The price of collecting life

Overcoming the challenges involved in computerizing herbarium specimens.

Gideon F. Smith, Yolande Steenkamp, Ronell R. Klopper, Stefan J. Siebert & Trevor H. Arnold

The digitization of natural history collections in museums — the capture of data from specimen labels — is an essential first step in transforming these vast amounts of data into accessible, usable and useful information products. This process has attracted a lot of attention lately[1], mainly because it is vital if these institutions are to remain relevant to their customers and stakeholders. However, very little has been done to accurately calculate the costs involved in such an exercise. What follows is a summary of our experience in southern Africa of computerizing herbarium (plant) specimens in the Southern African Botanical Diversity Network (SABONET) project. Our findings provide a valuable and accurate estimate of the costs involved for other, similar enterprises.

Computerization of natural history collections is an objective of the Global Biodiversity Information Facility (GBIF) (www.gbif.org), the All Species Project (www.allspecies.org), and a goal of numerous major natural history collections and taxonomic institutions for more than three decades. The Botany Department of the Natural History Museum in London, for example, has designed and is currently testing an experimental database to quantify the time, effort and money spent on indexing natural history specimens in this way (www.nhm.ac.uk/botany/cuttings/issue9/newsandviews/index.html).

With access to comprehensive, interoperable databases, scientists and other users of taxonomic information will be able to rapidly compile biodiversity catalogues, study biogeographical patterns, and assist in both long-term conservation planning and crisis conservation management. In addition, we taxonomists should bear in mind that our salaries and project running costs are often paid from public funds. We should therefore be prepared to make our data accessible to interested and affected parties.

Plants lead the way

Extensive work has already been done on animal collections, but the computerization of herbarium collections has mostly lagged behind. However, in southern Africa, the Pretoria National Herbarium Computerised Information System (PRECIS) database — one of the first of its kind and size in the world — has been under development since the early 1970s, providing the single largest digital repository of botanical information in Africa. Since its small beginning, PRECIS has extended to ten southern African countries, largely under the auspices of SABONET (Box 1).

Although PRECIS can house a myriad of information (Box 2), the capture of information into selected fields has been emphasised for the past few years by SABONET. These selected fields include accepted name, synonyms and importantly, geo-referenced distribution data (GPS or grid location data). This policy is in line with other major projects, including GBIF. GBIF focuses on species and specimen level data, and its DIGIT programme will concentrate on capturing and geo-referencing basic specimen label data that can later be extended to include other information.

It is easy to estimate the number of specimens housed in southern African herbaria. Following the publication of the comprehensive southern African Index herbarium and a few updates (ref. 2 and references therein), we know that 4,223,697 specimens are kept in the herbaria of the subcontinent (Fig. 1, overleaf). Since the early 1970s, 760,000 specimens kept in South Africa's national herbarium have been computerized — this figure being given a boost in the mid-1990s when SABONET extended the programme to nine other southern African countries. A further 374,532 specimens have been added to PRECIS since 1998, through the work of dedicated and

Several valuable lessons learnt by SABONET would be important for similar future projects.

Box 1 SABONET

SABONET (www.sabonet.org) is a successful donor-funded, capacity-building project operating in Angola, Botswana, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia and Zimbabwe. Botanists from these countries are involved in compiling specimen inventories, Red Data Lists and national plant checklists under the auspices of SABONET. Plant collecting expeditions target under-collected areas and under-collected taxa. The training of postgraduate systematics and biodiversity science students and herbarium personnel is also supported by SABONET. One of the project's principal objectives is the "computerization of plant specimens stored in national and regional herbaria and botanical gardens".

Projected outputs include:

- A checklist of the southern African flora, complementing ref. 3.
- A checklist of the endemic plants of southern Africa.
- A collated distribution map of the southern African flora indicating areas of high diversity and endemism.
- A comprehensive list of the grass taxa for all participating countries.

Box 2 PRECIS

The PRECIS database consists of a number of components, two of which house the specimen records and the taxon names, respectively.

In the specimen sub-database, information such as family name, genus name, specific epithet, type status, identification level (degree of certainty of identification), temporary name, name of the person who identified the specimen, name of the collector(s) with collection number and the date of collection can be recorded. In addition, the region and country where collected, longitude and latitude, and a 1/4-degree grid reference[4] or exact location (GPS-reading) are also recorded. Habitat information (such as soil type, substrate, moisture regime, vegetation and biotic effects), and phenological information (such as presence of flowers, fruit and seed, flower colour, growth form and height of plant) may also be recorded. There is also a 'notes' field to record any additional information on the plant's morphology, ecology, occurrence, common name, usefulness, toxicity, reproductive biology, horticulture, aroma and whether or not it is desirable (weed, invader, and so on). Finally, it can be noted if the specimen serves as a voucher for any particular study.

From the 'plant name' sub-database (taxon sub-database) one can obtain information on specific taxa, including synonyms, distribution within southern Africa and uses (for example, medicinal, food). Currently, only approved users have access to PRECIS, but the intention is for it to be available for anyone to use freely.

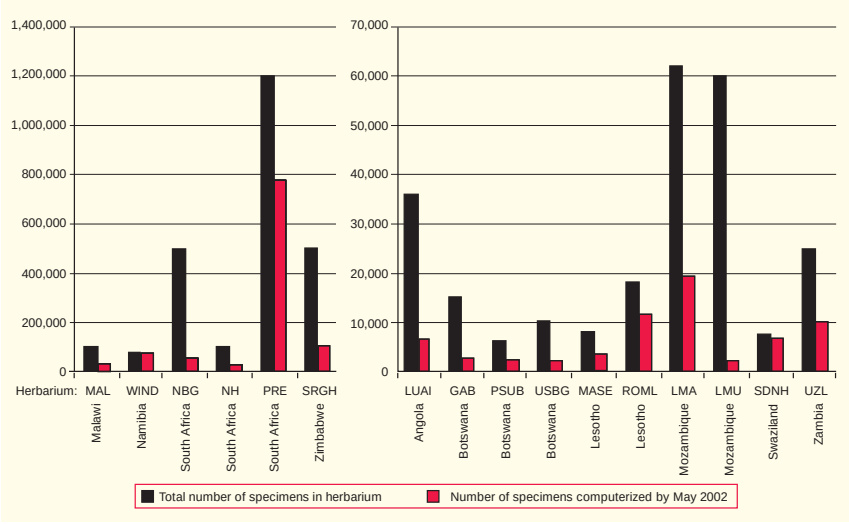

Figure 1 **Total number of specimens housed in participating herbaria, as well as the number of specimens computerized by May 2002. All herbarium acronyms follow ref. 5.**

part-time data capturers. A total of 1,134,532 of the specimens contained in southern African herbaria have therefore now been computerized (Fig. 1), and by extrapolation, between 481,250 and 604,681 specimens would ultimately be captured under SABONET.

From January 1998 to May 2002, it has cost an equivalent of US$1,376,223 to computerize these herbarium collections in southern Africa (see Table 1 for a breakdown of the expenses). This figure includes aspects such as staff remuneration, consultants' fees, acquisition of computer hardware and software, field/collection trips, acquisition of equipment such as microscopes, herbarium cabinets, vehicles for fieldwork, workshops and training courses. During this period, 374,532 specimens were digitized — hence it costs about US$3.67 per herbarium specimen in southern Africa. This figure may not exactly match costs elsewhere in the world — staff remuneration would be much higher in the developed world, but other expenses could be less, for example acquisition of microscopes, other laboratory equipment and computers, and the need for vehicles for fieldwork would not be as necessary or as extensive as in developing countries.

Very little comparative data are available from other institutions or similar projects. The Smithsonian Institution estimated the costs involved in digitizing its collections, including staff remuneration and the upgrading of computer hardware and software, but not including training, collection and so on. This fell between US$1.63 (cataloguing standards) and US$12.50 (enhanced standards) per specimen, depending on the amount of data captured. Computerization under SABONET requires the capture of information (for example geo-referenced distribution data) beyond mere cataloguing standards, but not necessarily to the level of fully enhanced standards, thus making it difficult to compare these figures accurately.

GALLO IMAGES/CORBIS

King Protea: the national flower of South Africa.

Lessons learnt

Several valuable lessons learnt by SABONET would be important in the consideration of similar future or ongoing projects.

The importance of capturing specimen data — for both the institution and the region — must be emphasised by the project management. When the participating institutions are not convinced of this importance, they do not see this activity as a primary objective, and concentrate resources elsewhere.

Table 1 **Breakdown of the cost of computerizing herbarium specimens in southern Africa, in US dollars**

Description	*1998 to May 2002*
Salaries	532,936
Evaluation	6,314
Meetings/workshops	70,020
Travel expenses	50,527
Consultation fees	41,233
Training courses	84,552
Field trips	35,792
Running expenses	189,002
Equipment	293, 401
Maintenance	7, 701
Publications	64 ,740
Total	1, 376, 223

Figures are the portion of SABONET expenditure on computerization. Total SABONET expenditure in this period was US$2,610,367. Specimen numbers from ref. 6.

Building capacity takes time. Newly appointed data capturers, electronic networks and system managers do not become confident overnight, and follow-up training courses will invariably be required (and should be budgeted for).

Participating centres should be encouraged to make use of local computer service providers. Institutions have on occasion waited weeks for SABONET to fix their 'faulty' database, when in fact the problem was a technical one that could have been fixed within a day by a local service provider.

A quick-response call centre is a necessity. When a problem with hardware, software, database construction or data gathering is encountered in a relatively isolated participating centre, a mechanism must exist to address and rectify it rapidly — for example, through a call to a central help desk, or courier service to send equipment, such as hard drives, to centrally based technical staff.

Data capturers should be encouraged to remain in their positions for at least a year. Data capturer positions have a high turnover, and new clerks are often appointed long before the next database training course is scheduled. As knowledge transfer is often inadequate, clerks remain unproductive until they have attended a course. New, untrained data capturers also do not usually understand the terminology in the database, and subsequently waste time contemplating and adding new variables.

Finally, a quality-control mechanism must be in place. Untrained data capturers are often responsible for many wrongly encoded entries and misspelt information, which must be corrected later, wasting resources and time. ■

Gideon F. Smith, Yolande Steenkamp, Ronell R. Klopper and Trevor H. Arnold are at the National Botanical Institute of South Africa, Private Bag X101, Pretoria 0002, South Africa. Stefan J. Siebert is at the SABONET Regional Office, Private Bag X101, Pretoria 0002, South Africa. Author for correspondence: G.F.S. at gfs@nibpre.nbi.ac.za.

1. Godfray, H. C. J. *Nature* **417,** 17–19 (2002).
2. Smith, G. F., Willis, C. K. & Behr, K. *SABONET News* **6,** 214–215 (2001).
3. Arnold, T. H. & De Wet, B. C. (eds). *Mem. Bot. Surv. S. Afr.* **62,** 1–825 (1993).
4. Edwards, D. & Leistner, O. *Mitt. Bot. Staatssamml. München* **10,** 501–509 (1971).
5. Holmgren, P. K., Holmgren, N. H. & Barnett, L.C. *Index Herbariorum, Part 1: The Herbaria of the World*. 8th edn (New York Botanical Gardens, New York, 1990).
6. Arnold, T. & Siebert, S. *SABONET News* **7,** 92–96 (2002).

Acknowledgements.
SABONET thanks GEF-UNDP for funding the computerization process of the participating southern African herbaria. The authors also thank H. Snyman (National Botanical Institute) and A. Weitzman (Smithsonian Institution) for their help.