabilities, and current status (John Irish)

Abstract:

The NBD (on-line at biodiversity.org.na) includes basic biodiversity information for Namibia and an adjacent portion of the southeast Atlantic Ocean. It aims to be the go-to site for determining Namibian species presence; current taxonomy, synonymy and classification; Namibian legal, conservation and endemic status; basic species context data (herbivore/carnivore, terrestrial/aquatic/marine, indigenous/alien, etc.); Namibian language common names; and interpretively collated published Namibian distribution data. The latter is being added chronologically, currently completed: 1760 to 1875. Current content: 13,869 taxa, 13,2626 records, and growing. All vetted data is visible on the site, and free. The web site is a
PHP front-end to a MySQL/MariaDB database. It is supported by an extensive editing, maintenance and security subsystem, and a variety of cron-based bash scripts that automate repetitive tasks. Only FOSS is used. The site can display in any indigenous Namibian language. The NBD originated in 2003 as a Government-supported, donor-funded project under the National Biodiversity Programme, but some data roots go back to 1983. There was no provision for continuity beyond the end of the NBP, but I kept it alive, inter alia with GBIF support, till 2009 when it was officially abandoned by Government. At that time it included 17,361 taxa and 270,669 records. The current version is an independent private initiative that was recreated from scratch starting 2010.

Questions & Discussion

**Question for John Irish:** There must be some kind of route for a more sustainable platform for the Namibian database. What’s the plan for the future?

**A:** It’s not ideal and sustainable at the moment, I agree. No plans at the moment but I know some plan has to be made and I’m open to possibilities.

**Question for John Irish:** What’s Namibia’s plans for joining GBIF? That would help with the sustainability aspect.

**A:** GBIF only works with government as far as I know, but I don’t work for the government so I can’t answer that question.

**Fatima Parker-Allie comment:** The GBIF Africa node can look at opportunities for how to take that forward.

**Question for John Irish:** How is data getting added to the site? Is it growing?

**A:** Yes it’s growing. Whenever the one person who works on it has time! This varies a lot.

**ACTIONS:**

- Take steps to get Namibia to plug into the GBIF Africa node.
Goal: To determine what the challenges are with sharing data with SANBI

Structure - Delegates were divided into 3 groups to discuss the following databases: SPECIFY, BRAHMS, and Red List.

DATABASE CHALLENGES highlighted in groups:

SPECIFY

- Capacity
- Data issues (we need to train scientists)
- In need of dedicated data managers
- Standards
- Intellectual Property (IP) to be protected
- Sensitive data
- Funding
- Taxonomic issues (lots of outdated records)
- Large quantities of data

BRAHMS

- New POSA is slow in terms of access
- Small herbaria do not see the need for a data manager
- Issues with regard to equipment: is it going to stay at the herbaria?
- How to link to the national checklist
- Structure of database
- Have to look at new table form

Red List

- How to do Red List assessment
- Who do we contact to do the Red List assessment?
- Sometimes there is a mismatch between fieldwork information and the data that is required for the Red List assessment
• Updating mailing system for data changes
• The data gathered is at different scales. Therefore, it is difficult to carry out the assessment
• Most assessments are based on what you can get at the time

BUILDING THE BIODIVERSITY INFORMATICS AFRICA AGENDA
SUPPORTING THE MOBILISATION OF BIODIVERSITY INFORMATION IN AFRICAN COUNTRIES

The SANBI Regional Engagement Strategy – Fatima Parker-Allie (SANBI)

Abstract:

Africa is one of the most megadiverse continents in the world. Here biodiversity plays a critical role in sustainable development, provides vital ecosystem services and is one of our greatest regional assets. The benefits of biodiversity are crucial to key economic sectors (i.e. forestry, agriculture, fisheries, tourism, health and energy) and to providing solutions to sustainable development and poverty alleviation on the continent. The management of our natural assets and the information related to this are crucial. Ultimately, we cannot manage what we don’t measure. The SANBI Regional Engagement Strategy has been developed to guide SANBI’s efforts in the region in support of national and regional priorities for biodiversity management. It provides a framework for the implementation of biodiversity priorities in the African region, as opportunities for collaboration on the continent are growing, due to an increase in emerging economies and investment on the continent. This provides an ideal opportunity for SANBI to consider its strategic role in the African biodiversity research, management, conservation and policy landscape. This regional engagement strategy identifies five strategic priority areas, to advance SANBI’s efforts across the value chain, to support the generation, management and use of biodiversity information for conservation, decision-making and sustainable development in Africa. The African continent is alive with opportunity and over the next five years, SANBI as an organisation will be expanding its efforts, across the Biodiversity Science and Policy Branch, to engage actively in this exciting evolving regional landscape.

Presentation summary:

• What is our role within the continent?

• We have strengthened our connections with other governmental departments to achieve this

• 5 strategic objectives
We had to ask ourselves hard questions – how engaged are we and how much do we want to be? What should be the scale of our impact? Do we have the capacity?

- Need to address transboundary issues
- Facilitate knowledge exchange
- Cooperation is cost-effective

What are we doing?

- Strategy 2: Encouraging efforts to get data onto GBIF (both historical and current – field trips etc.)
- Strategy 4: Play a role in coordinating continental processes and events. RLA/Sci authority, IPBES, Coordinate Africa and BHL.

Highlights

- SABONET (Southern African Botanical Diversity Network)
- Progress with botanical gardens with Botanic Gardens Conservation International (BCGI)
- NCAP workshops to strengthen capacity for assessments/Mapping Biodiversity Priorities
- Mobilising policy-relevant biodiversity data. African Biodiversity Challenge (ABC) project – replicate Biodiversity Information Management Forums (BIMFs).
- GBIF Africa: data mobilisation is happening in West Africa through GBIF nodes, underpinned by GBIF engagement and meetings. In 2015 SANBI-GBIF was requested to take forward Africa Coordinating Mechanism, which will support data science policy interface.
- More and more African countries are publishing data through capacity development initiatives – achieved a substantial amount of data mobilisation and grow our networks of partners.

Way forward: develop a business case for Biodiversity and Development Institute (BDI) in Africa

The African Biodiversity Challenge: An incentive-driven approach to mobilizing biodiversity data - Matthew Child (SANBI)

Abstract: Primary biodiversity data are essential to sustainable development in Africa. However, policy-makers will only incorporate such data if they are 1) accessible, 2) fit for use and 3) relevant to national development agendas. Enabling these conditions necessitates building functional biodiversity informatics networks comprising data holders, data managers and end users. Currently, only 3.7% of the records freely accessible on the Global Biodiversity Information Facility (GBIF) were from Africa, of which 49% have been published through African organisations and South Africa contributes 97% of the total. There is thus a critical need to capacitate African countries to mobilise biodiversity data. Simultaneously, as philanthropic funding is limited and volatile, national institutional demand for mobilised data, and willingness to support networks engaged with this work, must be cultivated to sustain the momentum generating and mainstreaming biodiversity information. Here we discuss...
unique project methodology, which falls within the GBIF Africa and SANBI regional strategy’s scope of work, which uses a competition format to incentivise self-organisation of biodiversity informatics networks; complements alternative funding models; and incentivises the mobilisation of policy-relevant data. We present a conceptual model of linked funding mechanisms to sustain biodiversity informatics networks.

Wildlife biodiversity for sustainable socio-economic development -
Kenneth Uiseb (Ministry of Environment and Tourism, Namibia)

Abstract:

Namibian conservation policy provides for private ownership and utilisation of wildlife resources. In the 1970s, a conservation policy was changed to give ownership rights over wildlife to private landowners. This policy change resulted in an increase of wildlife numbers on private land, so that wildlife is being adopted as an important form of land-use; changing wildlife’s status from a liability and competition to livestock to a valuable asset. After Independence, in 1996, conditional rights over wildlife were granted to inhabitants of the communal land. More than 50% of all wildlife is found on private and communal land outside of formally protected areas. The increase in wildlife numbers outside of the protected areas was a result of economic value gained by the wildlife because of the enabling policy framework. To ensure that the growing wildlife based industry remains sustainable, and meet the conservation objectives, the Ministry regulates the use of wildlife through the permitting system. Permits for various forms of wildlife utilisation, e.g. shoot and sell, live capture, transport, trophy hunt, import and export are all manually issued and are filed manually, including the report-backs on the permits. Valuable information on wildlife utilisation that may significantly influence wildlife utilisation policy and trigger targeted research and monitoring remain inaccessible in files stored in cabinets. Such wildlife utilisation data files dates back to 1975. There are over 4 000 private farms and 82 communal area conservancies whose wildlife utilisation data is available in a hardcopy form in the Permit Office. The biodiversity data mobilisation project aims to transform currently inaccessible biodiversity data for use to inform conservation policy while at the same time bringing more biodiversity data in the public domain.

Presentation summary:

Namibia Vision 2030: Industrialised country with decent living standards. For Namibia to capitalise on natural resources for sustainable development.

- They have wilderness areas and abundant wildlife population/uncontaminated fish and meat products/value traditional knowledge – competitive advantage.
- Various forms of conservancies (freehold and community) with different management systems – but all try to integrate biodiversity conservation and utilisation.
- 45% of land under some form of conservation management.
- Wildlife numbers same as cattle
- Wildlife numbers have been increasing each year (biomass increased from 8-29% between 1972 and 2009) – demonstration of having good policies.
- Government investment stable but contribution to GDP increasing in community based management areas.
- Wildlife has better returns in arid areas.
- Mobilisation project:
  - Online permitting system then created automated data capture and storage system
  - All hunter profile info captured too
  - Facilitate data mobilisation at other institutions (museum, herbarium etc.)
- Going to cooperate with John Irish and National Biodiversity Database – strong motivation and self-organisation occurring already.

- Potential benefits:
  - Use data to influence economic policies
  - Mainstreaming biodiversity into all spheres of society
  - Promoting climate resilient indigenous production systems
  - To inform EIAs. They have rich mineral deposits that mining companies are after – need data to mitigate.

---

An online resource for the flora of the western Indian Ocean region – Benny Bytebier (UKZN)

Abstract:

The Madagascar and Indian Ocean Islands Hotspot harbours some 15 000 plant species, 12 000 of which are endemic. The flora is only partially documented and floristic treatments are incomplete. To extend the botanical understanding of the region, digital access to locally deposited herbarium material is critical. The Biodiversity Project of the Indian Ocean Commission aims to empower local herbaria by creating a regional network and encourage all participants to actively disseminate information about their flora. The Botanical Research and Herbarium Management System (BRAHMS) was chosen as a common application for data management and exchange. BRAHMS offers efficient, fast and low cost image and data capture methods appropriate for the region. As a pilot project to digitising all the herbarium specimens deposited in the region, we focused on the large family Orchidaceae. Seven regional herbaria (DBEV, HKM, MAU, TAN, TEF, REU, SEY) imaged and digitised all their orchid specimens, which are now available on a common web platform (http://herbaria.plants.ox.ac.uk/bol/swioorchids). The availability and collation of digital herbarium specimens will lead to improved curation of the regional herbaria, and will allow for easy and useful outputs such as species reports, checklists, field guides and maps. It will facilitate research activities in systematics, evolution, pollination biology and conservation.

Presentation summary:
• Online resource for western Indian ocean
• Madagascar and Indian Ocean island biodiversity hotspot.
• Constraints for all country: incomplete botanical knowledge, limited local capacity to document and disseminate.
• High level of endemism in all countries.
• Only recently started a herbarium in Comoros.
• Madagascar – 10 613 plant species with high level of endemism.
• Institutions came together to create a virtual herbarium: digitised specimens, images, freely available, software to allow for user-friendly searches, using BRAHMS
  o Network of virtual herbaria each to establish their own BRAHMS database
  o All Western Indian Ocean (WIO) herbaria are linked in a common portal to aggregate search queries.
• Pilot project: Orchids
  o Expand to other families
  o Booklet on common orchards
  o Expansion of common web portal
  o Site visits to different herbaria to trouble shoot BRAHMS database setup
  o Third meeting in Madagascar

Microbes are biodiversity too: the African Soil Microbiology project – Don Cowan (UP)

Abstract:

The African Soil Microbiology (AfSM) project is an USAID-funded collaborative project focussing on sub-Saharan Africa and aimed at the acquiring the first-ever continental-scale survey of soil microbial diversity. With researchers from ten other partner nations scattered across the SADC, east and west African regions, we will undertake a low resolution phylogenetic survey (1 000 sample sites) of soil microbial communities from approximately 15% of the sub-Saharan land area. Sampling and data collection strategies have been designed and will be implemented through the second and third quarters of 2017. Metagenomic DNA from all samples will be extracted using a standard methodology and subject to MiSeq sequencing of 16S (bacterial) phylogenetic markers. With the involvement of researchers from each of the partner laboratories, the phylogenetic datasets will be analysed in the context of regional climate, soil, biome, physicochemical and land-use parameters. Among other objectives is a long-term aim to develop a new soil health metric, which includes metagenomic data. This project is a first for the African continent and is at the forefront of similar national and regional landscape-scale biodiversity surveys that are currently being planned or initiated in other continents.

Presentation summary:
• No comprehensive survey of soil microorganisms anywhere in Africa. Nothing was known previously.
• Aim: ‘low-resolution’ microbial community survey across Sub-Saharan Africa.
  o Interpret community composition in context of other datasets in terms of soil properties and macro-environmental parameters.
  o 10 nations participating. Need 1000 samples. Sample allotment defined by land area.
  o 50 km spacing between samples – wonderfully naïve, but exciting (should be doing 1 km spacing).
  o Going to get representative sample from major SA biomes.
• Processing is metagenomic DNA extraction. Phylogenetic assessments.
• Data interpretation:
  o Phylogenetic
  o Diversity analyses
  o Correlates with environmental variables
  o Identification of biomarkers (differently abundant taxa) / economically important taxa like cyanobacteria
  o Identification of core taxa
• Identification of microbiome fingerprints within regions.
• Future work: increase coverage, increase survey resolution, include fungi and viruses, and extract functional capacity data – soil health metrics.

QUESTIONS & DISCUSSION

Question for Don Cowan: What layer of soil will be collected and how representative will the samples really be?

A: Sampling strategy designed to avoid plants and to sample more open areas to minimise rhizome interference. However, they won’t know the extent of the confounding variance until phylogenetic studies. Soil sample resources will be open to other researchers.

Question for Don Cowan: Is there a long-term storage plan for the samples?

A: They are going to buy a freezer for soil samples. Beyond life span of centre – who knows…?

CLOSING...

JEFF MANUEL

The conference theme was natural science collections data management and efforts to build momentum in digitisation. There was also a focus on an enabling environment as well as how DEA is linking this work into international reporting (policy interventions). And how we as a community are linking into the rest of Africa.

• We can get more technical in future forums.
• Heartening to see much engagement in social media and networking happening – conference epitomised by “together we can do more” by Yonah Seleti.
• Many thanks to the Department of Science and Technology (DST) and SANBI
• SANBI has kept the Forum going for over 10 years and it intends to continue.