Best Practice Guide
for Compiling, Maintaining and Disseminating National Species Checklists

Michelle Hamer, Janine Victor & Gideon F. Smith

Version 1.0

October 2012

This document was produced in collaboration with the South African National Biodiversity Institute (SANBI).

Document Control:

Version	Description	Date of release	Author(s)
0.1	First complete draft, released for public review	April 2012	Hamer, Victor & Smith.
1.0	First public final version of the document.	October 2012	Hamer, Victor & Smith.

Cover Art Credit: *GBIF Secretariat, 2012. Image by Zsuzsanna Kilian (Hungary), obtained by stock.xchng (http://www.sxc.hu/photo/1024631).*
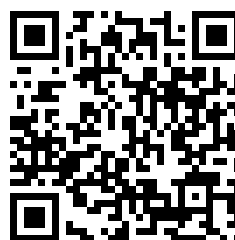
About GBIF

The Global Biodiversity Information Facility (GBIF) was established as a global mega-science initiative to address one of the great challenges of the 21st century – harnessing knowledge of the Earth's biological diversity. GBIF envisions 'a world in which biodiversity information is freely and universally available for science, society, and a sustainable future'. GBIF's mission is to be the foremost global resource for biodiversity information, and engender smart solutions for environmental and human well-being[1]. To achieve this mission, GBIF encourages a wide variety of data publishers across the globe to discover and publish data through its network.

---

[1] GBIF (2011). GBIF Strategic Plan 2012-16: Seizing the future. Copenhagen: Global Biodiversity Information Facility. 7pp. ISBN: 87-92020-18-6. Accessible at http://links.gbif.org/sp2012_2016.pdf

About SANBI

This guide has been produced by the South African National Biodiversity Institute (SANBI) through funding from the Global Biodiversity Information Facility (GBIF). SANBI serves the South African government and public as the primary statutory institution devoted to the study, conservation, display and promotion of the country's indigenous biodiversity. SANBI is tasked to lead the biodiversity research agenda, provide knowledge and information, give policy support and advice, manage gardens as 'windows' to our biodiversity for leisure, enjoyment, spiritual upliftment and education, and engage in ecosystem restoration and rehabilitation programmes and best-practice models to manage biodiversity better.

The Biosystematics Division has a staff component of 63 people spread across the three SANBI herbaria (National Herbarium in Pretoria, Compton Herbarium in Cape Town and the KwaZulu-Natal Herbarium in Durban). Staff members of the Biosystematics Division are responsible for discovering, documenting and describing the plants of South Africa, for maintaining and providing access to the collections, for identifying plants for stakeholders, and for the maintenance and expansion of the plant specimen and taxon database. SANBI was, until 2005, the National Botanical Institute. The Biodiversity Act required that the Institute broaden its focus to include all biodiversity. No checklist of animal or fungi species exists for South Africa, and this has been highlighted as a priority need by a wide range of stakeholders. SANBI does not employ animal or fungal taxonomists but it has responsibility for developing and maintaining checklists for these taxa. The checklist for South Africa's more than 60,000 animal species has now been initiated. The process of developing policies and procedures for both compiling and updating checklists, and the recent efforts to compile the animal checklists have provided a range of insights into developing species lists for a country.

The information presented in this document is based on personal experiences and opinions of the authors, on the input from the research staff of the Biosystematics Division through a workshop funded by GBIF, and on relevant documentation presented on the Internet or in the literature (which are referenced). Most of the examples used are South African, but they have relevance for other countries.

The document does not provide the technical / information technology details for checklist compilation, maintenance and dissemination because these were dealt with

in two GBIF documents on Publishing Species Checklists: Best practices[2] and Global Names Architecture Profile Reference Guide for Darwin Core Archives[3]. The current guide will also not provide detailed recommendations regarding the format or structure of checklists either for management or dissemination, because this is dealt with elsewhere by GBIF.

The main purpose of this best practice document is to provide guidance for policy and procedures relating to accessing and capturing information for national checklists, and to highlight some key issues which should be considered to promote standardization, quality and use of these products.

---

[2] GBIF (2011) Best practices in publishing checklists. Contributed by Remsen, D., Döring, M., & Robertson, T. Copenhagen: Global Biodiversity Information Facility, 20pp. Accessible online at http://links.gbif.org/checklist_best_practices.
[3] GBIF (2011) GNA Profile Reference Guide for Darwin Core Archives, version 1.0. Contributed by Remsen D.P., Döring, M, Robertson, T. Copenhagen: Global Biodiversity Information Facility, 28 pp. Accessible at http://links.gbif.org/gbif_gna_profile_reference_guide.

Table of Contents

1. Summary

National checklists are invaluable resources which serve a range of functions for research and biodiversity-related activities. Ideally national checklists should be integrated, coordinated and disseminated from a single platform.

National checklists should be compiled by expert taxonomists, but this is not always possible, and non-experts with knowledge of nomenclature, taxonomy and access to the relevant literature and to online biodiversity data resources such as GBIF can also adequately carry out this activity.

The national policy and procedures should be documented for compiling and maintaining national checklists in order to standardize and guide what is included and how this should be done.

A core set of fields for national checklists includes the seven levels of classification and the authority. Additional information for each species increases its value and credibility, but also increases the time and effort required.

The format for compiling the checklist should ensure that it is accessible to a wide range of compilers, in that it should be simple to use and easily modified. The format in which the checklist is disseminated should allow searching using different criteria, and should also enable download of datasets.

Recommended and optional fields for the checklist are as follows:

Recommended fields for checklist	Optional fields for checklist
Kingdom	Reference for original description of species
Phylum/Division	Common / vernacular name(s)
Class	Type locality
Order	Type specimen repository
Family	Type specimen catalogue number
Genus	Taxonomic note(s)
Species	Indigenous / exotic
Species name authority + date	Distribution at country level
Synonyms + reference for each	Distribution at provincial / state level
Unique identifier for species	Habitat
Source of species record (reference or website)	Threat status
Author/compiler of checklist record	Economic importance
Date of capture of species record	
Date of amendment to record	

Summary of the initial steps involved in compiling a national checklist:
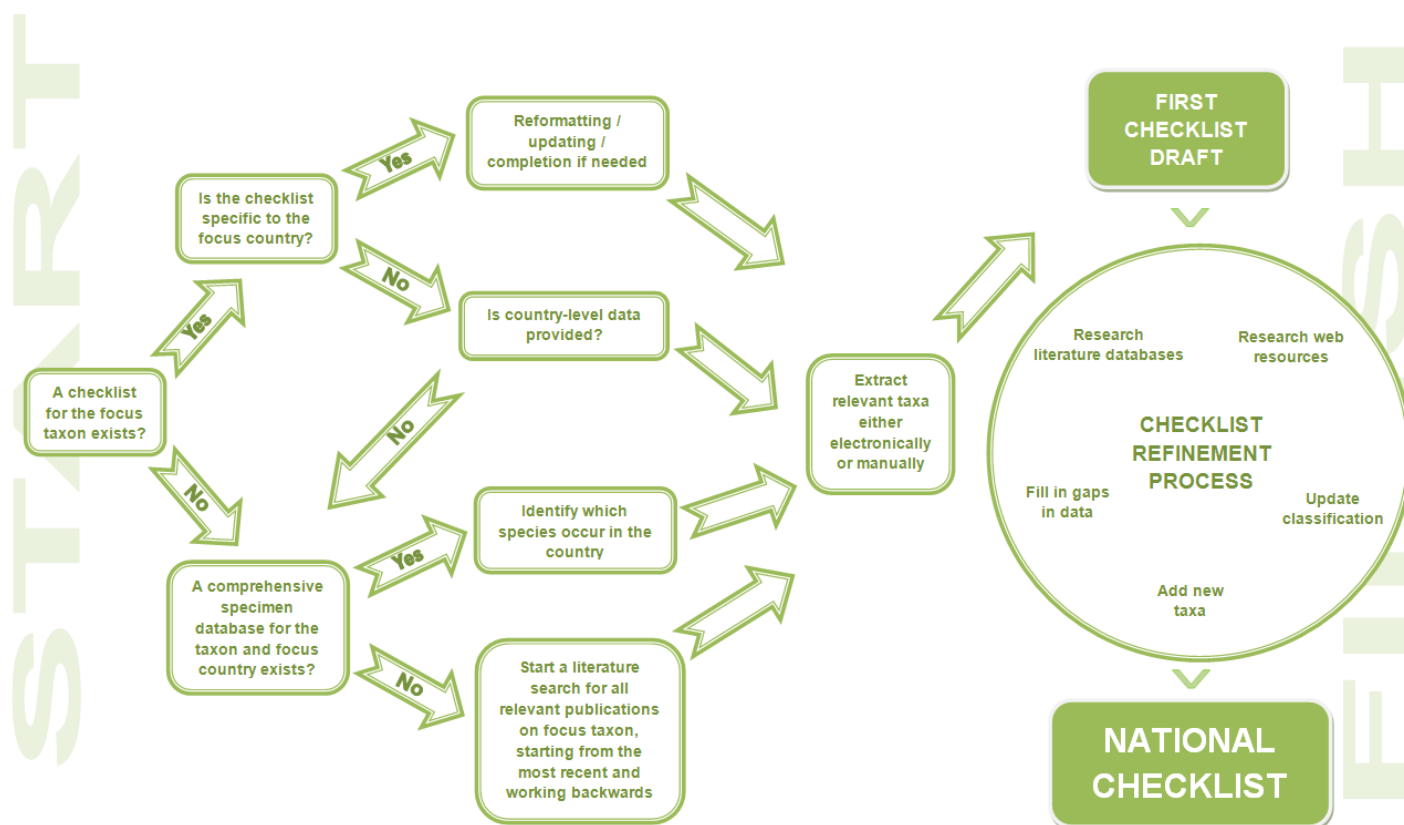
2. Background

2.1. What are national species checklists?

National species checklists can also be referred to as country-level species catalogues or inventories. In simple terms these can be described as *the list of species recorded from a country,* but they can vary from a simple list of species' names, usually scientific names, to an annotated list with various types of information associated with each species. A catalogue is usually considered to have more taxonomic and reference detail for each species and name than a checklist (see Catalogue of Life[4] as an example), and to be more authoritative because of the level of detail provided. An inventory is usually derived from a catalogue, and species are organized according to some classification system, but the taxonomic detail of the catalogue is lacking. GBIF's guide to best practices for publishing species checklists[5], which focuses on sharing taxonomic checklist information in a standard way, provides a more detailed definition for the different types of species checklists.

In this document a checklist is interpreted more broadly than as a simple list of names, and is probably more in line with the definition of a catalogue combined with an annotated checklist. This is because having a single national list, which includes several different types of information, is more practical than having a separate checklist, inventory and catalogue for a country, especially when resources are limited and species richness is high. *The objective of a national checklist should be to coordinate, consolidate and disseminate basic taxonomic and species information that is commonly required by a range of users*. The checklist should serve not only taxonomists, but also enable non-taxonomists to readily access information about species, including which scientific name to use for any given taxon, without having to understand the scientific intricacies, nuances and debates surrounding names and classifications. The requirements of the users must therefore be considered when a national checklist is designed and compiled.

One of the main requirements of a non-specialist for a particular taxon is to know which name and classification framework to use when referring to a species or other

---

[4] Catalogue of Life: http://www.catalogueoflife.org/

[5] GBIF (2011) Best practices in publishing species checklists. Contributed by Remsen, D., Döring, M., & Robertson, T. Copenhagen: Global Biodiversity Information Facility, 20pp. Accessible online at http://links.gbif.org/checklist_best_practices.

taxon (family, genus, subspecies for example). While taxonomists understand that there is no one 'correct' classification system, and that the status of taxa may vary according to which concept is applied, which characters are analysed, and how analyses are interpreted, this is not especially helpful for most users of the checklist. Probably the greatest challenge for anyone compiling and maintaining national checklists is how to ensure that the list is scientifically credible and accommodates different viewpoints, but at the same time provides the guidance required by non-specialists. A national checklist may be for a defined taxonomic group, for example the birds of South Africa, or it can be based on growth forms (e.g. the trees of South Africa), habitats (e.g. the marine invertebrates of South Africa) or functional or ecological roles (crop pests, alien invasive species); or it can be broader and cover all species, usually within at most a Kingdom (e.g. the animals of South Africa). In addition, while national species checklists are referred to here, *infraspecific taxa such as subspecies, varieties and even cultivars can also be accommodated, depending on country-specific needs, priorities and resources*.

Surprisingly few countries have compiled and coordinated checklists of species, even though they have enormous value and are often referred to as providing the basic information required for monitoring, conservation and sustainable use of biodiversity. A vast number of taxon-specific and other types of lists such as those for threatened species or for protected areas do exist and have been published or are available on the Internet, but these lists are fragmented, are generally scattered across institutions and may not publically accessible. *Ideally national checklists should be made available through one coordinated website, and one institution or organization should be responsible for the coordination of the lists*. The Netherlands[6] and Australia[7] each have a national biodiversity checklist site and several other countries have indicated an intention to develop national checklists, which is probably a reflection of the increasing recognition of the importance of such a resource.

National checklists should ideally be at as high a taxonomic level, or as inclusive as possible, for example all angiosperms, rather than fragmented at a lower taxonomic level such as by family. A more inclusive checklist, being a centralized source of information, will have a wider audience, and while there are benefits to focussing on lower taxonomic levels such as orders or families, this approach may lead to duplication, overlap or gaps and to inconsistency in what is included and in names or classification systems used when species occur in more than one list. Checklists at a

---

[6] Nederlands Soortenregister - Dutch Species Catalogue: http://www.nederlandsesoorten.nl/nlsr/nlsr/english.html
[7] ABRS | Fauna databases and online resources: http://www.environment.gov.au/biodiversity/abrs/online-resources/fauna/index.html

higher level (e.g. all animals or plants) provide a far more useful tool for monitoring diversity, and for coordinating all taxonomic changes using a standard set of principles and procedures. Compilation of the lists can be done at lower taxonomic levels by relevant experts, but they should be coordinated and made accessible through one platform.

The differences in botanical and zoological nomenclatural rules and taxonomy probably constrain the development of a checklist beyond a single Kingdom.

2.1.1. Form of the checklist

National checklists can be published in hard copy, or in electronic format on a website. National or country-wide checklists are generally not published in scientific journals, because they are usually long and do not conform to the requirements of journals, but components of national lists such as order or family level checklists may be accepted for publication. The benefit of publishing lists in journals is that they are peer-reviewed and therefore have more credibility, and they can still be incorporated into and distributed via a national checklist website. As a result of additions or changes to checklists, published versions quickly become outdated, but they still have some use as an assessment of the status of a taxon at a particular point in time. Similarly, some taxonomists favour national checklists published in book format because they represent a tangible output for which they receive credit, but the reality is that all scientific names presented in such books will still need to be checked by a user against a dynamic electronic list. Hard copy books also need to be purchased and so are not as readily accessible to all stakeholders as electronic lists, which can be quickly and freely disseminated via the web or by email.

Because of the dynamic nature of checklists, it is recommended that a national checklist intended to identify the most up-to-date status of names be provided electronically. Electronic format allows regular updating; it allows updating of components of the checklist, rather than requiring the entire group to be updated at once; it is more cost effective; and if published on the Internet in open access format, provides information to a wide range of target audiences. In addition, publication in electronic format may allow manipulation, including ordering and searching which is not possible in print format, and it may allow the presentation of species organized according to different classification systems.

2.1.2. Organization of the information

National checklists are usually organized according to a scientific classification system, with species referred to using their scientific name. Species are usually listed

alphabetically within each genus, with genera listed alphabetically within each family, but sometimes phylogenetic rather than alphabetic ordering may be used.

2.1.3. National checklists are dynamic and maintained

Compiling a list of species that have been recorded from a country and organizing them according to a classification system may sound simple, but many complexities have to be considered, associated with the description and naming of taxa and changes resulting from later revisionary research. While we cannot provide a complete review of taxonomic and nomenclatural principles and procedures, some of the main complexities relevant to compiling and maintaining national checklists are briefly mentioned here. For more detail, the reader should refer to the appropriate nomenclatural codes (International Code of Zoological Nomenclature[8], International Code of Botanical Nomenclature[9]).

There is no single classification system that is accepted by all biologists. Classification systems should reflect evolutionary relationships, but different types of data and analyses of data may suggest different groupings (relationships) of taxa. This means that any classification system should be considered only as a hypothesis and it should be recognized that there is no 'correct' or 'incorrect' classification system. Compilers and managers of national checklists must deal with the challenges of deciding which classification system to use, and how to reflect published changes to and rearrangement of taxa in the classification system.

The status of taxa and names, especially in the case of genera and species, may change for the following reasons:

A type specimen (i.e. the specimen on which the description is based) may have been described as two different species, usually by different researchers at different times. The type specimen can only have one 'correct' name, usually the first name by which it was described, and the later name is considered to be a synonym of the original name (an objective [zoology], or nomenclatural or homotypic [botany] synonym).

New combinations arise when a species is transferred from one genus to another.

Questions around the status of species include the following scenarios:

A taxonomist may refer to a specimen identified by one name, but another taxonomist may not agree with the name used because he or she believes that the specimen more closely resembles the type of a different species. This is important for a national

---

checklist, because one taxonomist might identify a species as occurring in a country from specimens identified, but another taxonomist may consider that the specimens have been misidentified, and therefore that the species may not actually be present in the country. In this case the presence of a species in a country, and in other cases the distribution of the species in the country, may be matters of opinion relating to identification and interpretation of characters.

A species may be described based on a type specimen, with other specimens also identified as the same species. Closer examination of the material may lead to the recognition that more than one species is represented, and several new species, each with its own type specimen, are then described. However, while some taxonomists may agree with the splitting of what was previously considered to be one species, others may not, and may simply consider the original species to be variable, with the new species considered as synonyms (subjective [zoology] or heterotypic or taxonomic [botany] synonyms). In this case a decision would need to be made about how to reflect the different opinions in the checklist. This type of dilemma is becoming increasingly common with DNA analyses which frequently indicate variation within a species, but the amount of genetic variation required to reflect distinct species is contentious.

There are different terms to reflect which taxa and which name associated with a particular taxon is 'correct' (in botanical terms 'accepted', and in zoological terms 'valid'). Each species can only have one 'correct' (i.e. valid or accepted) name, but the status of species can differ depending on the interpretation of the data by different taxonomists and their concept of a species.

2.2. Purpose and users of national species checklists

Regardless of the field of scientific endeavour in biological and life sciences and beyond, accurately recording and applying names to organisms are of primary and paramount importance. Once recorded and assembled into partial or comprehensive checklists or catalogues, names take on an infinitely greater significance.

Considering the wide range of users of a checklist, the purposes of creating national species checklists for organisms are:

To provide a list of the species known to occur in a country;

To provide clarity on the status of names and taxa;

To provide access to synonyms (names that are not considered to be valid or accepted) of taxa so that users can track the fate of all names and identify the 'correct' name for any organism;

To provide a standardized classificatory framework for the placement of species and higher taxa.

These four purposes are relevant for scientific publications, primary biodiversity databases, and for the organization and labelling of biodiversity specimen collections both living (in zoos, botanical gardens or fungal collections), and preserved (herbaria or museums). This implies that *all researchers, in any field, working with species and compiling databases or referring to species in manuscripts should use a national checklist*. In addition, legislators and regulators need to include the 'correct' name in documentation dealing with alien invasive, traded, cultivated, pest, harvested or threatened species.

In addition to providing the 'correct' name, national checklists can also be used as follows:

To enable the estimation of the diversity (richness) of a particular group, or overall diversity within the country;

To track changes in the diversity (through for example, the addition of newly described species) which can be used as an indicator of research effort; and

To identify knowledge gaps through flagging those higher taxa for which species information is lacking or where taxonomic and nomenclatural problems are highlighted in the list.

These uses are important for funding agencies and research managers in terms of understanding how to prioritize the allocation of resources and develop capacity for taxonomy.

An additional use of national checklists is to act as a planning and monitoring tool for:

Tracking national contributions to global programmes such as the Encyclopedia of Life[10] and the Barcode of Life[11]. A species checklist assists effective planning, for setting targets and identifying priority species/taxa to include in these programmes; and

Tracking progress in terms of representation for species or suites of species by habitat or by higher taxon for data mobilization (i.e. which species need to be focussed on for occurrence data to provide comprehensive data sets) for national nodes of the Global Biodiversity Information Facility.

---

[10] Encyclopedia of Life: http://eol.org/
[11] IBOL: http://ibol.org/

These activities would be important for national node / programme managers for global programmes.

2.3. Who compiles checklists?

Traditionally, specialist taxonomists in the fields of botany, zoology, mycology and phycology employed at herbaria, museums and universities would compile and update checklists. This is still the approach taken, but in many cases expertise does not exist within a country and this role falls to non-specialists, including citizen scientists and postgraduate students. While taxon experts are obviously the ideal choice to compile checklists for their speciality, ***anyone with a solid knowledge of taxonomic and nomenclatural principles, and access to relevant literature sources, can compile checklists***. Developing a national checklist that covers all taxa will probably require that non-specialists contribute to at least parts of the list.

3. Developing a checklist

In some countries, national checklists have a long history and have been expanded, improved and updated many times. The plant checklist of South Africa provides a good example of this situation.

3.1. History of South Africa's plant checklist: an ideal situation

One of the first efforts to catalogue the higher order plant taxa of South Africa was the 'Genera of South African Plants' (Harvey 1868) in which 1186 genera were listed. This was followed by the 'Genera of South African Flowering Plants' (Phillips 1926), in which 1645 genera were listed. The first single comprehensive list of the southern African flora including all known species was published in 1993, in which 2604 genera were published (Arnold & De Wet 1993), and this volume became known as the Black Book. It has since been replaced by, 'A Checklist of South African Plants' (Germishuizen et al. 2006).

The national plant checklist is available on the SANBI website[12], and is updated regularly. The mechanism used to update the list is the National Herbarium, Pretoria (PRE) Computerised Information System (PRECIS), a database that documents the primary collection data of all plant specimens housed in the three SANBI herbaria. The database has two main components, one of which houses the locality data from specimens collected. The second main component is taxonomic and comprises the electronic catalogue of taxon names, along with their authors, synonyms, references, and to which family they belong, for indigenous and naturalized taxa of the Flora of the Southern African region. From the latter, a checklist can be derived at any time, that will reflect the current accepted names and classifications of plants in South Africa. The website resource is derived from the taxonomic component of the PRECIS database. This illustrates a case where a comprehensive list has existed for over 100 years, and an electronic version can be compiled quickly using a database as a sound basis for the checklist, which only needs to be updated when changes are published.

The South African Plant Checklist is the exception rather than the general situation, and in most cases, lists have to be compiled from scratch, or from numerous lists for lower taxa, or from a completely outdated list.

---

[12] Plants of Southern Africa: http://posa.sanbi.org/searchspp.php

3.2. Compiling a checklist

3.2.1. Using taxon experts to compile the checklist

Taxonomists invariably develop competence in one or more families or groups. After a period of exposure to the constituent taxa, they are familiar with all the literature dealing with the taxonomy of the group, and they become experienced at recalling names of organisms and recognizing when an organism is new to science. Institutions that take on the responsibility for compiling checklists and inventories usually make these taxon-based experts responsible for creating, improving and maintaining names (accepted and synonyms) of groups they know and feel comfortable with.

A national checklist for a major taxonomic grouping (e.g. vascular plants, or all animals) can be compiled by getting specialist taxonomists to compile the list for one or more families or orders, and then combining these lists. It is most likely that there will be taxonomic gaps remaining, and these can be filled by non-specialist taxonomists, using a range of approaches and resources as detailed in the following sections.

There are two considerations when getting taxonomists to compile components of a major national checklist. Firstly, the format and content will need to be specified (section 3), and the conventions and philosophy for what will be included and which classification system will be used need to be standardized across taxa as far as possible. Secondly, grants to taxonomists may be required, so this should not be considered as a cost-free approach.

3.2.2. Compiling the list in the absence of experts

While taxonomists will use some of the resources and approaches outlined below in compiling a checklist, they are likely to have a sound understanding of existing literature, websites, and collection databases and their potential value for compiling a checklist, and they would probably have a complete set of relevant references and a comprehensive specimen database for the taxon. Non-experts, on the other hand, would need to locate and assess various sources of information.

3.2.3. Extracting a checklist from a specimen database

Natural history collections that have been databased include the names and usually some degree of classification information that has been captured for each specimen. A simple approach may be to extract the taxonomic list from such a database, at least as the basis for the full and updated list. This is especially true where the database is comprehensive, for example where it represents a major collection for a particular

taxon. Gaps will exist in the taxon list, where material of any number of species may not be represented in the collection. In addition, the taxonomic and nomenclatural data may be outdated, unless the collection and databases have been regularly checked and updated. In addition, the identity of specimens may not have been verified, so species may be erroneously included or excluded. Finally, most collections include material that is not from the country in which the collection is housed. This should be the simplest challenge to address, because non-focus country material can be excluded before the taxon list is extracted. The advantage of extracting the taxonomic list from a specimen database is that each name is associated with actual specimens, which means that the identity of the specimen can be verified if taxonomic expertise is available.

3.2.4. A major work exists, but this is broader than the country, is outdated, and is only available in print.

An example of this is the Catalogue of the Afrotropical Diptera (Crosskey 1980), which formed a sound basis for compiling a checklist of the Diptera of South Africa, but a considerable effort was still required. The text for each species (more than 16,000) included in the catalogue had to be examined to identify those species which were recorded from South Africa and these were then captured, with their associated information. All subsequent published literature which covers or includes South African taxa had to be identified and reviewed to add new species and to update any changes to the classification and nomenclature. The electronic version of Zoological Record (Thomson Scientific Inc.)[13] was used for this updating activity.

3.2.5. Global checklists exist for a particular taxon or habitat

The Catalogue of Life (CoL) may be a useful starting point for compiling a national checklist either in its entirety or for a particular taxon, but it must be recognized that the CoL list is far from complete.

There are many global efforts to compile checklists for particular taxa (e.g. various insect and plant families, fish), and for marine organisms. In several cases these checklists or databases are available on the web and are searchable by country, producing a list of species which can be downloaded. This can be an excellent starting point for these taxa. The content partners listed in the Encyclopedia of Life[14], in the

---

[13] Zoological Record: http://thomsonreuters.com/products_services/science/science_products/
[14] Encyclopedia of Life: http://eol.org/

checklist section of the GBIF Checklist Bank[15] website, the projects in the Taxonomic Databases Working Group (TDWG) [16] site and the Catalogue of Life information databases[17] provide useful links to these websites. An attempt should be made to determine the completeness of the list, how and when last it was updated and who compiled it.

When resources such as these are used to extract data, copyright restrictions and requirements must be considered.

3.2.6. No checklists have been compiled for a particular taxon.

Where no checklist exists for a particular taxon, the national list will need to be compiled from published taxonomic papers, electronic resources or a combination of the two.

One approach is to scan relevant literature to identify the most recent publication on the group, and to use the reference list from that publication to identify further relevant works. Citation and abstract databases such as the Thomson Reuters (ISI) Web of Knowledge[18], Scopus[19] and Zoological Record[20] are very useful for searching for relevant literature. Museum or herbarium libraries may need to be visited to access literature, but those with a focus on the specific taxon should be prioritized.

The Biodiversity Heritage Library[21] is a rapidly-growing resource that provides open access to the 'legacy literature' from major natural history libraries through scans of documents. It provides a search function by any taxonomic level or species name, or by country. JSTOR[22] is a similar resource and is also open access, but it covers a suite of journal titles, rather than focussing on legacy literature. Mendeley[23] is another literature database which can be searched by taxon and which provides open access to many more recent publications.

Taxonomy is not only perfectly suited to the electronic dissemination of nomenclatural, descriptive and illustrative material, but Smith & Figueiredo (2010) conclusively showed that it is equally possible to construct, virtually *de novo*, a near-perfect plant inventory for a country (or region of any size for that matter) using

---

[15] GBIF Checklist Bank: http://www.gbif.org/informatics/name-services/checklist-bank/
[16] TDWG: Projects Database: http://www.tdwg.org/biodiv-projects/projects-database/
[17] Catalogue of Life - 15th March 2012: Source databases: http://www.catalogueoflife.org/info/databases
[18] Web of Knowledge: http://wokinfo.com/
[19] Scopus: http://www.info.sciverse.com/scopus
[20] Zoological Record: http://thomsonreuters.com/products_services/science/science_products/
[21] Biodiversity Heritage Library: http://www://biodiversitylibrary.org/
[22] JStor: http://www/jstor.org/
[23] Mendeley: http://www.mendeley.com/

resources currently available on the world wide web. With resources such as GBIF[24], JSTOR Plant Science[25], TROPICOS[26], the International Plant Names Index[27] (IPNI), and several others (see Table 4 in Smith & Figueiredo 2010), 11,316 plant names were compiled for 7,296 plant taxa accepted for Angola, a south-tropical African country (Figueiredo & Smith 2008). The research required was conducted at a cost of only 1.4 Euros per name added to the checklist, which additionally makes this a comparatively affordable way of compiling a checklist.

3.3 Which classification system to use?

The classification system provides the overall hierarchical framework for a checklist.

The classification of biological organisms should reflect the evolutionary relationships between taxa and so *classifications should be considered as hypotheses that change on an ongoing basis,* mostly as a result of the application of new technologies to explore DNA and to analyse molecular data. There are often several different higher classification systems in use at any one time. None of these is 'correct' or 'incorrect', and it should be accepted that individual taxonomists will favour one system over another which will result in contention over which is the better system to use at a national level.

In many cases institutions or taxonomists select some components of an established classification system, but not others which they replace with systems from more taxon-specific publications. This approach can create problems when comparisons of richness or diversity need to be made with other countries for a particular taxon, and it will be problematic for data to be contributed to or extracted from one of the larger species databases.

For national checklists it is therefore recommended that *a generally accepted classification system be used, rather than one composed of components of systems,* or derived from a range of fragmented publications. The system for a national checklist can be decided on through a workshop or discussion involving the relevant taxonomists. A useful approach is to consider positive and negative aspects of different classification systems in terms of stability, extent of use by major collection institutions or global programmes, and the implications in terms of organizing the national flora or fauna. If necessary a national poll of taxonomists can be carried out

---

[24] GBIF: http://www.gbif.org/
[25] jstor plant science: http://about.jstor.org/content-collections/primary-sources/jstor-plant-science
[26] Tropicos: http://www.tropicos.org/
[27] IPNI: http://www.ipni.org/

to select a classification system. For a national checklist, probably the most important criterion is stability, i.e. it is unlikely to change with every new analysis, or to be completely rearranged on a regular basis. ***One of the classification systems used by the larger global programmes such as the Encyclopedia of Life***[28] ***and Catalogue of Life***[29] ***are recommended for national checklists***.

The ***classification system used can be reviewed when new systems become established,*** but the implications in terms of feasibility and resources required for migrating from one system to another should be considered. Review should be within a set time, preferably not more frequently than every five years to avoid continual change, which could result in loss of credibility among users of the checklist, and additional costs related to capacity to make the changes. Reviews do have to be done, however, since the classification system used must be valid.

If resources permit, web-based checklists can be presented using the classification system promoted as the 'national standard', and species lists can also be presented in alternative classification systems for the use of specialists. This is done in the Encyclopedia of Life and several other smaller checklists. The maintenance of multiple classification systems should, however, be seen as a useful add-on rather than as a core component especially for many of the biodiversity-rich countries with limited capacity.

A problem which will be encountered is allocating taxa included in one classification system to another one which may not include such taxa. It may be simplistic to include such taxa in the checklist with a note explaining the source and the reason why they have not been allocated to a particular higher taxon in the classification system used. The only real way to resolve such problems is through taxonomic study, but this is beyond the scope of the normal checklist compilation process.

3.4. A stepwise approach to compiling a national checklist

There is no 'recipe' for compiling a checklist and the details of how a checklist should be or can be compiled will depend on the resources available, which are usually related to the taxon being dealt with and the country for which the list is being compiled. We suggest some steps that may be a useful guide.

---

[28] Encyclopedia of Life: http://eol.org/
[29] Catalogue of Life: http://www.catalogueoflife.org

Determine whether there is an existing checklist for a particular taxon (family, order, class or other higher level). Interrogating citation and abstract databases, Google Scholar, or even Wikipedia is a useful way to start this search.

If a checklist does exist, access this through the website, a library or the Biodiversity Heritage Library[30].

Determine what format an existing checklist is in, and whether it could be used to extract the relevant data without having to recapture it.

Remsen *et al*. (2012)[31] described a pilot project in which an electronic checklist was extracted from a Word document, and data were reformatted into an electronic database for integration into GBIF, and into an Excel spreadsheet that was included as an appendix with the electronically published paper. In the case where a checklist exists as a Word document, or in a specimen database, the procedure described can be used to avoid recapturing the data.

There are various tools that can assist with finding and extracting names in PDF or other electronic documents [32]. These can be useful and save time, but the decision about how the name fits into the checklist still needs to be made by the checklist compiler.

If no published checklist exists, determine whether there is a comprehensive specimen database which covers the country or region for the focus taxon. The national GBIF node, or if this does not exist, the global GBIF website, is the most logical source of this information. Collection facilities such as herbaria and museums may also be useful. Records from the focus country can be filtered, and the taxonomic component of the database extracted. The GBIF data portal[33] provides such a download option. In many cases the identification of the specimens will not have been checked by an expert, and the specimens may be misidentified, or the name may have changed, so lists based on specimen databases should be treated with some caution.

Lists from neighbouring countries or broader regions that have checklists can also be useful because the species information may mention the occurrence of species in other countries. From this point, the literature will need to be used to identify gaps and

---

[30] Biodiversity Heritage Library: http://biodiversitylibrary.org/
[31] Remsen, D., Knapp, S., Georgiev, T., Stoev, P. & Penev, L. (2012) From text to structured data: Converting a word-processed floristic checklist into Darwin Core Archive format – Pensoft. Accessible at http://www.pensoft.net/journals/phytokeys/article/2770/
[32] Taxon name processing tools: http://code.google.com/p/taxon-name-processing/wiki/nameRecognition

check the validity and status of the names and higher classification used. Citation and abstract databases are the most efficient way of finding relevant publications.

Use the checklist or extracted taxon database as the basis, and then use various literature databases or search engines to find later publications on the taxon that may have added species, or changed the taxonomy of existing taxa. It is advisable to get the full reference for checking the taxonomic information.

If there is no existing checklist, all publications dealing with the particular taxon, and that include or focus on the particular country, will need to be sourced and accessed so that relevant species and classifications can be extracted and compiled into the national checklist.

Establish the classification framework for the checklist. An available source is the Catalogue of Life Annual Checklist[34].

Start with the most recent publication, capturing the genera and species and relevant associated information for each taxon covered. Details of what information to capture for each species is covered in section 3, and guidance for the format in which it is captured is provided by GBIF (2010a).

Verify the nomenclatural status of each name included in the list using a nomenclator (Index Fungorum[35] and International Plant Names Index[36] or ZooBank[37]). If the name does not appear in the electronic nomenclator, as is the case for most animals, it will need to be tracked through the literature.

Source additional information about each accepted or valid name, especially synonyms and references. The African Flowering Plants Database[38] is an example of a resource that provides this kind of information, but there are many others on the web. In the absence of an appropriate web-based resource, relevant literature will have to be reviewed.

Ideally a taxonomist with specialist knowledge of the relevant higher taxon should review the list before it is disseminated. Obviously if the list has been compiled by a specialist, this step is not necessary.

For plants and fungi, a different sequence of activities may be more efficient because of the comprehensive nomenclators that exist for these groups. A species checklist can

---

[34] Catalogue of Life: http://www.catalogueoflife.org/
[35] Index Fungorum: http://www.indexfungorum.org/names
[36] IPNI: http://www.ipni.org
[37] ZooBank: http://zoobank.org/
[38] African Flowering Plants Database: http://www.ville-ge.ch/musinfo/bd/cjb/africa/index.php

be extracted from a specimen database or GBIF for a country, and each name in the list checked against the relevant nomenclator to identify status, synonyms and references for each name.

3.5 Costs associated with compiling the lists

National checklists can be compiled without great expense. The main costs will be to provide incentive grants to specialist taxonomists, subscriptions to citation and abstract databases, to access publications through journals which might require purchasing PDFs, and, possibly the main expense, salaries for checklist compilers where they are contracted specifically to compile the lists and/or for integrating smaller checklists into larger, more comprehensive ones and for formatting checklists submitted by different contributors. Smith & Figueiredo (2010) calculated that it took one person one year to compile a country plant list that included 11,316 plant names for 7,296 plant taxa, and the *major cost was the salary* for this person.

Many scientific institutions have access to literature and citation/abstract databases, so this may not be an expense, and some resources such as the Biodiversity Heritage Library and components of JSTOR are open access. Wikispecies and Google are also freely accessible and often provide useful leads to relevant information or resources. There may be a cost for software development, but with the number of open source options, and the relative simplicity of the checklist structure, this should not be necessary.

4. Scope of the checklists

Checklists can be simple lists of scientific names organized within a classification system, but for national checklists this format would be of little use. *Increasing the information on each species can add considerably to the value of the list and thus increase its audience and use*. The amount of information provided has to be balanced with the effort and the time required to compile the list. Initiatives such as GBIF and the Encyclopedia of Life are compiling detailed information about species, and checklist compilers and designers should aim to link to these resources, rather than duplicate their efforts.

There are options for the categories of species the checklist covers, and how rigorous the requirements for validation of species' presence in the country will be.

4.1 Policy on which species to include in the checklist

Explicit policy on which species to include in the national checklist needs to be developed, stated with the checklist, and adhered to in a standardized way across taxa. Some of the decisions covered by the policy should include:

Whether only indigenous or also introduced and naturalized or alien invasive species are included;

Whether only those species currently present in the country or those recorded historically, but which are now extinct, are included;

Whether fossils will be included and if so, for which historical or geological period;

Whether species for which there is only weak evidence for occurrence in a country will be included, and what evidence will be required for making a decision about presence in a country. For example, will unverified observational data be accepted, or only species for which vouchered specimens have been verified by a taxonomic expert? And

What to do about species for which there is uncertainty about locality. For example, many old specimens in collections have broad locality data which may or may not allow definite placement of the species in a country, and in other cases locality names and country borders have changed and it is uncertain exactly where the specimen was collected.

4.2 Critical data to include in the list

GBIF (2010a,b) has produced two guides that include information about the type of data needed for checklists. One of these is a reference guide for the 'Global Names

Architecture (GNA) Profile for Darwin Core Archives[39]'. The Darwin Core is a set of standards for data, and is the recommended standard for submitting data to the GBIF network. The GNA is an effort to develop standardized methods, practices and shared resources specifically for the taxonomic component of biodiversity data. The reference guide presents the core terms and descriptions for the information that should be included in a species checklist, as well as some additional fields that could be used in an annotated checklist. The list of data required is long and complex, comprising 41 fields for basic information alone, and it is unlikely that all fields will be included in a national checklist, but it is worth reviewing the GNA structure. The second document (GBIF, 2010a) on best practices for publishing checklists[40] provides useful guidelines for the structure and formatting of the data.

The main data, which should be the first priority for inclusion in a national checklist, are as follows:

Classification levels (taxonomic ranks): There are various ranks which differ slightly between the different Kingdoms, and even between groups within Kingdoms. There are, however, seven classification levels which must be included in a checklist: Kingdom, Phylum or Division, Class, Order, Family, Genus and Species. Subspecies and varieties may also be included. The more taxonomic ranks that are included in the classification (e.g. sub-Kingdoms, Subphylum, Subclass, Infraclass, Cohort, Suborder, Infraorder, Tribe, Subfamily, etc.), the more effort will be required to compile and maintain the list, and there will be more conflicting systems for which decisions need to be made. For national checklists it is therefore advisable to include the major taxonomic ranks or levels, at least as the standard minimum, and only to include additional levels where there is a strong reason.

Scientific species name and author: An 'authority' must accompany a scientific name. The authority is the name of the scientist who first published the name. For example, the scientific name *Apis mellifera* is written as '*Apis mellifera* Linnaeus, 1758' because Linnaeus described and named this species in 1758. The names of authors are frequently abbreviated in botany and mycology and there is a regulated list of standard abbreviations for authors which is accessible through IPNI[41]. Abbreviations are not used for authorities of zoological names. There are other

---

[39]GBIF GNA Profile Reference Guide for Darwin Core Archives, version 1.0, released on 1 April 2011, (contributed by Remsen D.P., Döring, M, Robertson, T.), Copenhagen: Global Biodiversity Information Facility, 28 pp. Accessible at http://links.gbif.org/gbif_gna_profile_reference_guide

[40]Best practices in publishing species checklists. Contributed by Remsen, D., Döring, M., & Robertson, T. Copenhagen: Global Biodiversity Information Facility, 20pp. Accessible at http://links.gbif.org/checklist_best_practices

[41] IPNI author abbreviations: http://www.ipni.org/ipni/authorsearchpage.do

differences in citing authorities for plants and animals: it is standard that if a species' name or placement has changed since the original description, the original authority's name is placed in parentheses. In botany, in cases where a species is no longer placed in the genus in which it was originally described (i.e. a new combination of genus and species has been published), both the author(s) of the original genus placement and those of the new combination are given, with the former in parentheses. In zoology the publication year is given following author name(s), while this is often not done in botany, and the authorship of a new combination is not given for animals (see Welter-Schultes (2012)[42] for details of conventions for citing zoological names).

Synonyms: These are names which have been previously used for a taxon but which are no longer considered to be the accepted or valid name for the species. Incorrect spellings of the name can also be considered to be synonyms. For each species name considered accepted or valid, all synonyms, with the reference in which the synonymy was identified should be included. The type of synonym can also be included (e.g. homotypic synonym, see section 2.1.3 above). Where one name has many associated synonyms this can create structural challenges which will need to be considered, but it is important to include all synonyms so that the fate of any name that has been used can be tracked in the checklist.

Unique identifier for each species: As the electronic management and dissemination of species information has increased, the pressure to provide a unique number or code for all biological names has increased. There are various types of unique identifiers and no single widely-accepted or complete system. Life Science Identifiers (LSID) are used by the large nomenclatural databases, and if a particular species or higher taxon name already exists in the database, then the number should be available, and be included in the checklist. The Index Fungorum[43] and International Plant Names Index[44] provide the LSIDs for many names, and the Catalogue of Life[45] is probably the most useful for accessing the LSIDs for higher animal taxa. There is no single, coordinated source of LSIDs for animal species. ZooBank[46] provides LSIDs for names but currently has limited coverage. Where LSIDs do not yet exist, an effort should be made to

---

[42] Welter-Schultes, F.W. (2012). Guidelines for the capture and management of digital zoological names information, version 1.0, released on June 2012, Copenhagen: Global Biodiversity Information Facility, 126 pp, ISBN: 87-92020-44-5. Accessible at http://www.gbif.org/orc/?doc_id=2784.

[43] Index Fungorum: http://www.indexfungorum.org/names
[44] IPNI: http://www.ipni.org/
[45] Catalogue of Life: http://www.catalogueoflife.org/
[46] ZooBank: http://www.zoobank.org/

register them through nomenclature sites and include them in the checklist. The source of the unique identifier should be stated in the checklist.

Source of the checklist information: For each species on the checklist, a reference should indicate the basis on which it is included. If most of the species belonging to one family are in the list because they were extracted from a world catalogue for the particular family, then that catalogue would be the reference for the source of those species. The source of the species record may be a specimen database or a published journal article, where these are used for identifying species for inclusion in the list. The source reference may be the same as the original description reference used in the compilation of the checklist, but in many cases the source reference will be different from the original description. In some cases the bulk of or even the entire checklist will have been compiled from a single reference (e.g. a published checklist or a monograph), but this may change for individual species as updates are made using new publications. It is therefore inadequate only to include the reference source in the metadata profile for the list. While it may seem enormously time-consuming to include a source reference for each species, this is essential for the rigour and credibility of the list.

Author/compiler/editor: The individual responsible for adding the species or other name or changing status in the checklist should be documented. This information may be included in the metadata document (see below), but it is also important at a finer level such as for each species. Where checklists include a major or broad taxonomic grouping, there are likely to be many authors who have contributed to the list, which may be difficult to detail in the metadata. Over time, numerous contributors may need to be given credit for updates or additions.

Stating the author not only allows credit to be given to those who have contributed to the list, but also provides some indication of the credibility of the entry. Naming the individual responsible for the record also promotes accountability for its accuracy.

Date of capture/amendment: This allows users of the information to determine and track the number of new species added and changes made in a given time period, and to identify easily the set of most recent changes by searching the checklist by date. It also allows machine-mediated synchronization of data held in different systems.

4.3 Additional/optional data for inclusion

The critical or core components of a checklist are the classification and names of organisms, with the formality of the authority for the name, and the source of the species record. There are several other fields or types of information (extensions from

the core) that may not be considered to be critical in a checklist, but which do increase its value if included.

Common or vernacular names: These are useful especially for the non-scientific community and for communicating species information to a wider audience. For some taxa the common name is well-established, but in invertebrates many species lack common names, and for some taxa, including plants, there are many common names for one species, and many species with the same common name. The checklist can serve as a coordinating and standardizing mechanism for common names. When new species are described, formalizing the common name soon after the description of the species and making this available could limit the future proliferation of common names for a single species.For some taxa, especially plants, birds and mammals, species may have well- known names that are not English, and others have common names in a variety of different languages. In South Africa many medicinally important plants that are used for traditional healing have non-English names. The situation in other multi-cultural countries may be similar, and it will be necessary to consider including common names in the main languages of the country. The number of languages to include in a national list needs to be dealt with sensitively and sensibly, with feasibility being a major consideration. Consultation with relevant stakeholders to decide which languages to include and which sources of common names to use may be necessary.

References: Citation of the publication in which the species is described is very useful, and can at a later stage be linked to the actual publication through the Biodiversity Heritage Library or some other literature database. Including all taxonomic references for each species would provide a valuable additional resource, but will increase the time required for compiling and updating the list, and is probably more suited to the species pages in the Encyclopedia of Life than a checklist. The format used for the reference will need to be standardized in the checklist, but the usual authors, date, title, journal, volume and page numbers should at least be included.

Type information: The type locality and repository (museum or herbarium where the type is housed) for the holotype specimen are very useful additional pieces of information to include in a checklist. The catalogue or accession number of the type specimen, if available, is also ideal because it links the name to an actual vouchered specimen on which the name is based. If time is being spent, the references consulted while compiling a checklist often include information on type specimens, and it is therefore more efficient to capture such information at this initial stage.

Where type information is not available, or the type locality is outside the focus country for the checklist, then it may be useful to link the species name in the list to

at least one specimen in a local collection and from a locality within the country, whose identity has been verified by an expert. This may not always be possible, but it does improve the credibility of the list.

Distribution and Endemism: In deciding the scope of a national checklist, if all species present in a country are to be included it will be necessary to identify whether species are indigenous or introduced/exotic. This would be the simplest level of distribution information. Identifying which species are confined to the focus country (country endemics) is also important and can be done by having a field specifically stating endemic/non-endemic, or by including the countries in which a species occurs. For species that occur widely, it may be necessary to use broader categories, such as Africa, Global or Cosmopolitan. An additional level of distribution data could be to include the state(s)/province(s) in which the species has been recorded, so that at least checklists or inventories of species for major areas could be extracted. This is especially important in a country such as South Africa, where conservation is at a provincial level. Including more distribution information than this greatly increases the effort required for compiling the checklist. Ideally, at some stage in the future it will be possible to link species in the list directly to occurrence data in GBIF network.

Threat status: Including threat status as listed in the IUCN Red List of Threatened Species[47] and CITES[48] or according to the country's own threatened species lists is useful in that all information about all species would be coordinated through a single list, rather than having a separate list for threatened species that would need to be maintained in terms of taxonomic changes. Threat status will need to be updated when new assessments are carried out, and when revisions change the status of taxa or names.

Taxonomic note or comment: It is important for checklist compilers and contributors to be able to flag potentially problematic taxa in the list, to be able to highlight where there is debate or dilemma that require further research for resolution, or to make any other comment of relevance to the name and its taxonomy.

4.4. Taxon or country-specific options for inclusion in the checklist

There are bound to be additional data fields identified as being important, because of either country or taxon needs. For example, in plants, life form descriptors (herbaceous, shrub, tree, etc.) may be important, while in animals a broad habitat category (marine, freshwater, terrestrial, or free-living/parasitic) may be important.

---

[47] The IUCN Red List of Threatened Species: http://www.iucnredlist.org/
[48] CITES species index: http://www.cites.org/eng/resources/.../checklist11/CITES_species_index.pdf

Economic importance may be categorized and included (e.g. medicinal use, crop pest, livestock disease vector), but in any such case it is suggested that a standard set of terms be developed and adhered to so that searching for and extracting databases of subsets of the entire checklist (e.g. all crop pests) can be done quickly and effectively.

4.5. Metadata for the checklist

The GBIF (2010a) document suggests that a metadata profile should be developed for each checklist dataset and that this should include the individuals and organizations that may be associated with a dataset, the links to the homepage of the source, whether the checklist is linked to a particular project, and the citation (exactly how the checklist data should be cited when used). The list of references used to compile the checklist can also be included in the metadata document. There can be two levels of metadata profile for large, integrated checklists covering multiple higher taxa — one for the overall checklist and a finer detail for checklists of specific taxa which contribute to the major list.

5. Maintaining the checklist

Checklists are generally dynamic, and in order to ensure that they provide the most up-to-date information, they will need to be regularly maintained. *Documented and nationally- implemented policy and procedures for the maintenance of national checklists are critical* to ensure that all contributors and users understand the rationale for the inclusion of data and for changes and additions made, and also to ensure that there is some degree of standardization across taxa and compilers in terms of approach to updating. The procedures ensure transparency about how changes are made and who is responsible.

5.1 Reasons for changes in the checklist (types of changes)

Checklists will need to be updated in response to various research developments including:

The description of a new species, genus or higher taxon which needs to be added to the checklist. If the checklist includes infraspecific taxa, these may also need to be added;

When the status of a name changes, for example if it is synonymised;

When the spelling of a name is corrected;

When the name is changed because of other nomenclatural requirements;

The genus to which a species was assigned changes (i.e. new combinations);

The higher classification changes;

Additional species are recorded from a country, either through new arrivals or new data revealing the presence of a species not previously recorded, or new data reveal that a species included is not actually in the country; or

Other additional information of relevance to the checklist fields becomes available to fill an existing gap.

5.2 Approaches to updating the checklist and capacity required for different scenarios

Providing a national checklist which reflects the latest understanding of species, names and classifications is best achieved if the checklist(s) are coordinated through one organization or institution which applies a single policy and procedure in the compilation and updating of the checklist(s). If many individuals and institutions are

all responsible for maintaining components (either lower taxonomic levels, or smaller geographic regions) of a national checklist, then it is critical that a common policy and approach are accepted, understood and uniformly applied by all contributors.

A policy for national checklist maintenance is necessary to ensure broad understanding and acceptance of changes made to the checklist, and an understanding of the frequency with which the list is updated.

The <u>policy decisions</u> made by SANBI taxonomists for the South African National Plant Checklist include:

The checklist will be updated on an ongoing basis, with *changes identified in the literature being incorporated into the checklist within three weeks of receipt of the publication*. *Only published changes will be incorporated*, but where a publication leaves unresolved issues this may be stated in a notes column or field. For example, where a study has investigated the DNA of one of nine species in a genus, and the results indicate that this species should be placed in a different genus, only the results for this species can be reflected in the checklist. A note may be added to the other eight species indicating that they may also belong to the other genus, but that this requires investigation. The assumption cannot be made that, because the DNA of one species indicated a new combination, the same will be true for the other species.

All newly published names will be added to the list, whether resident scientists agree with the concepts or not.

All published changes will be incorporated into the checklist, but those published in non-peer-reviewed journals or informal sources will be indicated as such by having a field specifically for this kind of information.

Experts within SANBI or from other institutions can submit an application not to implement a change to the checklist if it will have a major impact on nomenclature and classification and it is likely to be repudiated within a short period of time by other published research. This application has to be submitted to an internal research committee before acceptance, and this type of appeal will likely be an exception rather than a regular occurrence. An example of this would be where the author of a molecular study splits a very large genus (eg. more than 200 species), but the phylogeny is considered by another taxon expert to be weakly supported. In such a case a time frame for not capturing the change will be set (two to three years), and if no other studies are published on the genus to refute the findings then it will be adopted in the species checklist.

The status of names will be indicated based on the Plant List[49] information and not unpublished opinion. Experts will not be consulted on inclusion of the published taxonomy because this may lead to conflict among specialists. In cases where the name or taxonomy needs to be treated with caution this can be indicated in the comments field.

One higher-level classification system will be used in the checklist and adhered to in its entirety. (For vascular plants the Angiosperm Phylogeny Group [APG] II system was selected). The classification system and the policy will be reviewed every three to five years and may be amended then.

It should be noted that this policy may not be ideal for all countries or taxa, and different approaches may be adopted. The South African example helps to illustrate the type of decisions that should be included in an explicit policy document for national checklists, rather than prescribing what the decisions should be.

The SANBI procedures for maintaining the South African National Plant Checklist may also be useful as a guide for other taxa and for other countries, but details may need to be modified.

The decision was made to appoint *a single staff member*, with a sound knowledge and understanding of plant nomenclature and literature, to take full responsibility for maintaining the checklist. This 'checklist co-ordinator' will:

Scan the published literature that SANBI subscribes to for identification of relevant publications;

Establish and communicate regularly with a network of researchers who publish taxonomic papers on South African plants to request copies of newly-published work not in the list of journals to which SANBI subscribes. The editors of some of the more obscure journals will be contacted and a similar request made;

Use an accessible literature database to search at least biannually for new species and other relevant publications;

The global Plant List[50] will be scanned on a six-monthly basis to identify any publications that might have been missed, and to check the status of names; and

---

[49, 33] The Plant List: http://www.theplantlist.org/

The following will be recorded for each change, as appropriate: new species or species whose status has changed; synonyms; reference; date that the change was made; a comment, where required, explaining the change.

Dissemination:

The updated electronic checklist will be disseminated through SANBI's website, with a function that allows a search by date so that the latest changes can be identified by users.

A feedback function that allows input on errors is also required on the checklist. The checklist coordinator will be responsible for responding to comments.

The system formerly in place for maintaining the checklist involved different taxonomists being allocated responsibility for several plant families for which they had to identify publications that included changes. They then assessed the paper, and made a decision on whether to accept the taxonomic change of new species/genus/classification. If they accepted the change, it was then recommended to a staff member who captured the change in the checklist, printed out the change and referred it back to the expert for checking. Once approved by the expert, the change was finalized in the database. This type of system creates four problems:

Published names or taxa are not captured in the checklist if they are not agreed upon, and this may cause confusion for users of the checklist attempting to check the status of the name or taxon;

It allows subjective decisions to be made about changes, without any standardization. Personal preferences often cloud judgements and loss of credibility of the checklist results;

Different families are updated at different frequencies and at different times because some staff may be away or consider this responsibility a low priority; and

This is not a cost effective approach to maintaining the checklist since the time, and thus salaries (estimated 10%) of all 20 staff members, most of who are senior scientists have to be considered, as opposed to a system where the salary (probably 60%) of a single, mid-level staff member is required to implement an explicit procedure and policy.

In the case of animals, the maintenance of a national checklist covering all taxa may be considered as being too broad a task for a single staff member, and in this case individual experts could take responsibility for identifying changes. The experts can either update the list online, or send changes through to a checklist coordinator. The latter is probably preferable because of different levels of database and web expertise

amongst participants, different interpretations of the fields/information required, and varying frequencies at which the changes are made. If individual experts maintain the checklist, it is advisable to hold a workshop to ensure that all contributors agree on the policy and procedures, and are competent in the technologies that are used for submitting the changes. It is also advisable to have a checklist coordinator, even if on a part-time basis to ensure that the format is standardized, to send regular reminders to contributors, to deal with technological problems and to maintain those components of the checklist for which there are no experts. Without a coordinator, the checklist maintenance process is likely to collapse or lose credibility because it is inconsistently maintained.

6. Format Options

There are two aspects relating to the format of the checklist, i.e. the format in which the information is captured and managed and the format in which it is made accessible and disseminated. The advantages of an online electronic national checklist over a hard copy publication have been discussed already, and only the electronic format is considered here. The GBIF (2010a) document provides details for the format in which data should be captured, and also provides some guidelines for how a checklist should be structured. The recommendations given here relate more to some key principles for the format of the checklist, rather than detailed instructions.

6.1 Options for capturing the checklist and considerations in selecting format

There is no standard database design available for checklists. Several collection/specimen databases do have a taxonomy component. The Botanical Research and Herbarium Management System (BRAHMS) and the SPECIFY databases structures have taxonomic sections which allow the compilation and extraction of checklists, but BRAHMS is not open source and both softwares require relatively intense training to become proficient in their use.

The Integrated Taxonomic Information Systems website [51] (ITIS) does provide a downloadable workbench which was created using Visual Basic and Microsoft Access. Information is quite easily entered, copied, and manipulated using standard Windows commands and menus.

The Biodiversity Information Standards (TDWG) Subgroup on Biological Collection Data site lists collection management database software[52] that includes taxonomic or nomenclatural components. Several of the databases are open source, and are potentially useful for compiling and managing checklists.

Databases can be designed specifically for compiling and managing taxonomic information relevant to checklists. These types of structures and formats are acceptable if the checklists are compiled and managed by one person who is familiar with the operations of the database, and who has the software on a desktop, but if the expectation is that taxonomists scattered at various institutions will manage the checklists, then this is a risky approach since it is unlikely that all taxonomists will be

---

[51] Integrated Taxonomic Information Systems: http://www.itis.gov/twb.html
[52] Biodiversity Information Standards (TDWG) Subgroup on Biological Collection Data: http://www.bgbm.org/tdwg/acc/Software.htm

familiar with specialized software, and they may not have it set up on a desktop. Problems may also occur if the developer is not available to assist when there are problems with the database, or to deal with requests for additional functions.

There is a wide range of software that can be used to produce a simple checklist database or spreadsheet and individuals are likely to have personal preferences. Microsoft Access and Excel can both be used for checklists with the advantage that many people have access to them and are able to use them to capture and manipulate taxonomic information. However, there are costs associated with the initial purchase and upgrading of Microsoft, and for checklist compilers wanting open source software, SQLite databases can be created and managed with a Web browser add-on. *The key considerations are that the data must be readily transferable to other software structures and platforms and whatever is used must be simple to design and manipulate even by those users with limited IT knowledge or experience*.

6.2 Options for providing online access to the checklist and considerations in selecting format

There are many different database structures and web interfaces that are suitable for disseminating checklists. Some key requirements from a user perspective are discussed below:

Searching by name alone limits the use of the checklist. It is important that *the list can be searched by any taxonomic level, and that a search is able to draw up a list of species* and not only information on a single name. For example a search by family should provide the option for listing all species within a family, rather than only providing information about the family. The *checklist should also be searchable by type repository, distribution (country and state/province), and importantly, by date of last update/change*. Having a number of different search criteria increases the usefulness of the checklist.

Checklist users will often want to download portions or even the entire checklist and allowing this option also increases the value of the list. A simple comma-separated values (*CSV*)/spreadsheet is usually acceptable to most users because this format allows manipulation and searching of the data. The GBIF (2010a) document provides an internationally-accepted standard for sharing checklists in a CSV format. A PDF format on the other hand does not offer these options, but it does have the advantage of not enabling changes to the list which might be preferable if standardization of the national list is an objective.

7. Copyright and data ownership issues for national checklists

National species checklists should be considered as a national and even as a global resource. *Ideally users should be able to download the checklist or a component of it and use the data freely but acknowledge the source of the data*, and where relevant the author or compiler of the particular component of the checklist that is used. Obviously extraction and publication in a different format and forum by someone other than the compiler should not be permitted without the permission of the compiler. *The checklist website should explicitly state any limitations for use and the requirements for acknowledgement*.

Published and web-based resources will be used in the compilation of a checklist and the source should be referenced in the checklist. Some resources may require that permission to use data be obtained or have other copyright requirements to be complied with.

8. References

Arnold, T.H. & De Wet, C. (eds). 1993. *Plants of southern Africa: names and distribution [Issue 62 of Memoirs of the Botanical Survey of South Africa](#)*. National Botanical Institute, Pretoria, 825pp.

Crosskey, 1980. *Catalogue of Afrotropical Diptera*. British Museum (Natural History), London. 1235pp.

Figueiredo, E. & Smith, G.F. 2008. *Plants of Angola / Plantas de Angola*. Strelitzia 22: 1–282. South African National Biodiversity Institute, Pretoria.

GBIF (2010a). Best practices in publishing species checklists. Contributed by Remsen, D., Döring, M., & Robertson, T. Copenhagen: Global Biodiversity Information Facility, 20pp. Accessible online at http://links.gbif.org/checklist_best_practices.

GBIF (2010b). GBIF GNA Profile Reference Guide for Darwin Core Archives, version 1.0, released on 1 April 2011, (contributed by Remsen D.P., Döring, M, Robertson, T.), Copenhagen: Global Biodiversity Information Facility, 28 pp,. Accessible at http://links.gbif.org/gbif_gna_profile_reference_guide.

Germishuisen, G. & Meyer, N.L. (eds). 2003. *Plants of southern Africa: an annotated checklist*. Strelitzia 14.

Germishuizen, G., Meyer, N.L., Steenkamp, Y. and Keith, M. (eds). 2006. *A checklist of South African plants*. Southern African Botanical Diversity Network Report 41 SABONET, Pretoria.

Harvey, W.H. 1838. *The Genera of South African Plants*. Cape Town, 429 pp.

Phillips, E.P. 1926. *The Genera of South African Flowering Plants*. Cape Times Ltd, Cape Town. 702pp.

Remsen, D., Knapp, S., Georgiev, T., Stoev, P. & Penev, L. 2012. From text to structured data: Converting a word-processed floristic checklist into Darwin Core Archive format. *PhytoKeys* 9: 1–13, doi: 10.3897/phytokeys.9.2770.

Smith, G.F. & Figueiredo, E. 2010. E-taxonomy: an affordable tool to fill the biodiversity knowledge gap. *Biodiversity and Conservation* 19: 829–836.

Welter-Schultes, F.W. (2012). Guidelines for the capture and management of digital zoological names information, version 1.0, released on June 2012, Copenhagen: Global Biodiversity Information Facility, 126 pp, ISBN: 87-92020-44-5, accessible online at http://www.gbif.org/orc/?doc_id=2784.