A NEW ERA FOR SPECIMEN DATABASES AND BIODIVERSITY INFORMATION MANAGEMENT IN SOUTH AFRICA

WILLEM COETZER, OFER GON

South African Institute for Aquatic Biodiversity, Somerset Street, Grahamstown, South Africa.
Email for correspondence: w.coetzer@saiab.ac.za.

MICHELLE HAMER

South African National Biodiversity Institute, Cussonia Ave, Pretoria, South Africa

FATIMA PARKER-ALLIE

South African National Biodiversity Institute, Rhodes Drive, Cape Town, South Africa

Abstract. – We present observations and a commentary on the inherited legacy and current state of biodiversity information management in South African natural history museums, and make recommendations for the future. We emphasize the importance of using a recognized database application, and training and capacity development to improve the quality and integration of biodiversity information for research.

In the last decade, biodiversity information in specimen databases of natural history museums has seen renewed interest and much innovation and development (Bisby 2000, Soberón and Peterson 2004, Johnson 2007, Peterson *et al.* 2010). Biodiversity Informatics has been defined as 'application of informatics to recorded and yet-to-be discovered information specifically about biodiversity, and the linking of this information with genomic, geospatial and other biological and non-biological datasets.' The mission of the Global Biodiversity Information Facility (GBIF[1]) is to 'facilitate free and open access to biodiversity data worldwide via the Internet to underpin sustainable development.' As of January 2012 the GBIF data portal provided access to >317 million primary biodiversity data records.

By January 2011 ~7.1 million records on the GBIF data portal were contributed by the South African Biodiversity Information Facility (SABIF), to which natural history museums in South Africa contribute their biodiversity information. These records originated from 8 South African data providers (mostly natural history museums) and 14 collections. Eighty per cent of the records were unvouchered occurrences, mostly observation records from the South African Bird Atlas Project, contributed by the Animal Demography Unit of the University of Cape Town. In January 2011 the digital records of

approximately 26% (Hamer 2011) of vouchered specimens in South African zoological collections could be queried through the SABIF Data Portal or the GBIF Data Portal.

The vast majority of information about South African biodiversity, which is relatively well sampled (Figure 1), originates from South African natural history museums and the South African Bird Atlas Project (Table 1).

Biodiversity information, including specimen records from natural history collections, is used in:

- Biodiversity monitoring (e.g., Reyers and McGeoch 2007);
- Bioregional planning (e.g., Smith and Wolfson 2004);
- Identifying and categorizing threatened species (e.g., Tweddle *et al.* 2009);
- Understanding the impacts of global change on biodiversity (e.g., Skelton and Coetzer 2011, Cherry 2009, Skelton *et al.* 1995) and developing mitigation strategies;
- Informing sustainable harvesting programs;
- Control of alien, invasive species (e.g., Foxcroft *et al.* 2009) and disease vectors;
- Environmental impact assessments.

Ecological niche modeling (Phillips *et al.* 2006) is a productive research area that relies heavily on high-quality biodiversity data, especially with respect to taxonomic precision and precision of georeferencing. Maintaining such

---

[1] http://www.gbif.org/.

high-quality data requires a well-designed and well-managed relational database and application, tailor-made for biodiversity information. Perhaps the most practical use of biodiversity information is in systematic conservation planning, to identify areas that need to be protected for the persistence and spatial continuity of genetic variability (populations), species, communities or ecological services (Nel *et al*. 2011). The biodiversity community has recently drawn extensively on biodiversity information held by South Africa's natural history museums for important, national biodiversity projects. Examples of these are the Alien Zonation Project (NEM:BA, 2004), the National Freshwater Protected Areas Project (Nel *et al*. 2011) and the National Spatial Biodiversity Assessment (Reyers *et al*. 2007).

1-9
10-99
100-999
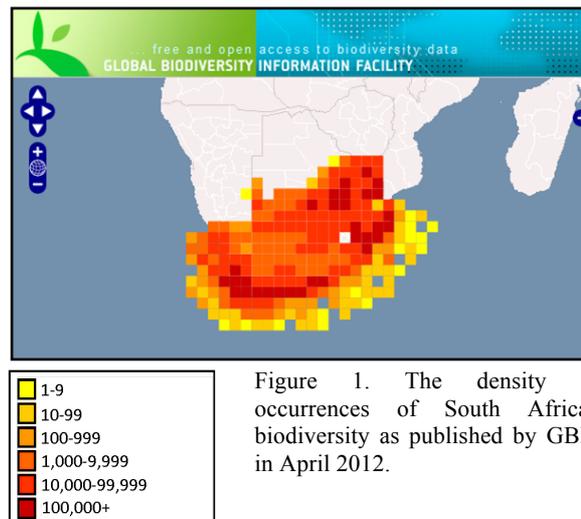1,000-9,999
10,000-99,999
100,000+

Figure 1. The density of occurrences of South African biodiversity as published by GBIF in April 2012.

Table 1. Numbers of specimen occurrences of South African biodiversity contributed to GBIF as of April 2012. Numbers in parentheses refer to vouchered specimens.

Country	Georeferenced occurrences	Georeferenced as % of total	Not georeferenced
Argentina	14	<0.05	0
Australia	3,676	<0.05	2,395
Austria	775	<0.05	4,352
Belgium	4,003	<0.05	3,406
Canada	1,501	<0.05	1,019
Colombia	15	<0.05	0
Denmark	1,284	<0.05	319
Estonia	6	<0.05	139
Finland	733	<0.05	1,179
France	612	<0.05	24,754
Germany	35,604	0.05	21,403
Hong Kong	2	<0.05	0
India	236	<0.05	0
Japan	6	<0.05	1,556
Mexico	56	<0.05	176
Netherlands	2,056	<0.05	6,201
New Zealand	0	<0.05	105
Norway	376	<0.05	0
Poland	529	<0.05	764
Portugal	0	<0.05	38
South Africa	7,115,071 (2,295,813)	98.3 (94.8)	638,843
Spain	132	<0.05	531
Sweden	1,643	<0.05	36,594
Switzerland	2	<0.05	16,014
Taiwan	1	<0.05	0
United Kingdom	2,166	<0.05	37,773
United States	69,496	1	71,476
TOTAL	7,239,995 (2,420,737)	-	869,037

As the ultimate, verifiable source of much biodiversity information, natural history museums are now not only responsible for the curation, preservation and management of collections of physical specimens, but also for capturing, managing and disseminating accurate, precise,

and current biodiversity information in the form of up-to-date species names and specimen records. Similarly the perception of a specimen database is changing from that of a pure collection management tool to include the concept of a 'biodiversity specimen database': a repository of useful research data relating to occurrences of biological species in spatial, environmental and temporal contexts. As the former, the database is used independently of other organizations' specimen databases, in an inward-looking fashion (e.g., Coetzer *et al.* 2009), but as the latter, the information from different databases and organizations needs to be integrated in interoperable systems, to discover and analyze patterns in biodiversity, and to detect changes in these patterns, including those caused by global change. In this article we consider some of the challenges faced by the South African biodiversity community with respect to biodiversity information management, and emphasize the importance of adopting a more systematic and rigorous approach to formal biodiversity information management in South African natural history museums.

We believe that training and skills-development will be central to any strategy that seeks to improve biodiversity information management in South Africa. In the last decade, we have witnessed the development of a staggering array of standards, recommendations, products, tools, initiatives, and collaborations focusing on biodiversity informatics. We should not postpone introducing young technical staff into this new world any longer.

Information management is not the only challenge faced by natural history collections in South Africa. In 2010, a survey of zoological collections was undertaken to assess the state and sustainability of collections. The report representing the initial outcome of the survey (Hamer 2011) found that collections could be consolidated to achieve a critical mass of staff and economies of scale. There can be no doubt that any investment in biodiversity information management in museums will have complementary positive spin-offs for sustainability in curation and collection management, especially at this critical time.

COORDINATED BIODIVERSITY INFORMATION MANAGEMENT IN SOUTH AFRICA

Drinkrow et al. (1994) were among the first to highlight the need for greater recognition of the value of South African natural history collections in biodiversity research, specifically the need for coordinated biodiversity database management through a national biodiversity network. In the late 1990s, such an initiative was launched. Named BioMAP (and later renamed SA-ISIS), it was the first coordinated program to capture, collate and disseminate South African biodiversity information, including specimen records from natural history collections.

SABIF was established as a program under the management of the National Research Foundation (NRF) in 2003, following the signing of a Memorandum of Understanding with GBIF. In 2006, SABIF was incorporated into the South African National Biodiversity Institute (SANBI). As a node of GBIF, the objectives of SABIF are similar to those of GBIF: to mobilize biodiversity data, provide a data-sharing platform, promote data standards and tools, and develop capacity in, and raise awareness of, biodiversity informatics.

Progress in digitizing South African natural history collections can be traced back almost two decades (Gon and Wertlen 1996), but the recent development of information technology has far outpaced the development of skills in museums. South African custodians of natural history collections have been unable to address this disparity, particularly in the present context of underfunded and understaffed natural history museums (Cherry 2009). This situation contrasts with the recognition of South Africa as a megadiverse country with a tradition of excellence in collection management and research in biodiversity science. In the last five years, however, a user-community of people directly involved in biodiversity informatics has been formed under the auspices of SANBI to address this need. Practitioners meet annually in the Biodiversity Information Management Forum (BIMF [2]) to exchange information on the acquisition, maintenance, sharing and use of biodiversity information. Among the many subjects discussed at BIMF meetings is the array

---

[2] BIMF: Biodiversity Information Management Forum. http://www.infoforum.org.za.

of recently developed standards and protocols in biodiversity informatics.

Despite significant achievements, most notably the successful establishment of SABIF, the state of digitization and web mobilization of South African biodiversity collections leaves much to be desired. At least 18 organizations and 63 collections exist that could potentially contribute ~6.5 million as-yet undigitized zoological specimen records to SABIF/GBIF. The quality of biodiversity information contributed to SABIF is highly variable, and in some cases substandard due to a history of inadequate control over data quality by contributing organizations, a result of inadequate human capacity, infrastructure, training, and capacity development, specifically in museums.

If we increase the magnification and examine what is happening at the individual workstation, we still see a lack of fundamental skills and inability to adopt new technology. We suggest that standardization and formalization of biodiversity information management are therefore needed in the museum and at the individual workstation. The proposed mechanism to achieve this objective should be based on the common use of a recognized database product, such as Specify6, a new version of the Specify biodiversity collections management software platform developed by the University of Kansas Biodiversity Institute. Of the ~2.8 million records in South African collections that are already digitized, only about a third are presently managed using Specify6 databases, representing ~900,000 specimen records (Hamer 2011; Table 2).

Table 2. The South African natural history collections currently using Specify6.

Organization	Collection	No. of specimen records	Using Specify since
South African Institute for	Fish	100,000	August 2001
Aquatic Biodiversity	Amphibians	500	January 2009
	Diatoms	50,000	January 2011
	Total	**150,500**	
Ditsong National Museum of	Reptiles and Amphibians	74,000	January 2008
Natural History	Birds	44,000	January 2008
	Mammals	46,000	January 2008
	Archaeozoology	2,500	January 2008
	Total	**166,500**	
Albany Museum	Aquatic Invertebrates	67,000	January 2009
	Fish	15,000	January 2009
	Terrestrial Insects	33,000	December 2009
	Total	**115,000**	
KwaZulu-Natal Museum	Diptera	54,000	September 2010
	Oligochaeta	5,000	September 2010
	Total	**59,000**	
Agricultural Research Council	Apoidea and Chalcidoidea	30,510	January 2011
	Homoptera	8,347	April 2012
	Total	**38,857**	
Iziko Museums of Cape Town	*Paleontology and Geology:*		June 2011
	Plant fossils	2,295	
	Karoo vertebrate fossils	7,748	
	Invertebrate fossils	16,694	
	Rocks and minerals	3,592	
	Biodiversity:		
	Arthropods (incl. Arachnida)	245,890	
	Invertebrates (incl. Crustacea)	48,230	
	Fish	19,826	
	Reptiles and Amphibians	13,700	
	Birds	17,048	
	Mammals	13,590	
	Total	**388,532**	
	Grand Total	**918,470**	

A PLATFORM FOR BIODIVERSITY INFORMATION IS
A PLATFORM FOR BIODIVERSITY SCIENCE

Berendsohn (2003) listed 26 software applications that could be used particularly for paleontological collections, entomological collections, botanical collections, or for many kinds of collections. An application qualified for the list if it could be used to manage specimens or observations, was available free or for purchase, was in use by at least one collection, did not require re-programming, and came with some support from the software developer. Fourteen applications could handle many kinds of collections, a requirement of most South African natural history museums. Current websites could be found for 11 of these applications, and only two could be downloaded and used free of charge. These were Specify6 (University of Kansas Biodiversity Institute) and Biótica5 (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Government of Mexico). Many of the world's largest natural history museums, including those of the Smithsonian Institution, American Museum of Natural History and the Natural History Museum, London, use commercial software (Knowledge Enterprises' Electronic Museum, or KE-Emu, is popular) or custom software, developed in-house. As recently as 2009, Naturalis, the central natural history museum of the Netherlands, was finalizing a contract with a commercial company to develop a new system in-house.

SPECIFY6 SOFTWARE

Together with its predecessor, MUSE, Specify has been developed in Lawrence, Kansas, since the early 1990s. Specify6 was released in April 2009. The Specify Software Project is funded by the Advances in Biological Informatics Program of the U.S. National Science Foundation (NSF) and has received NSF support since 1987. The Java application can be used on Microsoft Windows®, Mac OS X or Linux operating systems, and future releases will be capable of connecting to any relational database management system. Specify6 is free and open source software used by 404 collections in 43 US States and 26 countries worldwide (e.g. countries in South America and Europe, as well as India and Kenya). A complete list of the collections using Specify

software as of November 2011 may be found on the Specify Software website (http://www.specifysoftware.org).

In addition to conventional functionality expected in a collection database, such as report design for loan invoices and labels, Specify 6.4, released on 22 November 2011, included a new module specifically designed for the spatial visualization of specimen records and ecological niche modeling. Through the use of state-of-the-art web services, the Lifemapper module allows the user to visualize not only data in the local database, but also data currently served by GBIF. The need to search for, compile, collate and assemble static copies of datasets has therefore been eliminated. This illustrates that Specify6 optimally marries the concepts of collection management and biodiversity analysis in a single repository-and-workbench. An upcoming web interface will closely reflect the functionality of the desktop application, and will greatly facilitate the management of off-site collections and the dissemination of information.

THE MUSEUM DATA MIGRATION PROJECT

In early 2009 the first author began to clean the specimen data of selected collections in four of the country's large natural history museums, and migrate the data to Specify6 databases that were to be installed in the museums. Funding was secured from SABIF for a pilot project involving Ditsong National Museum of Natural History, Albany Museum, KwaZulu-Natal Museum and Iziko South African Museum.

The objectives of this work were to:

- Clean legacy biodiversity data and migrate the data to Specify6;
- Install the customized Specify6 databases in the four museums;
- Train museum staff to use and manage the databases;
- Train museum staff to export formatted data to contribute to SABIF;
- Stimulate interest in, and understanding of, and develop skills for, biodiversity information management in the South African biodiversity community, especially zoological museum collections.

During the project every effort was made to communicate and demonstrate data cleaning and data-migration techniques, and to provide end-user training. It was during this data-cleaning and data-migration work, and during these interactions, that the insights expounded here were formed, and where a discussion and opinions among various role-players and organizations originated. While the learning curves at all the museums were steep, the Specify 6 databases continue to be used successfully in all cases.

The training workshops that were presented during the MDM project were, however, not comprehensive nor sufficient. No other training program exists, across the various museums, in the use of a particular biodiversity database application. The reason for this is that very few collections use a particular database product; many collections relying on flat Microsoft Access tables or Microsoft Excel spreadsheets exhibiting a wide range of expertise in design (Hamer 2011). There is much enthusiasm in the South African biodiversity information community, including within SANBI and SABIF, to design a sustainable biodiversity-database training program. At the very least there are now five museums using a particular database application that did not do so before the MDM project, and it can be argued that any database training that is conducted will therefore not be merely theoretical but will actually make a difference to existing collections and practices, and have a more meaningful and lasting effect.

THE SOUTH AFRICAN BIODIVERSITY INFORMATION LEGACY

Evidence suggests that the erosion of data quality and data integrity observed in some databases is directly attributable to the common practice of indiscriminately importing legacy Disk Operating System databases (e.g. DBase, PC File etc.) into Microsoft Access files (usually as single tables), and the use of these with no user-interface, or with a poorly designed user-interface. Among many symptoms, arguably the worst symptom of this disease is the truncation of text fields, or perhaps the formatting of date fields as text, caused by a complete lack of design or planning, and possibly the result of inappropriate manipulation by unskilled people. The tendency of these Microsoft Access files to multiply is also well known, with current information becoming distributed among the copies of files. Microsoft Access files can easily become corrupted. The maximum file size of a Microsoft Access file is 2GB whereas that of a MySQL file is 8TB. There is ample evidence of all these afflictions, and more, in the databases of our natural history museums.

Arguably the Pretoria Computerized Information System (PRECIS) of SANBI, designed in 1974 (Morris 1974) and now considered technologically out of date, is nevertheless capable of far better data validation than the hundreds of duplicated Access files, with scant validation rules, of our zoological museums. In 1974 information was managed according to strict system constraints that users accepted implicitly–they had no choice. The ill-considered and inappropriate use of Microsoft Access in South African natural history collections, however, probably since the early 1990s, has created a need for a large amount of data cleaning, validation and restoration of data integrity, even before migration to an appropriate platform can begin. This work will require highly skilled database analysts who understand biodiversity information–a domain that is attractive to researchers in applied computer science and ontology engineering due to its complexity. It is unclear who will do this work, how much it will cost, who will pay for it or how long it will take to complete. Yet this work is urgently needed to bring South Africa's collections, biodiversity information, information systems, analysts and users up to date, and to allow researchers to more easily and effectively use the biodiversity information.

There is a risk of losing biodiversity information, especially in South Africa where there is a shortage of appropriate skills and the future of some natural history museums and collections is in question (Hamer 2011). This risk has been recognized by GBIF, which has initiated a global program (ReBIND [3]) to rescue biodiversity data at risk of being lost because they are neglected in outdated systems and are not properly documented.

---

[3]Biodiversity Needs Data. http://rebind.bgbm.org/.

THE SCALE AND INTEGRATION OF BIODIVERSITY DATA

Presently the manual collation of data, usually in *ad hoc* spreadsheets, is accepted by many as the way to digitize specimen information. Once the species- or specimen list or analysis is published, however, the 'flat', static dataset resulting from the short-lived research project gets copied and distributed (unless it is lost), and the copies get edited in an uncoordinated fashion. This perpetuates the cycle of poor information management practices, and begets poor data quality, which makes these data inaccessible in the future. Rather, data need to be captured in a central museum information system designed for managing natural history specimens and collections, which can accommodate the needs of a curator as well as the needs of a biodiversity scientist. Thereafter, the data need to be managed on an ongoing basis in the museum. Standardized biodiversity information for research can be simply extracted from a good collection management tool such as Specify6. In other words, data integration is an automatic consequence of good database design and management, rather than a process itself requiring design.

Keeping a national system of integrated biodiversity databases up-to-date will be greatly facilitated by the standard implementation of a recognized product, such as Specify6, in museums. GBIF and the collaboration known as Biodiversity Information Standards (BIS [4]), formerly the Taxonomic Databases Working Group (TDWG) of the International Union of Biological Sciences, have developed data standards and data exchange protocols. If these protocols are followed properly there is no need to manipulate data at the record-level as a means to integrate and distribute data. For example, Specify6 includes a field-mapping and data-export utility. All that is needed is an initial mapping of fields from the specimen database schema to the set of standard terms chosen by the data analyst, such as the 'Simple Darwin Core' (Wieczorek *et al*. 2012). Exporting updated data from the database in the future is then merely a matter of updating the data cache by clicking a button. In contrast to this, we presently see enormous investments of effort and time in manipulating individual rows and columns manually in huge datasets, even at a national scale. This practice is doomed from the start because it relies on the literal interpretation of data by humans rather than the computation, by computers, of the arbitrary unique identifiers that make a relational database useful and necessary. We are not harnessing the real power of the relational database if, in the final stage of the process, we are second-guessing it or circumventing it completely. If we fail to scale up our systems to match the overwhelming tide of incoming biodiversity information we shouldn't be surprised when, in a few years, we are drowning in data and yet the data remain inaccessible.

THE USABILITY OF DATA

Like other information that is awash on the web, biodiversity information is distributed and heterogeneous, and is in need of improvement and integration. South African researchers in biodiversity science desperately need vouchered, improved, standardized and integrated, high-quality biodiversity information that originates from across the board and through time.

Well-managed, and therefore clean, biodiversity specimen information has been easy to integrate into derived analyses in projects which have become important contributions to our knowledge of the current distribution of, and threats faced by, South African biodiversity. This readiness was demonstrated by the 2006 conservation assessments of freshwater fishes conducted by the South African Institute for Aquatic Biodiversity (SAIAB) and IUCN (Tweddle *et al*. 2009). In contrast, data that have become corrupted due to poor information management practices need to be cleaned at great cost before being used or are not used at all. Cherry (2009) lamented the lack of examples of the effect of global change on South African biodiversity (see Foden *et al*. 2007 for an excellent exception). We believe that the inaccessibility of South African biodiversity information may partly explain such information shortfalls. Inadequate information management systems and practices result in data of a poor

---

[4]BIS: Biodiversity Information Standards, http://www.tdwg.org/.

quality. There is no doubt, at least in the case of the SAIAB/IUCN project, that the use of Specify software contributed to the quality and integrity, and therefore the usability, of the data.

<div align="center">

THE PERCEPTION OF THE MUSEUM SPECIMEN DATABASE AND THE RESPONSIBILITY FOR BIODIVERSITY INFORMATION

</div>

The museum specimen database has not received, and is still not receiving, due recognition as the origin of species- and specimen-related biodiversity information. As individual researchers come and go, datasets tend to become more idiosyncratically designed, more fractured, less consistent, separated from the voucher specimens and more isolated from one another. For this reason, paradoxically, information management has become more difficult since the 1970s because there is now no control over how individuals manage or mismanage information. A useful thought experiment is to imagine leaving the responsibility for an organization's financial data to the whims of individual employees. Unlike budget projections and actual expenditure figures, however, biodiversity information is not only complex and interesting, but also remains useful after three years (provided that it is managed well), and its value increases with time, even after, and especially when, the voucher specimens have dried up or been reduced to dust by museum bugs despite our best efforts. Biodiversity information management in South Africa needs to be formalized and recognized as a profession, or we face a future of ever-more-haphazard datasets and practices, and our return on investment will be in glitz (a 'Facebook of Biodiversity'?), and not in data, information, research, knowledge, conservation policy or conservation action.

Vouchered biodiversity information originates in the museum. As long as museums exist, biodiversity information will continue to originate in museums. As the origin of the information, museums are the only places where the information *can* be created and managed in parallel with the physical specimens that it describes and documents. Anything that happens to the information downstream from the museum can only be to facilitate the discovery, transmission, and use of the information.

The challenges we face in capacity development are not insurmountable if we recognize the role of the biodiversity information manager in the museum, a competency that is presently missing between users and senior management, or between users and the network administrator or IT support staff.

RECOMMENDATIONS AND CONCLUSIONS

South African natural history museums need to employ qualified biodiversity information managers. The responsibilities of the biodiversity information manager (a kind of a biologist) are to:

- understand the particular meaning of this museum's biodiversity information;
- maintain oversight and control over data quality in the museum;
- train museum staff to use the biodiversity database and information;
- manage and administer the biodiversity database;
- develop the museum's broader biodiversity information management system (e.g. expand the scope to include new kinds of data);
- curate and analyze the biodiversity information using appropriate technology;
- communicate with users, management and IT staff in the museum, and with the wider community.

The way we work with biodiversity information has been revolutionized in the last decade and has continued to change significantly in the last few years. Specimen records used to be manipulated exclusively by collection managers and systematists, who often became database designers out of necessity. Information management is now a specialized discipline, requiring qualified, skilled and experienced designers and analysts to work together with collection managers and expert systematists to develop and maintain sophisticated, web-enabled biodiversity information management systems. Our hope is that by publicizing the importance and vulnerability of biodiversity information, we will highlight the need to develop formal information technology skills among our young biologists, who already have some understanding of the biological and information domains. We need to train specimen cataloguers, georeferencers,

taxonomic editors, information analysts and information managers.

We propose not only that the adoption of a recognized biodiversity database product, such as Specify6, will be a solution for information management in museums, but that the very use of a recognized product will facilitate the development of capacity to manage biodiversity information in museums. For example, the developers of Specify6 have employed a language that can be used to communicate effectively about biodiversity specimens and information. Words and phrases like 'collecting event,' 'collection object,' and 'preparation' are not only a part of conceptual data modeling but are also concise human expressions of the concepts that we need to represent in biodiversity information. The wider use of this 'language' on an everyday basis in museums will allow data capturers, specimen handlers, collection managers and information managers to understand their work and communicate with each other more effectively and easily.

The subject of support, training and capacity development in collections information management and biodiversity informatics has never been more important. The emergence of relational databases in specimen collections in the 1980s was an opportunity that the South African collections and biodiversity community largely missed. Today the ability to manipulate specimen records and species records is still seen as a mystifying trick on the side-line rather than the mainstream profession that it should be seen as. Delaying the introduction of state-of-the-art tools will only delay efforts to support, train, and develop capacity among curators and collection managers. While we are not recommending blindly enforcing superficial software standardization, we believe that standardizing on an established product will greatly facilitate training, specifically through enhanced communication through the use of a 'database language' across different museums.

It's no wonder that the DINA project,[5] a collaboration of museums in Sweden, Denmark and Estonia, is choosing to stand on the shoulders of giants by adopting the Specify6 schema as a standard for developing a web-based information system to allow scientists and amateurs to manage collections and distribute biodiversity information. Not only is the Specify6 schema an integral part of the developing DINA system, but during a transitional phase the Specify6 application itself will be implemented wherever an adequate biodiversity database is lacking. South Africa has the potential to continue this trend, and become one of the first countries to adopt a national database training, capacity development and database implementation program.

Smith *et al*. (2003) investigated the value of South African natural history collections and their associated information, a subject that has been studied in detail previously (Pietsch and Anderson 1990, Allmon 1994, Davison 1994). Natural history collections and biodiversity information are highly valuable to science and society (Scoble 2010). Local practitioners and professionals ought to demonstrate this value by managing collections and biodiversity information using the right tools. Certainly, nobody else is likely to fly the flag of natural history collections on their behalf.

We owe it not only to the legacies of the thousands of naturalists and curators who built these priceless collections to look after the specimens and information, but to the research community as well. The present and future scientific research and conservation action that stand to benefit from a more rigorous approach to biodiversity information management are too important to sacrifice on the altar of absent or second-rate biodiversity database design and management. By stimulating interest in, and appreciation of, natural history collections, taxonomy, systematics, biodiversity informatics and biodiversity science through the use of sophisticated tools, especially among the younger generation, we will be able to rise to the challenge of a new era. In doing so we could simultaneously address the country's shortage of skills, particularly by training and employing skilled technicians.

---

[5] Digital Information System for Natural History Collections. http://www.dina-project.net.

LITERATURE CITED

Allmon, W.D. 1994. The value of natural history collections. Curator 37:83-89.

Berendsohn, W. 2003. Survey of existing publicly distributed collection management and data capture software solutions used by the world's natural history collections: Global Biodiversity Information Facility, Copenhagen.

Bisby, F.A. 2000. The quiet revolution: Biodiversity informatics and the Internet. Science 289: 2309-2312.

Cherry, M.I. 2009. What can museum and herbarium collections tell us about climate change? South African Journal of Science 105:87-88.

Coetzer, W., O. Gon, and P.H. Skelton. 2009. Moving into the future: Relocating the national fish collection to a new, dedicated collection facility. Collection Forum 23:1-10.

Davison, P. 1994. Museum collections as cultural resources. South African Journal of Science 90:435-436.

Drinkrow, D.R., M.I. Cherry, and W.R. Siegfried. 1994. The role of natural history museums in preserving biodiversity in South Africa. South African Journal of Science 90:470-479.

Foden, W., G.F. Midgley, G. Hughes, W.J. Bond, W. Thuiller, M.T. Hoffman, P. Kaleme, L.G. Underhill, A. Rebelo and L. Hannah. 2007. A changing climate is eroding the geographical range of the Namib Desert tree *Aloe* through population declines and dispersal lags. Diversity and Distributions 13:645-653.

Foxcroft, L.C., D.M. Richardson, M. Rouget, and S. MacFadyen. 2009. Patterns of alien plant distribution at multiple spatial scales in a large national park: implications for ecology, management and monitoring. Diversity and Distributions 15:367-378.

Gon, O., and R. Wertlen. 1996. Fishnet, a computerized database management system for the national fish collection at the J.L.B. Smith Institute of Ichthyology. South African Journal of Science 92:117-121.

Hamer, M. 2011. An Assessment of the Zoological Research Collections in South Africa: Report to the National Research Foundation (Unpublished).

Johnson, N.F. 2007. Biodiversity informatics. Annual Review of Entomology 52:421-38.

Morris, J.W. 1974. Progress in the computerization of herbarium procedures. Bothalia 11:349-354.

NEM:BA - National Environmental Management: Biodiversity Act, 2004: Draft Alien and Invasive Species Regulations, 2009. Government Gazette, South Africa 32090:3-37.

Nel, J.L., K.M. Murray, A.M. Maherry, C.P. Petersen, D.J. Roux, A. Driver, L. Hill, H. van Deventer, N. Funke, E.R. Swartz, L.B. Smith-Adao, N. Mbona, L. Downsborough, S. Nienaber, S. 2011. Technical Report for the National Freshwater Ecosystem Priority Areas project: Water Research Commission Report No. 1801/2/11. Water Research Commission, Pretoria.

Peterson, A.T., S. Knapp, R. Guralnick, J. Soberón and M.T. Holder. 2010. The big questions for biodiversity informatics. Systematics and Biodiversity 8:159-168.

Phillips, S., R. Anderson, and R. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

Pietsch, T.W. and Anderson W.D. 1990. Collection building in ichthyology and herpetology. Bulletin of Marine Science 62:957-958.

Reyers, B., M. Rouget, Z. Jonas, R.M. Cowling, A. Driver, K. Maze, and P. Desmet. 2007. Developing products for conservation decision-making: lessons from a spatial biodiversity assessment for South Africa. Diversity and Distributions 13:608-619.

Reyers, B. and M.A. McGeoch. 2007. A biodiversity monitoring framework for South Africa: progress and directions. South African Journal of Science 103:295-300.

Scoble, M. 2010. Rationale and value of natural history collections digitisation. Biodiversity Informatics 7:77-80.

Skelton, P.H., J.A. Cambray, A. Lombard, and G.A. Benn. 1995. Patterns of distribution and conservation status of freshwater fishes in South Africa. Proceedings of the Zoological Society of Southern Africa 1994 30:71-81.

Skelton, P.H., and W. Coetzer. 2011. Changing patterns of freshwater fish diversity in South Africa. Pp. 190-192 in Observations on Environmental Change in South Africa (L. Zietsman (ed.). Sun Press, Stellenbosch.

Smith, G.F., Y. Steenkamp, R.R. Klopper, S.J. Siebert, and T.H. Arnold. 2003. The price of collecting life. Nature 422:375-376.

Smith, G.F. and M.M. Wolfson. 2004. Mainstreaming biodiversity: The role of taxonomy in bioregional planning activities in South Africa. Taxon 53:467.

Soberón, J. and A.T. Peterson. 2004. Biodiversity informatics: managing and applying primary

biodiversity data. Philosophical Transactions of the Royal Society B 359:689-698.

Tweddle, D., R. Bills, E. Swartz, W. Coetzer, L. da Costa, J. Engelbrecht, J. Cambray, B. Marshall, D. Impson, P.H. Skelton, W.R.T Darwall, K.S. Smith. 2009. The status and distribution of freshwater fishes. Pp. 21-37 in The Status and Distribution of Freshwater Biodiversity in Southern Africa (W.R.T. Darwall, K.G. Smith, D. Tweddle, and P.H. Skelton, eds.). International Union for the Conservation of Nature, Gland.

Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. Darwin Core: An evolving community-developed biodiversity data standard. PLoS ONE 7:e29715.