SAFARIS Brief Tutorial

**Southern African Friends and Researchers Indexing Specimens**

When biodiversity specimens are collected there is typically specific information recorded by the person doing the collector that includes the collector, the date of the collection, the place and habitat where the collection was made, and abundance of the species in the area, and sometimes a description of the individual organism. Plants are generally prepared as dried, pressed specimens and then added to a herbarium. They are typically mounted on board with a label that gives information about who collected the specimen, the date and place of collection, the identification of the plant species, some information about the locality and plant, and possibly some other information. Mammals, insects, fungi and other organisms would have many similar details. A collector will often keep a register of his or her collections. This SAFARIS project involves text information from images of specimens or field notes being transcribed so that the image can be located during online searches. Often a user viewing the image of a specimen can see sufficient detail in the image in order to do the research and not always having to either travel to, or loan the specimen from, the herbarium or museum where the specimen is housed.

This document gives a brief introduction to the process of transcribing information. At present the images are hosted on internet servers in Australia.

Index all information from the image

Index all the information on each record of the collecting register as highlighted in Figure 2. Enter the text in the box ‘Index all text’. Hover over the field help to see key combinations to add special characters.

A spreadsheet template simplifies data capture, especially for information repeated from record to record. Type the information into the green cells and then paste the formatted data from the red cells into the transcription window.
The image to be indexed. Using the spreadsheet, one types the information into the green cells. This can be used on the lower half of the screen and the image in the upper part of the screen using the to the right to view the image in a new window while transcribing the text.

Information transcribed from the image.

Green cells with transcribed text and red cells with formatted text.
- Please index all of the information as it was written.
- We suggest that you not expand on abbreviations.
- Simply record what is written (or what it appears that Acocks intended as his edited end-result).
• Additional items under the species name such as ♂ (alt-11) and ♀ (alt-12) would get in the way of matching names to the checklist, so please record them under Subtext.
• It is not necessary to expand on names. We shall use a control checklist for matching names to valid species names.

• Please add an Altitude heading before the F [Frequency] heading for recording elevation [altitude] if recorded.

• It seems reasonable that if the Habitat is the same as a preceding record (it has a Do. for ‘ditto’) then, if there is an altitude, it can be repeated for the same habitat, although Acocks does not repeat the altitude.

• Repeat the text from the line above instead of typing Do. for ‘ditto’. The spreadsheet template generally does this automatically.

• Use the | (vertical bar) to separate the information under each column heading. This is specifically for headings given by Acocks (e.g. Herb No: and Date:).

• Symbols such as ✓ [ticks] and • [dot] should be represented as v (lower case V) in place of tick marks, and o (lower case O) in place of dots to simplify reading of characters by the computer. Their meaning is not always clear, so it is probably not critical how or if they are captured.

• If a herbarium name is included at the top of the Herb column, it should be captured in the herbarium number column as shown. This is accommodated as a separate column in the data-entry spreadsheet to facilitate the default automatic sequential numbering.

• Text written between lines refers to the line above. This is the smaller text slightly below the main text and should be indexed as Subtext in a field following the F [Frequency] heading as a separate entry following a | (vertical bar).

• F [Frequency] heading as a separate entry following a | (vertical bar). Use [ to signify overstripe, > to signify upper case, ^ to signify superscript, / for up arrow, \ for down arrow.
• Note it is not necessary to expand abbreviations, e.g. no. and R.

Fish R. valley at Hunts' Drift. (Flrd. In Estcourt, Aug., 1944)
Misgunst: along railway on hillside in T.G.V.

• Use the colon (:) to separate information under the same column heading (specifically under the heading Locality, habitat etc.). It is captured in separate columns in the spreadsheet template and the template then inserts the colons.

Umhlali: along railway through sugar-cane: bushy banks
Stanger: by roadside: in patches

• A character crossed out such as Augrabies should not be typed, e.g. Augrabies.

• Text crossed out such as Doryalis potulidifolia would be transcribed as Doryalis tristis (Sond.) Sim.

• Indecipherable text can be typed as [???].

[???]ents

• Smudges, or what do not appear to be written information, can be typed as [xxx].

• If dates or numbers are out of sequence, simply record them in the sequence as written.

• Comments made by the transcriber for a particular record can be added between [] in the Sub text column.

[date out of sequence]

• A sketch can be recorded as [sketch].

[simple sketch of leaf at end of Name column]

• Some functions of the Field notes template are not relevant to the collecting register tasks. For example, '2. Where a species or common name appears in the text please enter any relevant information into the fields below' is not active for this project.

• A spreadsheet template simplifies data capture, especially for information repeated from record to record. Type the information into the green cells and then paste the formatted data from the red cells into the transcription window.

• It may be useful to keep this tutorial or a document open with reminders of how to represent characters such as ✓ for [ticks], • for [dots], o+ for ♀ and o/ for ♂, 1/2 for Half (½)
• In Excel you can type `ctrl-' (ctrl-apostrophe) to copy the information from the cell above. It may help, if not using the spreadsheet, to use the copy (ctrl-c) and paste (ctrl-v) keystrokes to copy text from the previous line, e.g. where Acocks wrote Do. (ditto) for the District and Locality, habitat, etc.

• Please read off a map latitude and longitude coordinates (in decimal degrees) and a precision (radius in metres) for the locality, and the name of the person who did the georeferencing. Type these into four columns LatDD|LongDD|Prec(m)|GeorefBy. Maps and an extensive gazetteer are available for use in GIS (ArcGIS, QuantumGIS, DIVA GIS or other) that will be of great value in georeferencing Acocks collections. A GIS shapefile giving Acocks’ collecting and sampling record by date also help to indicate where he was likely to have been at a given time.

We continue to learn in this indexing work and will try to use observations from indexers to improve the SAFARIS tutorial. The ultimate aim of this project is to index the information in these images so that the images are searchable online, and so that the information is available for various applications such as species modelling and vegetation classification.

Locality information is particularly useful to researchers, so please be sure to record it as accurately as possible. Names, places and routes would best be found on field maps from about the time of the collection rather than using current maps. For example, it is useful to use the Topographical map of the Union of South Africa 1:500 000 map series of 1936 with marks made by Acocks for finding most places that he described. You can refer to these maps on BGIS.SANBI.ORG.ZA together with other scanned maps for southern Africa.

We recommend using the following column headings.

Book|Task|LHPage|RHPage|Collector|CheckFlag|Collector No|HerbName|Herb No|Date|Name|Family|Det by|District|Locality|Habitat 1|Habitat 2|Habitat 3|Altitude|F|Sub text|Remarks|LatDD|LongDD|Prec(m)|GeorefBy|LatDDAudit|LongDDAudit|PrecAudit|GeorefAuditBy|TmscrbdBy

Type text continuing across from left to right across the two pages as single entries for each record.

Keep indexing each record into a new row until all the records on the page have been entered. Copy the formatted text in the red cells of the spreadsheet template if you use that, and paste it into the Verbatim Text box of the web browser.

Once all the rows have been completed, type additional comments in the Notes box at the bottom that may assist with the validation of the task.