Bothalia 13,1 & 2: 149–160 (1980)

Encoding the National Herbarium (PRE) for computerised information retrieval

J. W. MORRIS*

ABSTRACT

This paper describes the operation of encoding information from 470 000 extant herbarium labels for accession to a computer data bank with a view to aiding curators and systems analysts who may wish to apply similar procedures in their herbaria. The preparation of the herbarium, encoding procedures used, verification of encoded information, costs and remuneration of encoders are described. A complete list of items encoded and their code equivalents is given. Some of the difficulties faced during the operation are discussed and the success of the project to date is evaluated briefly. Its success is attributed to a bonus incentive scheme and availability of a comprehensive instruction manual.

RÉSUMÉ

ENCODAGE DE L'HERBIER NATIONAL (PRE) POUR EXTRACTION DE L'INFORMATION PAR ORDINATEUR

On décrit ici l'encodage de l'information contenue dans 470 000 *étiquettes d'herbier pour leur incorporation à une banque de données d'ordinateur; cette publication vise à aider les conservateurs et analystes qui souhaiteraient appliquer des procédures similaires à leurs herbiers. La description comprend: la préparation de l'herbier, les procédés d'encodage utilisés, la vérification de l'information codée, le coût de l'opération et la rémunération des encodeurs. On donne une liste complète des éléments encodés et de leurs équivalents dans le code. Certaines des difficultés rencontrées en cours d'exécution sont discutées et le succès du projet a ce jour est évalué brièvement. Ce succès est attribué à un plan d'encouragement financier et à la disponibilité d'un manuel d'instructions complet.*

INTRODUCTION

One of the most expensive and logistically-complicated parts of the operation to set up a computerised information storage and retrieval system for the National Herbarium was the encoding of the extant 470 000 specimen sheets. While it was acknowledged that without inclusion of these specimens, known below as the backlog, a data bank of new accessions would have limited value, the complete encoding of a herbarium of this size had not previously been attempted (Morris, 1974). In this paper, the procedures developed for the encoding of backlog material are described in detail, as it is considered that these will be of use to all curators of collections and systems analysts contemplating such an operation in future. The benefits being accrued from computerisation are such that it is hoped that other herbaria will follow this lead (see Brenan *et al.*, 1975). Preparation of the herbarium, training of staff for encoding and the development of a remuneration system for the part-time bonus incentive encoding scheme were some of the essential aspects which required attention and are reported here. Another reason for recording these procedures, and the explanations accompanying items and their codes in particular, is that users of the data bank will then know the conventions adopted during coding. Without this knowledge, they may have difficulty formulating their queries for optimum results.

Backlog encoding was one aspect of the development of an integral computer-based system for the retrieval of information from specimen labels, taxonomic data and literature references. The whole system, as it exists at present, is described by Morris & Glen (1978). The part described here deals only with that information recorded by collectors on their labels together with the state of the specimen and its scientific name, including only specimens collected prior to the completion of the project reported here.

* Botanical Research Institute, Department of Agricultural Technical Services, Private Bag X101, Pretoria, 0001.

THE HERBARIUM AND ITS PREPARATION FOR ENCODING

The National Herbarium is located in four wings of the building housing the headquarters of the Botanical Research Institute in Pretoria, South Africa. Specimens are housed in some 500 cabinets, each with 24 to 36 shelves, and are arranged in the taxonomic sequence of Dyer (1975, 1976). Specimens are filed in blue, heavy paper species covers which are themselves contained in light card, brown genus covers. Type specimens are generally located at the start of a species and are filed in easily-identifiable special folders with red edges. Within a species, specimens are arranged by province and country with the oldest specimen from a province or country at the top and the newest accession at the bottom. Only one province or country is contained in a blue cover and the province or country is written on the outside of the cover. It was decided that all material of indigenous taxa from the Flora of Southern Africa Area (Ross *et al.*, 1977) as well as all African type material and photographs of African type material, when the actual specimen was housed elsewhere, would be encoded. Specimens having labels with fewer than two items of information were not encoded.

A twelve-digit code number, consisting of seven digits for the genus and five digits for the species was developed (Morris & Leistner, 1975; Morris & Glen, 1978). A computer card was punched for each valid combination and four lists of codes, one for each wing of the herbarium, were printed by computer. It was found that there were more than 17 000 Southern African species names included in these master lists. Once these lists had been checked and corrected they were used to number species and genera folders. All name changes accepted by the herbarium from that time were then entered by hand on these lists.

Before encoding commenced in a wing, curatorial assistants paged through the entire collection, checking that all the herbarium sheets within a species cover

were of the same taxon and that the specimens were correctly identified to genus, species and sub-specific epithet (where appropriate). Where a name could not be assigned, assistants ensured that the unidentified material within a folder was all of the same taxon. The species code number was written on the outside of the species covers and the genus code number was written on the outside of the genus covers.

All herbarium cabinets were numbered by wing and cabinet to facilitate replacement of specimens after encoding.

ENCODING PROCEDURES, VERIFICATION AND REMUNERATION OF ENCODERS

Introduction

Once it had been decided that the backlog housed at PRE would be encoded, the Workstudy Section of the Department of Agricultural Technical Services together with officers from the Botanical Research Institute, investigated ways in which the task could be efficiently, rapidly and cheaply completed (Pieters, 1974). Requirements of the operation were that it was to be carried out within a relatively short period, say two to three years, and that minimal disturbance be caused to routine herbarium curatorial activities. The procedure outlined below was considered superior to the alternative of appointing a smaller number of encoders to work during office hours without a bonus incentive. Direct encoding from specimens onto computer punch cards by means of IBM 029 punches and onto magnetic tape by means of Olivetti key-punch stations were also tried but found to be inefficient, relatively expensive and encoded information was difficult to verify in comparison with encoding onto data sheets and submitting batches of data sheets for punching as a separate task.

Procedures, verification and remuneration

Twenty-eight people were employed to carry out the encoding operation for up to four hours each evening, Mondays to Thursdays. Of these, five teams, each consisting of four encoders and a team checker, completed and checked data sheets, respectively. Two bonus controllers were responsible for vetting a 10 per cent sample of the completed data sheets and a supervisor was responsible for training new encoders, interpreting the encoding instructions written by Morris & Du Toit (1976), when necessary, deciphering unreadable handwriting on labels, translating foreign-language labels and resolving differences of opinion between encoding teams and bonus controllers. Staff consisted of volunteers drawn from the Public Service and no botanical background was expected, except in the case of the supervisor who was expected to have had experience with the project and to have taxonomic knowledge.

Although most of the encoders had no prior botanical background, it was found that, in general, they could be trained within 12 hours (three evenings). Clear, written instructions were a prerequisite for both training and the bonus incentive and penalty scheme and as a result a handbook, which went through a number of editions (Morris & Du Toit, 1976), was compiled. Such a manual also ensured that encoding standards and conventions were maintained throughout the period. Labels in Latin and German were particular problems as they were next most common after English and Afrikaans and most encoders were not familiar with these languages. Lists of words commonly occurring on labels with their meanings and code equivalents (see below) were drawn up for use by encoders (Tables 1 & 2). These *ad hoc* lists should not be considered as definitive translations (the ablative case of nouns, for example, is often given as the most common case occurring on labels was usually listed) but indicate the use made of such aids by encoders.

Another problem faced by encoders was that of illegible handwriting and obvious copy typing errors. Where encoders could not decipher writing, the supervisor was asked to assist. In some cases, words could not be deciphered and were not encoded.

TABLE 1.—Alphabetical list of Latin terms and their English equivalents and numeric codes. Nouns are given in the case (often ablative) in which they are most often encountered on labels

ad ripam=on river bank (moisture regime 07)
ad rivulum=on stream bank (moisture regime 07)
ad viam=at road (biotic effect 06)
albi/albus=white (colour)
anno=year (date of collection)
apertis=open (not coded except *apertis silvae*=vegetation 5)
arenosus=sandy (soil type 02, substrate 1)
argillaris=white clay (soil type 04, soil colour A 1)
aridulus=slightly dry soil (substrate 1)
aridus=dry soil (substrate 1)
campester=flat area (marsh) (moisture regime 05)
capensis planitie=Cape Flats (locality)
CBS=*C.B.Esp.* (=*erance*)=Cape of Good Hope (locality)
circa=about (not coded)
clivus=slope (not coded)
clivus supra=upper slopes (not coded)
collibus/collinus=hill (not coded)
convalle=river valley (not coded)
corolla=part of flower (flower colour)
corona=part of a flower (flower colour)
declivis=downward slope (not coded)
ericifruteceta=fynbos (vegetation 8, Veld type 69)
ericetis=fynbos (vegetation 8, Veld type 69)
flavi=yellow (colour)
flore=flower (possibly flower colour)
flumen=river (moisture regime 08)
fructus/fructificatio=fruit (possibly fruit colour)
fruticis=scrub (vegetation 6)
frutex=shrub (life form 02)
fusci rubi=dark red (colour)
graminosis=grass (vegetation 3)
humidus=damp ground (marsh) (not coded)
lactei=white (colour)
lapidos(is)=rocky (substrate 2)
leg.=collector (collector)
litor(al)ibus=sea shore (moisture regime 12)
lutei=yellow (colour)
luteus=clay (soil type 04, substrate 1)
maris=sea (moisture regime 12)
maritimus=sea shore (moisture regime 12)
montanus=mountain (not coded)
montibus=mountain (not coded unless part of locality)
montis=mountain (not coded)
occidentalis=western (possibly locality)
orientalis=eastern (possibly locality)
paludibus=(in) marsh (moisture regime 05)
paludosus=marshy (moisture regime 05)
Planitie=flats (possibly locality)
planitie=flat (slope 1)
pratis=meadow (vegetation 3)
purpurei=purple (red) (colour)
ripis=river bank (moisture regime 07)
rivulus=stream (moisture regime 08)
rubi=red (colour)
sabulosis=marsh (moisture regime 05)
saltosis=bushy rocky mountain (vegetation 4, substrate 2)
saxosis=rocky (substrate 2)
silvae=forest (vegetation 7)
silvis=forest (vegetation 7)
s.n.=no collector's number—assign a PRE number
suffrutex=shrub (life form 02)
Tabularis=Table (Mountain) (locality)
umbrosis=shady area (not coded)
virides=green (colour)

TABLE 2.—Alphabetical list of German terms and their English equivalents and numeric codes

ausdauernd=perennial (annual etc. 4)
Felsen=rocks (substrate 3)
Geröll=gravel (substrate 1, soil type 1)
Glimmerschiefer=mica slate (rocky) (substrate 2)
haeufig=abundant (abundance 7)
Halbstrauch=halfshrub (life form 03)
Kies=gravel (substrate 1, soil type 1)
lachs=salmon pink (colour)
Quelle=fountain (moisture regime 07)
Rinnsal=watercourse (moisture regime 10)
rosa=pink (colour)
Senke=depression (moisture regime 03)
Staude=shrub (life form 02)
Sumpf=marsh (moisture regime 05)
Ton=clay (substrate 1, soil type 04)
Tor=gate/*poort* (not coded unless part of locality)
Wiese=grassland (vegetation 3)
zart=soft (not coded)
zwischen=between (not coded unless part of locality)

Encoders were further instructed to correct obvious typographical errors.

Encoders collected specimens from the herbarium cabinets, a shelf of specimens at a time. For each shelf a Daily Production Sheet (Fig. 1) was completed which accompanied the specimens, held in a parcel by means of a rubber band, through the process until the specimens were returned to the cabinet after encoding and checking. As some shelves required more than one evening to encode, provision was made on the Sheet for encoding to cover as many as four evenings. Most items required to be entered on the Sheet do not require explanation. New encoders were allowed 12 hours in which to work at their own speed without being financially penalised for errors and speed. During this period they were considered as 'learners' (see Fig. 1). Where more than one specimen was mounted on a sheet, a coding form for each specimen was completed. Information from all the labels attached to a specimen were used and in cases of contradiction, information from the apparently-oldest label was favoured.

After being encoded, the specimens, together with their encoding forms, were passed to the team checker. The checker was responsible for checking all the encoded information and correcting any discovered errors. No penalty was incurred for errors discovered and corrected by the team checker. The checker also rubber-stamped each specimen with a small 'ENCODED' stamp.

From the team checker, specimens were submitted to the two bonus controllers. They took a sample selected at random of 10 per cent of the encoding forms and made a careful comparison of the specimens with the completed data sheets. Errors discovered were of two degrees of severity: full errors (being such things as date of collection omitted or province being incorrectly encoded) and minor errors, counting as one fifth of a full error. A list of full errors was drawn up and it was understood that all other errors counted for one fifth of an error. The number of specimens checked (10 and 3 in Fig. 1) as well as the number of errors found ($\frac{1}{5}$ and $\frac{2}{5}$) were recorded on the Daily Production Sheets. These errors were corrected by the bonus controllers. Specimens were returned to the cabinets by the bonus controllers. Daily Production Sheets were collected until the end of the encoding period, which was usually three or four, four-evening weeks.

At the end of each encoding period the production of each encoder and of each team was calculated and a Monthly Team Summary was drawn up (Fig. 2) for each team. For each encoder, the number of specimens encoded and the number of errors found by the bonus controllers was recorded. The norm was the number of specimens expected from the encoder, based on the number of hours worked and the hourly norm. At the termination of encoding the norm was 16 specimens per hour. Initially it was 10 and it was raised gradually by the supervisor in consultation with the bonus controllers so that the average production per cent of all teams remained in the range from 120 to 140. Production and error per cents were calculated according to the formulae given in Fig. 2. Gross bonus per cent was read from a degressive bonus graph (Fig. 3), using the team's total production per cent. The penalty per cent (using Table 3) was subtracted from the gross bonus per cent to give the net bonus per cent for the team. Bonus controllers made use of a table based on the degressive bonus graph instead of the graph itself for greater speed and accuracy. Although production and error per cents were calculated individually, the team average net bonus per cent was used to calculate payment

DAILY PRODUCTION SHEET

NAME E KING TEAM 3 NO. 027

DATE	HOURS WORKED	TOTAL CODED	LEARNER ×	CABINET NUMBER	SHELF NUMBER
1976.01.19	4	104		30026	4
1976.01.20	1	26			
	5	130			

10	1/5
3	2/5

0.6

13

× If Learner, mark with 'x'.

Initials of Coder E.K. Initials of Checker J.W.M.

FIG. 1.—Example (slightly reduced) of completed Daily Production Sheet.

MONTHLY TEAM SUMMARY

TEAM NUMBER 3 PERIOD 1976.01.19 - 1976.02.13

ENCODER'S NAME	SPECIMENS ENCODED	NORM	PRODUCTION %	NUMBER INSPECTED	ERRORS	ERROR %	GROSS BONUS %	PENALTY %	NET BONUS
B. JONES	443	272	163	44	1,4	3,2			
C. DICK (MRS)	740	528	140	76	3,4	4,5			
D. JAMES	248	144	172	35	4,4	12,6			
E. KING	1438	768	187	154	6,2	4,0			
TOTALS	2869	1712	168	309	15,4	5,0	48,6	7,0	41,6

PRODUCTION % = (SPECIMENS ENCODED/NORM) x 100
ERROR % = (ERRORS/NUMBER INSPECTED) x 100

FIG. 2.—Example of completed Monthly Team Summary.

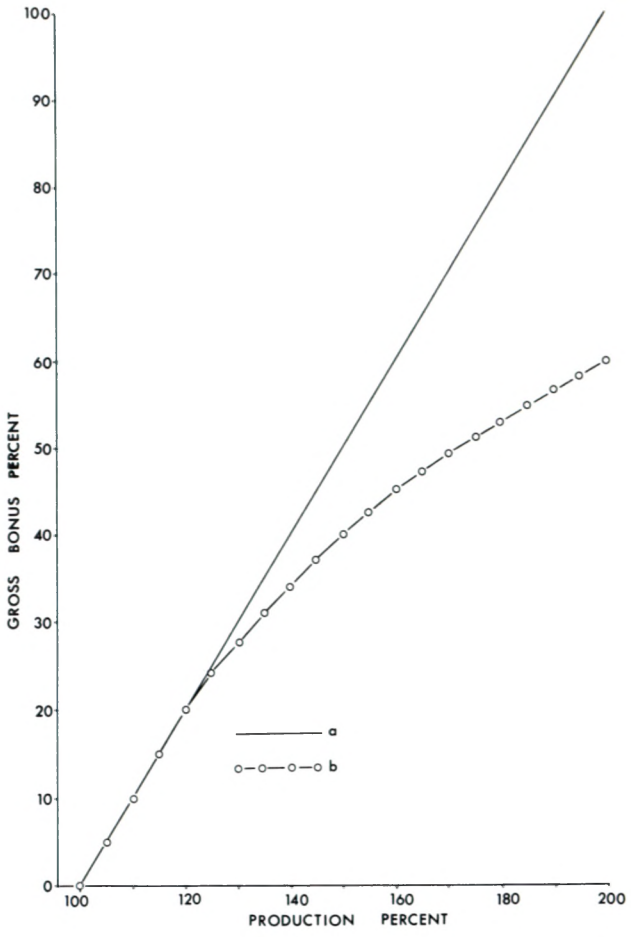

FIG. 3.—Degressive bonus graph for estimation of gross bonus per cent from production per cent: (a) Linear relationship, (b) degressive relationship used in this project.

for all members of a team. Thus a fast, accurate encoder raised the bonus per cent for the whole team while a slow, inaccurate encoder had the reverse effect. Individual statistics were calculated and tabulated (Fig. 2) so that the supervisor could investigate slow or inaccurate encoding at monthly intervals. The degressive bonus graph (Fig. 3) was proposed by the Workstudy Section as a necessary part of such a bonus incentive scheme (Pieters, 1974). Beyond a production of 120 per cent, the team benefits the employer to an ever increasing extent with this relationship. The area between lines a and b (Fig. 3)

TABLE 3.—Penalty corresponding to error per cent. See text for explanation

Error (%)	Penalty (%)
1	—
2	—
3	3
4	5
5	7
6	9
7	11
8	14
9	17
10	20
>10	no bonus

represent the profit of the employer at the expense of the employee. Such a graph is used principally to protect the health of the encoders (Pieters, pers comm. 1978) and to maintain the gross bonus at 20–30 per cent. Bonus calculation per team and not individual is another prerequisite of such a scheme. The penalties corresponding to error per cent (Table 3) were devised so that a small error had at most a very limited effect on the net bonus while at higher error per cent levels the net bonus was drastically reduced and finally no bonus was paid for an error per cent greater than 10. The scale was intended to encourage accurate encoding above mere speed.

The number of hours worked by each encoder and each team checker was entered daily on an Attendance Register and Pay Sheet (Fig. 4). At the end of the encoding period the hours worked were totalled and payment calculated according to the formula given in Fig. 4. Net bonus per cent was obtained from calculations described above. Basic hourly rates of payment at the termination of encoding were: encoder: R2,50; team checker: R3,20; bonus controller and supervisor: R3,75. Only encoders and team checkers participated in the bonus incentive scheme.

DATA CODING FORM AND KEY-PUNCHING OF DATA

An example of a completed encoding form is given in Fig. 5. The form was specifically designed to expedite the encoding operation and be suitable for key-punching on 80-column cards. The first line of the form was completed for the first sheet in each species cover only and this information was duplicated to following forms by computer until replaced by the next occurrence of a first line. Information common to all the specimens in a species folder, including identi-

ATTENDANCE REGISTER AND PAY SHEET

MONTH: JANUARY - FEBRUARY 1976 TEAM: 3

PERSONNEL NUMBER: S 100762431

SURNAME AND INITIALS: E. KING

HOURLY RATE: R2,50 A

NO. OF HOURS: 48 B

NET BONUS %: 41,6 C

TOTAL PAY: R169,92 D

$D = A \times B + (A \times B \times \frac{C}{100})$

WEEK	1	a	b	2	a	b	3	a	b	4	a	b	5	a	b	SUB-TOTALS
MONDAY	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS			INITIALS			INITIALS			8
	FROM/TO			FROM/TO			FROM/TO			FROM/TO			FROM/TO			
	HOURS	4		HOURS	4		HOURS	0		HOURS	0		HOURS	0		
TUESDAY	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS			INITIALS			12
	FROM/TO			FROM/TO			FROM/TO			FROM/TO			FROM/TO			
	HOURS	4		HOURS	4		HOURS	4		HOURS	0		HOURS	0		
WEDNESDAY	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS			15
	FROM/TO			FROM/TO			FROM/TO			FROM/TO			FROM/TO			
	HOURS	4		HOURS	4		HOURS	4		HOURS	3		HOURS	0		
THURSDAY	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS	E.K	E.K	INITIALS			13
	FROM/TO			FROM/TO			FROM/TO			FROM/TO			FROM/TO			
	HOURS	4		HOURS	4		HOURS	2		HOURS	3		HOURS	0		
FRIDAY	INITIALS			INITIALS			INITIALS			INITIALS			INITIALS			0
	FROM/TO			FROM/TO			FROM/TO			FROM/TO			FROM/TO			
	HOURS	0		HOURS	0		HOURS	0		HOURS	0		HOURS	0		
TOTALS		16			16			10			6			0		48

CERTIFIED CORRECT : BONUS CONTROLLER HFG.

FIG. 4.—Example (slightly reduced) of completed Attendance Register and Pay Sheet.

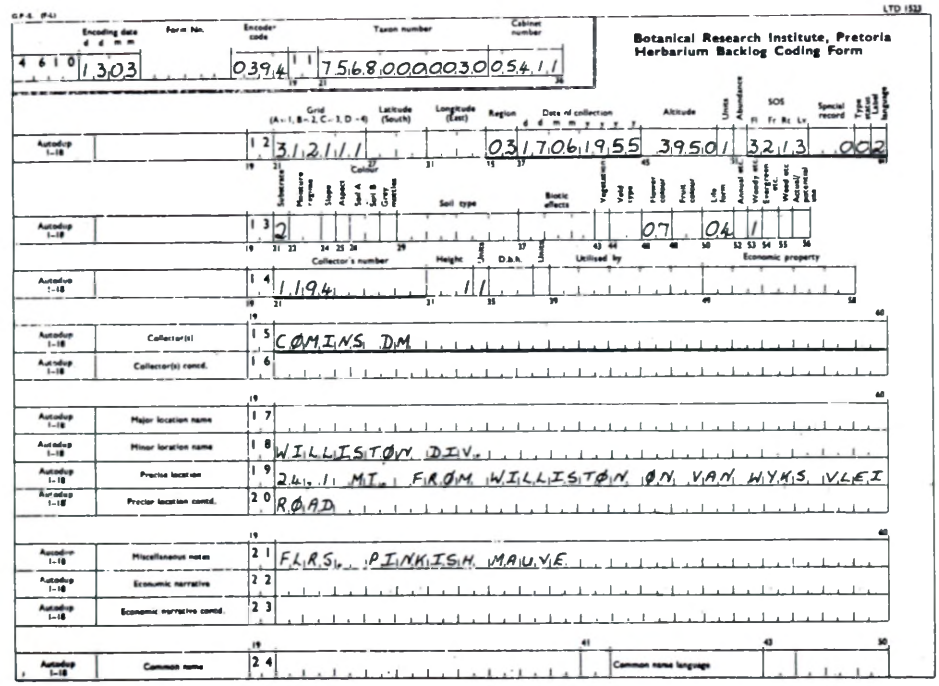

FIG. 5.—Example (reduced from A4 size) of completed Backlog Coding Form.

fication of encoder, genus and species code numbers and cabinet and shelf numbers, was included in this line. Subsequent lines contained information unique to the specimen and had to be completed for each specimen.

Shortly after encoding commenced, a 240-character per record key-to-tape data entry system replaced the IBM 029 card punches. If it had been known that such a system was due, a form with longer records and no 'auto dup' fields would have been designed. It is fortunate that it was not necessary to punch computer cards throughout the project as more than 2 300 boxes of cards would have been required and the task of processing and storing them would have been much more difficult than that of a few hundred magnetic tapes.

Encoding forms were numbered consecutively by the encoding supervisor by means of a numbering machine (Fig. 5) and then stapled into bundles of about 50. Each bundle was given a header form and consecutive bundle number. Batches of 20 bundles were submitted for key-punching and computer processing.

Specimens were accessioned piecemeal to computer disk packs in three groups of 150 000 and the remainder (about 60 000). The accessioning programme edited each specimen and listed errors. Errors were corrected by means of the same encoding form preceded by a different header form. Another programme finally converted correct and corrected specimens to the PRÉCIS data base.

ITEMS AND THEIR CODE EQUIVALENTS

It was decided that as far as possible, items would be given a numeric code (see Morris, 1974 for discussion). Exceptions were collectors' names, common names and locality names. Some descriptive text would be entered in full but would not be searchable. A list of items captured during the backlog encoding is given in Table 4 and an example of a complete backlog coding form is given in Fig. 5. Codes used under each item are listed below, together with notes on their use and conventions used by the encoders, where appropriate. The sequence is that of the encoding form.

TABLE 4.—Items encoded during backlog encoding (U=uncoded, C=coded, T=text)

1 Encoding date	26 Main fruit colour (C)
2 Encoder	27 Life form (C)
3 Taxon number	28 Annual/perennial (C)
4 Cabinet and shelf number	29 Woody/herbaceous (C)
5 Quarter-degree grid (C)	30 Evergreen/deciduous (C)
6 Latitude and longitude (U)	31 Weed/encroacher (C)
7 Region (C)	32 Actually or potentially used (C)
8 Date collected	33 Collector's number (U)
9 Altitude and units (C)	34 Height and units (C)
10 Abundance (C)	35 Diameter at breast height and units (C)
11 State of the specimen (C)	36 Utilized by (C)
12 Special record (C)	37 Economic property (C)
13 Type status (C)	38 Collector's name (U)
14 Label language (C)	39 Major locality name (U)
15 Substrate (C)	40 Minor locality name (U)
16 Moisture regime (C)	41 Precise location (T)
17 Slope (C)	42 Miscellaneous notes (T)
18 Aspect (C)	43 Economic narrative (T)
19 Soil colour (A and B horizons) (C)	44 Common name (U) and language (C)
20 Presence of grey mottles (C)	45 Number of duplicate sheets
21 Soil type (C)	46 Specimen or photograph of specimen (C)
22 Biotic effects (C)	
23 Vegetation type (C)	
24 Veld type (Acocks, 1953)	
25 Main flower colour (C)	

1. *Encoding date*

Two-digit codes for the month and day of encoding were recorded. The year of encoding was obtained from the batch header form.

2. *Encoder code*

Each encoder was assigned a unique, consecutive, three-digit code and each of the five encoding teams had a one-digit code. Thus 0394 represents encoder 39 from team four.

3. *Taxon number*

The seven-digit genus code and four-digit species code (see Morris & Glen, 1978) were taken from the outsides of the folders containing the specimens. An extra trailing zero was added to the species number by computer on conversion to allow for the future subdivision of an extant species into 100 new specific and sub-specific categories instead of only 10. The following special species codes were used:

9999 Tropical African type specimens and southern African material identified to genus only
9998 Hybrids (some hybrids also coded '9999')
9997 Cultivated, exotic species

Although a few were encoded, it was decided that in general exotic species would be omitted.

4. *Cabinet number*

Cabinets in each of the wings were given consecutive three-digit numbers, starting with '001' in each wing. Within each cabinet, shelves were numbered from top to bottom and left to right with two-digit numbers.

5. *Quarter-degree grid*

The latitude and longitude part of the code was entered as it appeared on the specimen. The letters, A, B, C and D, were converted to 1, 2, 3 and 4, respectively, before encoding. Where half- and quarter-degree codes were missing, a code of '0' was used.

6. *Latitude and Longitude*

Seconds, if given, were rounded to minutes before encoding. Where a range was given, the midpoint was coded.

7. *Region*

Black States within South Africa were coded as major or minor locations and for pragmatic reasons were given the region code of the province of which they were previously part. The codes used were:

01 Angola
02 Botswana / Bechuanaland/Ngamiland
03 Cape Province / *Kaapprovinsie*
04 Lesotho/Basutoland/*Basoetoeland*
05 Mozambique / *Mosambiek*
06 Natal
07 Orange Free State / *Oranje-Vrystaat*
08 Rhodesia/*Rhodesië*
09 South West Africa/Caprivi/Namibia
10 Swaziland
11 Transvaal
12 Not these
00 Unknown

Code 12 was used for other countries, in which case the name of the country was given as a major or minor location. The following region code conventions were adopted:

British Bechuanaland (usually Mafeking and surroundings)=03
Boesmanland (without mention of Province or country)=03
Great Namaqualand=09
Klein Namaqualand=03
Maputoland=06
Namaland=09
Namaqualand=03
n'Gamiland=02
Northern Ngamiland=09
Sekukuniland=11
Tembuland=03
Tongaland=06
Zululand=06

8. *Date collected*

The date on which the plant was collected was encoded in full. Missing dates, or parts of dates, were coded as '00'. In the case of a range being given, the earlier date was coded. The date on which the plant had flowered, according to the label, was used if the specimen had flowers and no other date was given. If the specimen was not flowering and only flowering date was given, date collected was entered as '00000000' (i.e. unknown) and flowering date was given as a miscellaneous note.

9. *Altitude and units*

Altitude was entered, right-justified. The midpoint of an altitude range was coded. The codes '1'= feet and '3'=metres were used for the units code.

10. *Abundance*

Four basic codes were used, each of which could be prepositioned by 'locally'. 'Locally', 'scattered' and 'local' without accompanying indication of abundance were, however, not coded. The codes were:

1 rare/very occasional/scarce/uncommon/*seldsaam/skaars*
2 locally rare/*plaaslik seldsaam*
3 occasional/infrequent/*toevallig/matig/nie volop nie*
4 locally occasional/very localized/*plaaslik toevallig*
5 common/fairly common/co-dominant/frequent/*algemeen/wydversprei/volop*
6 locally common/localized/*plaaslik algemeen*
7 abundant/frequently abundant/very frequent/*baie volop*
8 locally abundant/*plaaslik baie volop*

Additional codes for very rare and locally very rare have been added subsequently and distribution (i.e. local or widespread) has been separated from abundance. (In all cases where alterations have been made

to code lists the necessary alterations have been made to the data bank and the 'new' codes will be used exclusively in future.)

11. *State of specimen*

The presence and maturity of flowers, fruits, roots and leaves were coded after inspection of the specimen. The code for 'present' was not used if the encoder could determine whether the organ was mature or immature. For specimens of grasses and other families with reduced or small flowers, flowers and fruits were coded 'present' if either was visible. State of specimen of bryophytes was always coded '0000'. Side roots and root hairs had to be present before roots were coded 'present', 'mature' or 'immature'. Photographs of types were always coded '0000'. The following codes were used for the state of flowers, fruits, roots and leaves:

1 absent/*afwesig*
2 immature/*jonk*/*onvolwasse*
3 mature/*volledig ontwikkel*/*volgroei*
4 present/*teenwoordig*

12. *Special record*

A code was allocated for specimens collected for a special purpose. A rubber stamp or special label on the herbarium sheet usually indicated a special record. Codes allocated were:

000 not special record/*nie spesiale versameling nie*
001 Eland food study/*Elandkosstudie*
002 SKF alkaloid study/*SKF alkaloïedstudie*
003 Lamziekte survey/*Lamziekte opname*
004 Stijfsiekte survey/*Stijfsiekte opname*
005 Dunsiekte survey/*Dunsiekte opname*
006 Bovine staggers survey/*Stootsiekte opname*
007 Pollen studies (UOFS)/*Stuifmeel studies* (*UOVS*)
008 Anatomy study/*Anatomiese studie*
009 CSIR alkaloid survey/*WNNR alkaloïed opname*
010 Bushmen food study/*Boesmankosstudie*
011 Cancer research/*Kankernavorsing*
012 Economic plants of the Kung Bushman/*Ekonomiese plante van die Kung-Boesman*
013 Study of Tswana names and uses/*Studie van Tswana name en gebruikte*

13. *Type status*

All specimens found in type covers as well as specimens with a note of type status in ordinary folders were allocated a type status code. Where more than one type status was given for a specimen, the lowest code number was assigned, except for 'isotype' which was coded in preference to 'neotype', 'lectotype' and 'type'. The codes were:

0 no status/*geen status nie*
1 holotype/*holo-tipe*
2 syntype/cotype/*syn-tipe*/*ko-tipe*
3 paratype/*para-tipe*
4 neotype/*neo-tipe*
5 lectotype/*lekto-tipe*
6 type/*tipe*
7 isotype/*iso-tipe*

14. *Label language*

The language in which the majority of the original label was written was coded as follows:

1 Afrikaans
2 English/*Engels*
3 Latin/*Latyn*
4 Dutch/*Nederlands*
5 French/*Frans*
6 German/*Duits*
7 Italian/*Italiaans*
8 Portuguese/*Portugees*
9 Spanish/*Spaans*
0 other/unknown/*ander*/*onbekend*

15. *Substrate*

This item was coded as follows (present code numbers in parenthesis):

1 soil/mud/sandy flats/*grond*/*modder*/*sandvlaktes* (01)
2 stony soil/between rocks/*klipperige grond* (02)
3 bare rock/*blootgestelde rots* (03)
4 talus/scree (04)
5 cliff face/rock crevices/*kranswand* (05)
6 termite mound/termitaria/*termiethoop* (06)
7 beach dunes/dune forest/*kusduine*/*duinwoud* (07)
8 desert dunes/*woestynduine* (08)
9 other/plant growing on another plant/*ander*/*plant groei op ander plant* (10)

An additional code, 'in water' (09) has been added subsequently. In the case of coding 'other', details were entered under miscellaneous notes. Plants noted to be growing in or on granite, dolomite, sandstone or another rock type were coded '2' and if the colour of the rock was given, it was coded as the colour of the B-horizon soil (see below). Sandy flats were coded '1' for substrate, '02' for soil type and '1' for slope. Dune forest was also coded as vegetation '7'. Code '1' (soil) was only used if one of the following words, or their Afrikaans equivalents, occurred on the label: clay, gravel, sand, sandy, soil, turf.

16. *Moisture regime*

This item was coded in conjunction with substrate, using the following codes:

01 poorly drained soil/*swak gedreineerde grond*
02 well-drained soil/*goed gedreineerde grond*
03 pan/depression edge of pan/*holte*
04 seepage area/*syfergebied*
05 marsh/swamp/bog/*vlei*/*moeras*
06 floodplain/*vloedvlakte*
07 river/streambank/near river/*rivier*/*spruitoewer*
08 river/stream/burn/*rivier*/*stroom*
09 river or stream bed/*rivier- of spruitbedding*/omurumba/oshona
10 ditch/donga/furrow/water course/*sloot*/*spoelsloot*
11 lake/dam/weir/*meer*/*dam*/*stuwal*
12 estuary/sea/lagoon/river mouth/littoral/mangroves/*strand-meer*/*riviermond*/*see*
13 in water
14 other/*ander*

Code '07' was also used for 'next to river', 'near creek', 'riverine' (except 'riverine forest' and 'riverine bush', which were coded '7' under vegetation only, 'waterfall' and 'stream bank'. 'Above river' and place names (e.g. Tugela Mouth) were not coded here. All specimens coded '13' have been given substrate code '09' and code '13' has been removed from moisture regime.

17. *Slope*

The codes for slope were:

1 plain/flat/sandy flats/*vlakte*/*gelykte*
2 gentle/*effens*
3 moderate/*matige*
4 steep/*steil*

'Plateau' and 'slope' without qualifiers were not coded.

18. *Aspect*

Aspect was coded on the following eight-point scale, provided it was clear that the collector was referring to aspect and not location (i.e. south of . . .):

1 north/*noord*
2 north east/*noordoos*
3 east/*oos*
4 south east/*suidoos*
5 south/*suid*
6 south west/*suidwes*
7 west/*wes*
8 north west/*noordwes*

19. *Soil colour*

On very few occasions, soil colour was mentioned on labels. Where the soil horizon was not named, the colour was coded as that of the A-horizon. Rock colour was coded as that of the B-horizon. The codes used were:

1 white/*wit*
2 light grey/*lig grys*/*vaal*
3 grey/grey-brown/beige/fawn/*grys*/*grysbruin*
4 yellow-brown/*geelbruin*
5 red/*rooi*
6 black/*swart*

20. *Grey mottles*

A code for the presence of grey mottles within the soil profile was included on the advice of soil scientists but as this information was virtually never given by collectors, the item was not recorded after about 50 000 specimens had been encoded. It is now included in the data bank as code '11' of soil type.

21. *Soil type*

The soil type was described by up to four of the following codes:

01 gravel/shale/gritty/*growwe sand/gruisgrond*
02 sand/sandy/*sand*
03 loam/*leem*
04 clay/turf/*klei/turf*
05 humus-rich/peaty/*humus-vrugbaar*
06 salty/*brak*
07 on calcrete/limestone/calcareous soil/*kalkklip*
08 on laterite/ferricrete/*ouklip*
09 disturbed soil/*versteurde grond*
10 eroded/*geërodeer*
11 other/*ander* (12)

'Grey mottles' (11) were subsequently added and code '11' was changed to '12'. 'Sandveld', 'sand flats' and 'sand forest' were coded '02'; 'kalkveld' was coded '07' under this item and vegetation type was coded '4'. 'Sand on calcrete' was coded '02' and '07'. 'Dolerite soil' and 'alluvial soil' were coded as substrate '1' only. 'Disturbed soil', included as code '09' of soil type was not coded again under biotic effects.

22. *Biotic effects*

Biotic effects noted on labels were coded by means of up to three of the following codes:

01 cultivated land/ploughed/*landerye*
02 abandoned land/fallow/*ouland*
03 planted pasture/*aangeplante weiding*
04 plantation/plantation margin/*plantasie*
05 garden/lawn/*tuin*
06 roadside/railwayside/*langs pad/langs treinspoor*
07 heavily grazed/trampled/*swaar bewei/oorbewei/uitgetrap*
08 recently burnt/*onlangs gebrand*
09 disturbed-other/*versteurd-ander*
10 no effect seen/undisturbed/rested/*geen invloed waargeneem nie*

The following conventions were adopted in the coding of biotic effects. A firebreak was not considered as '08' and disturbed soil was not '09'. In a National Botanic Garden, or similar institution, code '05' was only used if it was stated that the specimen was cultivated. If the locality was a domestic garden, or the locality could be assumed to be other than a National Botanic Garden or similar institution then code '05' was assigned even though the specimen was not specifically said to be cultivated. Code '08' was used for burns of up to one year old, if the age of the burn was given.

23. *Vegetation*

The following codes were used for describing the vegetation in which the plant was collected:

1 desert/semi-desert/*woestyn*
2 karoo/*karooveld*
3 grassland/sandveld/*grasveld/veld*
4 savanna/bushveld/thornveld/tree veld/open veld/parkland/*grasveld met bome/kalkveld*
5 woodland/bush/*boomveld/bos*
6 scrub/thicket/among shrubs/*digte bos/kreupelhout/ruigte*
7 forest/sand forest/dune forest/riverine forest/forest margin/*woud*
8 fynbos/heath/macchia/sclerophyll scrub/*fynbos*
9 other/*ander*

The term 'bush' was coded '5' unless associated with a measurement of height, in which case the term was taken as a life form. If vegetation had been recently burnt, the vegetation type was assumed to be grassland unless another vegetation type was given.

24. *Veld type*

The two-digit veld type number assigned by Acocks (1953) was coded when given on the label.

25. *Flower colour*

The dominant flower colour, as noted on the label, and not as observed on the dried specimen, was coded as follows:

01 white/cream/*wit/room*
02 grey/silver/*grys*
03 yellow/saffron/*geel*
04 pink/rose/cerise/*pienk*
05 orange/amber/*oranje*
06 red/magenta/scarlet/maroon/*rooi/persrooi/skarlaken/wynkleur/karmosyn*
07 mauve/purple/violet/heliotrope/lilac/vermilion/*pers*
08 blue/*blou*
09 green/*groen*
10 brown/buff/*bruin*
11 black/*swart*

26. *Fruit colour*

The flower colours given above were also used for fruit colour. The codes were in each case greater by 20. Thus, white fruit was '21' and black '31'.

27. *Life form*

The following codes were used:

01 tree/*boom*
02 shrub/undershrub/suffrutex/*struik*
03 dwarf shrub/semi-shrub/half shrub/*dwergstruik/halfstruik*
04 herb/forb/hemicryptophyte/*kruid*
05 geophyte/bulb/bulbous plant/*geofiet/bolplant*
06 epiphyte/*epifiet*
07 climber / creeper / scandent / vine / twiner / liane / *klimplant/ranker*
08 parasite/*parasiet*
09 succulent/semi-succulent/*vetplant*
10 aquatic/hydrophyte/*waterplant*

Plants with succulent or fleshy leaves were not coded '09' unless life form was also specified.

28. *Annual etc.*

The life cycle of the specimen was coded as follows:

1 annual/*eenjarige*
2 ephemeral/*efemeer*
3 biennial/*tweejarige*
4 perennial/perennial rootstock/*meerjarige*

When 'herbaceous' was coupled with 'woody rootstock' the specimen was not coded as '4'.

29. *Woody etc.*

This item was coded as follows:

1 woody/*houtagtig*
2 herbaceous/*kruidagtig*

'Woody rootstock' was not coded under this item. An additional code, half-woody, has been added subsequently.

30. *Evergreen etc.*

This item was coded as follows:

1 evergreen/*immergroen*
2 semi-deciduous/*half bladwisselend*
3 deciduous/*bladwisselend*

31. *Weed etc.*

This item was coded as follows:

1 weed/ruderal/*onkruid/puinhoopplant*
2 encroacher/*indringer*
3 weed & encroacher/*onkruid en indringer*

An additional code for exotic plants has been added subsequently.

32. *Actual/potential use*

Actual uses as noted on labels (e.g., used as a poison) were separated from potential uses (e.g., could be a food plant) by means of this code. Use of this item meant that utilized by and economic property had to be entered as well. 'Poison' without further elaboration was taken as an actual use. Codes were:

1 actually used/*werklik gebruik*
2 potential use/*potensieël bruikbaar*

33. *Collector's number*

The collector's number, exactly as given on the label, was entered, left-justified. If no collector's number was given on the label, a PRE accession number, prefixed by 'PRE' was used for this item. If the specimen had a printed label of the Transvaal Museum (TM), Herbarium of the University of the Witwatersrand (MOSS), South African Museum (SAM), National Botanic Gardens (NBG), Herbarium of the University of the Orange Free State (BLF), Stellenbosch Unit of the National Herbarium (STE), Durban Unit of the National Herbarium (NH), Herbarium of the Forestry Department (FD) or Bolus Herbarium (BOL) and the number was that of the herbarium and not the collector, the 'collector's number' was prefixed with the letters given above in parentheses. Where specimens had both a personal collector's number as well as a herbarium accession number, the former was always used. If the initials of the collector were given as part of the number, they were omitted.

34. *Height and units*

The height of the plant, rounded to three digits if necessary was coded. The mid-point of a range was entered. When two or more distinct heights were given, the greater height was used. Units, for both height and diameter at breast height, were coded as follows:

1 feet/*voet*
2 inches/*duim*
3 metres/*meter*
4 centimetres/*sentimeter*

35. *Diameter at breast height and units*

The conventions noted under height also applied to the coding of this item.

36. *Utilized by*

Up to five codes were used to describe the uses of the plant. If more from one code was used then only one economic property was allowed, and *vice versa*. The following codes were applicable:

81 man/*mens*
82 stock/*vee*
83 cattle/*beeste*
84 sheep/*skape*
85 goats/*bokke*
86 horses/donkeys/*perde/donkies*
87 other mammals/game/*ander soogdiere*
88 birds/*voëls*
89 fishes/reptiles/amphibians/*visse/reptiele/paddas*
89 fishes/reptiles/amphibians/*visse/reptiele/amfibieë*
90 invertebrates (excluding bees)/*ongewerweldes* (*uitsluitende bye*)
91 honey bees/*heuningbye*

Codes '90' and '91' were subsequently reversed so that 'bees' became '90' and 'other invertebrates' '91'. Unless otherwise specified, 'eaten' and 'not eaten' on labels were assumed to apply to stock (code '82') and 'poison' was not coded under the utilized by item unless qualified.

37. *Economic property*

Up to five of the following codes were used to describe economic properties of specimens (present code numbers in parentheses):

01 poison/*gif* (01)
02 poison: arrow/*pylgif* (02)
03 poison: criminal/*gif: kriminele doeleindes* (03)
04 medicinal/*medisyne* (06)
05 drug/*verdowingsmiddel* (07)
06 irritant/allergy/*prikkelmiddel/allergie* (08)
07 tainting e.g. milk/*bysmake aan melk ens.* (09)
08 magic/ritual/*toor kuns/ritueel gebruik* (10)
09 eaten/palatable/*geëet* (11)
10 not eaten/*nie geëet nie* (12)
11 beverage/*drank* (13)
12 cordage/*toue/vlegwerk* (14)
13 paper/*papier* (15)
14 clothing/*klere* (19)
15 structural/baskets/mats/brooms/*gebruik vir konstruksie/mandjies/matjies/besems* (20 and 21)
16 aromatic/snuff/*aromaties/snuif* (22)
17 cosmetic/*grimering* (23)
18 beads/*krale* (24)
19 soap/*seep* (26)
20 oil/*olie* (27)
21 gum/resin/*gom/harpuis* (28)
22 dye/*kleurstof* (29)
23 fuel/*brandstof* (30)
24 sand binder/*sandbinder* (31)
25 ground cover/lawn grass/*grondbedekking* (32)
26 hedge/windbreak/*heining/windskerm* (33)
27 shade/*skaduwee* (34)
28 garden ornamental/*sierplant* (35)
30 other—see miscellaneous notes/*ander—sien algemene notas*
30 other—see miscellaneous notes/*ander—raadpleeg algemene notas*

Where more than one distinct economic note was given for a specimen, the first was coded and the subsequent notes entered in the economic property narrative field only. The following codes have been added subsequently:

medicine—internal (4)
medicine—external (5)
thatching (16)
timber (17)
other building (18)
other decorative (25)
other horticultural (36)
crop (37)

38. *Collector(s)*

In the case of one-person collectors, the surname, followed by initials without full-stops were entered. Where more than one collector's name was given on the label, surnames only were coded, each separated by two blank spaces. Where no collector's name was given, the collector was coded as 'PRE'. When a specimen was collected by one person (usually a layman) for another (usually a well-known botanist) and both names appeared on the label, the name and number of the botanist were encoded and the name of the actual collector was given in a miscellaneous note.

39. *Major location name*

Difficulty was experienced by encoders with the choice of major and minor location names. Pages of examples were drawn up to illustrate some of the many possibilities.

A major location name was only given if a minor location name was also coded; otherwise the major location name was entered in the minor location name field. It was intended that a large geographical unit, but not a region (see above) be entered. Examples are mountain ranges and magisterial districts.

40. *Minor location name*

It was intended that a small geographical unit be coded as the minor location name. Examples would be towns and cities. In general, a minor location is

TABLE 5.—Alphabetical list of abbreviations used in major, minor and precise locality fields

Afdeling	=AFD.	*Landbou*	=LANDB.
Afdelingsraad	=AFD. RAAD	Metres/*Meters*	=M.
Agriculture/Agricultural	=AGR.	Miles/*Myl*	=MI.
Between	=BETW.	Mount/Mountain	=MT.
Bosbou	=BOSB.	*Navorsingstasie*	=NAV. STA.
Cultivated	=CULT.	*Nasionale*/National Park	=NAT. PARK
Centimetres	=CM.	Nature Reserve/*Natuurreservaat*	=NAT. RES.
District	=DIST.	Near	=NR.
Distrik	=DIST.	North/*Noord*	=N
Division	=DIV.	Northwest/*Noordwes*	=NW.
Duim	=DM.	Pasture	=PAST.
East/*Oos*	=E.	Research	=RES.
Experimental	=EXP.	South/*Suid*	=S.
Feet	=FT.	Station/*Stasie*	=STA.
Flowers	=FLS.	University	=UNIV.
Forestry/Forest	=FOR.	*Voet*	=VT.
Inches	=IN.	West/*Wes*	=W.
Kilometres	=KM.		

situated within a major location. The collecting site had to be within the arbitrary limit of 16 km (10 miles) from the minor location or 'near' the minor location, if an exact distance was not given. Farm names were not used for major and minor locations except in South West Africa and Botswana, or where no other locality name was given. A list of abbreviations was prepared (Table 5) and these were used throughout for the spelling of location names. Points of the compass in names (e.g. Pretoria North) were not abbreviated. Specimens collected between two towns or on the boundary of two magisterial districts were not given major or minor location names through the use of these town or district names.

41. *Precise location*

The precise location, as given on the label, was recorded on two lines of the coding form. This entry was made as complete as possible without unnecessarily repeating the major and minor locations. The standard list of abbreviations was used in this field as well.

42. *Miscellaneous notes*

The most important information not coded anywhere on the data form was entered in this field.

43. *Economic narrative*

The utilized by and economic property items were elaborated upon in this field.

44. *Common name and language*

The common name was entered in this field as it appeared on the label. The language of the entered common name was coded from the following list (present code numbers in parentheses):

01 Afrikaans (01)
02 English/*Engels* (02)
03 German/*Duits* (03)
04 Dutch/*Nederlands* (04)
14 Basotho/*Basoetoe* (05)
05 Bushman/*Boesman* (06)
06 Damara (07)
07 Herero (08)
08 Himba (09)
09 Kavango (10)
10 Mafwe (11)
11 North Sotho/Sipedi/*Noord-Sotho* (12)
12 Ovambo (13)
18 Shangaan (19)
13 South Ndebele/*Suid-Ndebele* (14)
14 South Sotho/*Suid-Sotho* (15)
15 Subia (16)
16 Swazi (17)
17 Tjimba (18)
18 Tsonga (19)
19 Tswana (20)
20 Venda (21)
21 Xhosa (22)
22 Zulu/*Zoeloe* (23)
00 other/unknown/*ander/onbekend*

The label language codes were kept separate from the common name language codes.

45. *Number of duplicate sheets*

If there was more than one complete specimen of a collector's number within a species folder, the number of duplicates was entered in column 43–44 of the last line of the form. Where parts of a specimen were mounted on more than one sheet, these were not considered as duplicates and all such sheets were taken into account in the determination of the state of the specimen (see above).

46. *Photograph code*

Specimens encoded from photographs mounted on herbarium sheets and filed in species covers were identified by means of a code 'PH' on the data form in columns 45–46 of the last line of the form.

COSTS AND DURATION OF OPERATION

It was anticipated that encoding of all the estimated 500 000 specimens housed in the National Herbarium would take the encoding team two to three years. Coding was begun during June 1974 and was effectively completed by December 1976. As a result of continuing specimen collection (some 20 000 new accessions per annum) encoding will continue until all collectors are using the new collector's form (see Morris & Glen, 1978: Fig. 3). A total of over 470 000 specimens were encoded (Table 6) and 38 500 man-hours were worked. The average cost per specimen was R0,258 and the average cost per man-hour was R3,14.

TABLE 6.—Statistics of the backlog encoding project

Encoding initiated	1974/06/03
Encoding terminated	1976/12/06
Number of specimens coded	470 243
Total man-hours worked	38 557
Total labour cost	R121 215
Average cost per specimen	R0,2578
Average cost per man-hour	R3,1438

Costs and statistics for each of the encoding periods for which Monthly Team Summaries were calculated are summarised in Table 7. Initially, costs per specimen were high as a result of intensive training of encoders without botanical background and the use of inefficient procedures. Costs dropped rapidly and stabilised at R0,23 to R0,26 per specimen while the norm was gradually increased. Towards the end of the period, costs rose slightly as a result of basic pay rate increases being awarded to encoders.

TABLE 7.—Herbarium backlog encoding monthly statistics

Report number	Month and year	Norm	Number of specimens coded	Man-hours worked	Amount paid	Cost per specimen
					R	R
1	6/74	9	3701	1347,5	2926,50	0,79
2	6/74	9	14554	1804,0	5604,05	0,39
3	7/74	11	25957	2106,0	6919,96	0,27
4	8/74	12	17073	1486,0	4811,30	0,28
5	9/74	14	19854	1605,0	4836,50	0,24
6	10/74	14	28400	2070,0	6379,30	0,22
7	1/75	14	22152	2064,0	6503,06	0,29
8	2/75	14	18832	1560,0	4914,75	0,26
9	3/75	14	25509	1995,0	6560,96	0,26
10	5/75	14	14053	1013,0	3557,06	0,25
11	5/75	14	23471	1669,0	5136,24	0,22
12	6/75	16	20566	1613,0	5119,18	0,25
13	7/75	16	14479	1144,5	3584,12	0,25
14	8/75	16	20649	1520,5	4763,84	0,23
15	9/75	16	20191	1580,0	5137,60	0,25
16	10/75	16	25615	1800,5	6198,43	0,24
17	1/76	16	13234	1219,0	3591,89	0,27
18	1/76	16	20858	1609,0	4699,44	0,23
19	2/76	16	21296	1643,0	4913,60	0,23
20	3/76	16	14837	1132,0*	3453,48	0,23
21	4/76	16	13560	1062,5*	3346,38	0,25
22	5/76	16	10192	767,0*	2507,17	0,25
23	6/76	16	13672	1165,5*	3633,97	0,27
24	7/76	16	12655	956,0*	3024,77	0,24
25	8/76	16	8594	625,0*	2163,98	0,25
26	9/76	16	9934	715,5*	2635,29	0,27
27	10/76	16	10638	864,0*	2776,53	0,26
28	11/76	16	5717	420,5*	1515,77	0,27

* Encoding for 3 hours per evening instead of 4 as previously.

DISCUSSION

The success of the project is attributed primarily to the bonus incentive scheme and the availability of a comprehensive encoder instruction manual. The bonus scheme and the part-time staff of 28 ensured that the task was completed within a relatively short period. For about the same cost, the operation could have been carried out in about 10 years by a smaller, full-time staff. Under those circumstances, it is likely that interest in the project would have waned and also that encoding standards and conventions would have altered more than under the circumstances actually imposed. Availability of an instruction manual facilitated the training of staff, enabled encoding standards to be maintained and allowed the use of an enforceable penalty scheme.

The codes, including synonyms and Afrikaans equivalents are given in full in this paper together with the coding instructions given to encoders. In this way, users of the data bank are informed about the conventions adopted and generalisations made. Any future operation of this kind should find this a useful foundation on which to build. The tables of Latin and German equivalents are included for the same reasons. It is not intended that where more than one word is given for a code they should be considered as synonyms in the strict sense but they indicate the range of variation within the concept included in the code. Translation lists from foreign languages should be used to obtain an understanding of the philosophy behind the use of ite n codes; my lists are not dictionaries. Use was made of inferences and assumptions in order to define codes clearly and to spell out encoder instructions. In a few cases, the codes indicated will be incorrect or misleading to a trained botanist in the context of all the information contained on a label. It was, however, impossible to define all possible situations and a few such errors will obviously have escaped detection.

Most problems were experienced with the computer processing of the data and the physical handling of large volumes of data sheets, printouts and punch cards. Although computer processing falls outside the scope of this contribution, these problems are mentioned here. At initiation, this was the largest system on the Burroughs B5700 computer and although the computer was large enough, the operations room was not capable of handling the volume efficiently and keeping concise records of processed data tapes and system backup tapes. Although the problems diminished as experience was gained, a great deal of time was lost as a result. The need for efficient computer room control procedures with a project such as this cannot be over-emphasized. The volume of paper output to be scanned for errors was also underestimated and as a result handling of large stacks of paper was difficult. The single printout of every specimen which was made for manual checking purposes used over 100 boxes of computer paper. Storage of paper and punched data sheets in an accessible manner was also difficult to plan. Solutions would have been the use of computer output microfilm and a re-designed output format requiring fewer lines per specimen so that more specimens could have been listed on each page. It is fortunate that after about 200 boxes of computer cards had been punched a switch was made to key-punching onto magnetic tape. As mentioned earlier, handling of magnetic tapes is far easier than boxes of cards. It was found that computer processing could not keep pace with encoding and, similarly, error correction could not keep pace with computer processing. Processing and data correction therefore lagged further and further behind as the project continued.

It was hoped that manual vetting and correction of computer printouts could be made part of the bonus incentive scheme so that it could be completed in a short period as a distinct operation but this was not possible and vetting was carried out by the small team of permanent encoders who have taken over from the backlog encoding team.

The data bank is open-ended in that new accessions (about 20 000 a year) are being added continually. A special field data sheet has been designed for use by collectors who regularly submit specimens to the National Herbarium to obviate the need for encoding. A small team of permanent encoders has been responsible for encoding donations and exchanges not accompanied by the new field data sheet since the conclusion of the project reported here. This team is also responsible for the continuing correction of errors in the PRÉCIS data bank.

The error detection systems included in the bonus incentive scheme were successful in reducing the errors to an acceptable level. With the large volume involved, however, a certain number of errors were bound to slip through. A continuing programme of error correction will be maintained for a few years while the data bank is used and errors are discovered. The global editing of localities and collectors' names was planned from the outset and is proceeding satisfactorily. From alphabetical lists of all collectors and localities, mis-spellings are being removed by a qualified botanist. The addition of quarter-degree grids by computer from the list produced by Leistner & Morris (1976) at a later stage is being considered.

The costs given in this paper do not include those of key-punching, computer program writing and computer processing time. Also excluded are the costs of the manual checking of computer printouts and completion of error correction data forms. The costs of the systems analysis and time spent by the writer on the planning of the project are also omitted. These costs are not available but it is estimated that, if included, they would double or treble the expenses of encoding reported here. The benefits which the botanical community will gain from the project are, however, such that the expense will be amply justified.

ACKNOWLEDGEMENTS

It is most unlikely that this project would have been initiated without the enthusiastic support of Dr B. de Winter, Director of the Botanical Research Institute. The Workstudy Section of the Department of Agricultural Technical Services, and Messrs G. P. Kleynhans, F. J. S. van Biljon and J. L. Pieters in particular, assisted by streamlining procedures and proposing the bonus incentive scheme. Mr G. J. Smit and the staff of the Data Processing Section of the Department contributed to the success of the project by making computer-related facilities available when required. The ALGOL computer program for accessioning data was written by Mrs E. de Bruyn. The staff of the National Herbarium, and in particular the then Curator, Dr O. A. Leistner, and Mrs E. van Hoepen, were of assistance in the preparation of the herbarium for encoding. I thank Drs J. H. Ross and A. V. Hall for encouragement and advice during the planning stages of this project and Drs O. A. Leistner and H. F. Glen for commenting on drafts of this manuscript.

UITTREKSEL

Die enkodeering van inligting vanaf 470 000 *bestaande herbariumeksemplare vir gebruik in 'n rekenaardatabank word beskryf as hulpmiddel vir kuratore en stelselontleerders wat gelyksoortige prosedure in hul eie herbariums wil toepas. Die voorbereiding van die herbariums, die enkoderingsprosedure, die nagaan van geënkodeerde eksemplare, koste en vergoeding van enkodeerders word beskryf. 'n Volledige lys van geënkodeerde items en hul kodes word gegee. Van die probleme wat aangetref is, word bespreek en die sukses van die projek tot op datum word kortliks geëvalueer. Sukses word toegeskryf aan die bonus-aansporingsstelsel en die beskikbaarheid van 'n volledige instruksieboek.*

REFERENCES

ACOCKS, J. P. H., 1953. Veld types of South Africa. *Mem. bot. Surv. S. Afr.* 28: 1–192.

BRENAN, J. P. M., ROSS, R. & WILLIAMS, T., 1975. *Computers in botanical collections.* London: Plenum Press.

DYER, R. A., 1975. *The genera of Southern African flowering plants.* 1. *Dicotyledons.* Pretoria: Botanical Research Institute.

DYER, R. A., 1976. *The genera of Southern African flowering plants.* 2. *Gymnosperms and Monocotyledons.* Pretoria: Botanical Research Institute.

LEISTNER, O. A. & MORRIS, J. W., 1976. Southern African place names. *Ann. Cape Prov. Mus.* 12: 1–565.

MORRIS, J. W., 1974. Progress in the computerization of herbarium procedures. *Bothalia* 11: 349–353.

MORRIS, J. W. & DU TOIT, P. C. V., 1976. *Handbook for the encoding of herbarium specimens.* Pretoria: Botanical Research Institute (unpublished report).

MORRIS, J. W. & GLEN, H. F., 1978. PRÉCIS the National Herbarium of South Africa (PRE) computerized information system. *Taxon* 27: 449–462.

MORRIS, J. W. & LEISTNER, O. A., 1975. Progress with computerization of the National Herbarium, Pretoria. *Boissiera* 24: 411–413.

PIETERS, J. L., 1974. *Verslag oor die ondersoek na die kodering van plantmonsters in die Nasionale Herbarium van die Navorsingsinstituut vir Plantkunde met die oog op rekenarisering gekoppel aan 'n direkte finansiele aansporingstelsel.* Pretoria: Department of Agricultural Technical Services (unpublished report).

ROSS, J. H., LEISTNER, O. A. & DE WINTER, B., 1977. *A guide to contributors to the Flora of Southern Africa* (*F.S.A.*). Pretoria: Department of Agricultural Technical Services.