

## PANDORA PMERGE software

This document is a response to the document titled "The Merge process", which is dated 23 August but was not received in Edinburgh until mid-September. The same paragraph numbers are used but without quoting verbatim. A CDROM was mailed on 24<sup>th</sup> October, with the new software and 3 databases. PMERGE was originally written to function as a Taxonomic Editor, such that inconsistencies between existing and incoming taxa were meant to have been resolved interactively while PMERGE was running. Postponing these changes until later has made the software more complex.

### 1. ESFEDS database

1.1 The starting point was an updated version of the FE database. The expansion of 'non' authors in authority strings was included. This means that authority strings of the form

Fictitius vulgaris Sm. non L

were replaced with

Fictitius vulgaris Sm. and Fictitius vulgaris L. (taxon added if not already present).

In other words, all homonyms were normalised. 'Nec' and 'vix' and 'sensu' were also removed, but not sens.str. and sens.lat. The name types 'A' and 'S' were restored, 'K' changed to 'P' and the synonymies and the distribution data reinserted. The geography was changed slightly (see below). Various minor corrections to ESFEDS were also included. The database is referred to as FE2 (for version 2).

A number of minor corrections were made to the Macaronesia database. The records for Cape Verde were left in place, but a program was written to select only those taxa which had distribution in the other four areas (Azores, Canaries, Madeira, Selvagens) and which were not recorded from Cape Verde alone. The list NOCV specifies these taxa. Records of distribution from Cape Verde were copied over into the merged database but then deleted. The new version of the MACAR database has been included.

### 1.2 No news of Medclist V.2

1.3 The bibliography in FE is not standardised to either BPH or TL2 but follows the abbreviations used in Flora Europaea. The correction work was scheduled in Reading for December 2000 but I believe was not done. This standardisation is urgent and will cause many difficulties if put off any longer. Once the bibliography has been standardised it will then be necessary to check for duplicated publications and authors against the data from the centres. Many of the books are not in TL2 (which stops at 1940) but abbreviations for journals should follow BPH.

1.4 In order to merge Macaronesia with FE I found it necessary to alter the GAZETTEER data in both databases. The PMERGE software assumes that the GAZETTEER data is identical in both datasets, which it wasn't. The GAZETTEER included in FE2 is, according to me, the correct one, but will certainly not be the same as any GAZETTEER at Reading, and this will need to be checked carefully. Please also refer to my email message of 19/9/2001 which refers to other errors in GAZETTEER and ZONES (which are corrected in my version).

## 2. CONFLICTS

- 2.1 The file CONFLICTS already existed in E+M PANDORA before this discussion began, and had been proposed about 6 months before. I have retained it, for several reasons. One is that many of the conflicts involve synonyms and are therefore not part of the data for a taxon in PTAXON, but a relation between two such taxa, which needs a separate table. Another reason is that Anton in Berlin contacted me several months ago, asking how his editing software could find out which taxa needed editing. In reply, I explained about CONFLICTS and his response was that that was what he needed.
- 2.2 The PTAXON and CONFLICTS files will go together and can be included in standard copying procedures. I do not think that this will complicate matters very much.
- 2.3 No, I don't agree.
- 2.4 Requests from users for 'notes' in databases are generally a sign that the database needs to be redesigned with extra features. At best, notes fields should be used only temporarily until the noted data can be dealt with, preferably promptly. As explained above, a taxonomic editor can function much more easily with structured data than with free form notes. CONFLICTS has a viewing window with indexes on the key field where conflicts on both accepted names and synonyms can be searched for and viewed.
- 2.5 Conflicts are also visible in the PTAXON window, where an option can be used to browse the CONFLICTS, similar to the way that homonyms can be browsed in the HOMONYM field. Both accepted names and synonyms are included, with accepted names first. To detect which taxa have conflicts, do

```
SELECT PTAXON WITH CONFLICTS # ''
SAVELIST listname
```

And browse PTAXON with this list. Also, a quick report on CONFLICTS can be made with

```
LIST CONFLICTS ESTATUS NS1 NAME_TYPE NS3 BY NS1
```

A sample of such a report is supplied (CONF.REP). The CONFLICTS window has been added to the Data Entry menu, and is also available on the F10 menu within PTAXON and by F2 on the CONFLICTS field at the end of the PTAXON screens (deliberately put next to NOTES).

3. Distribution data. DIST records based on synonymous names are retained after substitution with the current accepted name. A new data field AS\_TAXON shows the synonyms used previously. Since synonymy can change repeatedly during editing, it seemed sensible to remember this in case these records are questioned. PMERGE gives a warning if there is more than one accepted name for a synonym, and only the first accepted name is used.
- 3.4 On the contrary, CONFLICTS data will be visible in PTAXON for both taxa in this example, so that editing can be approached by either name.

3.5 See the document CONF.REP. Yes, reports can mix data from any files

#### 4. Merging strategy

- 4.1 All existing and incoming names are left with the status they already had, unless you change it. The only exception is for incoming names which are in conflict with existing names, and in that case, their incoming status is saved in CONFLICTS and their existing status is left alone.
- 4.2 No
- 4.3 Names with status 'K' are changed to 'P' in order to draw attention to the fact that some decision needs to be made about them.

#### 5.

5.3 When a taxon such as a genus is detected with the wrong status, then you can get messages asking whether members of this taxon should have their status reversed. If you say NO, then no taxon is given 'X' status, and incoming members of this taxon may be found to be in conflict if their names are not new, but otherwise will pass unmarked. This may not matter, especially if the genus has to be reviewed anyway. If you say YES, then status of member taxa is changed to 'X' and when a member taxon appears in the input later on, you will be asked if you want to change the status. If you don't change the status, it will remain as 'X' in the output, which might be a good idea. It also goes into CONFLICTS.

#### 6

##### 6.1

There are many homonyms in the databases already. I can write a report to show them if you wish.

#### 7.

7.1 See 5.3

8. This example would be detected by the existing utility program for checking basionyms

9. A new situation occurred during testing which I had not thought of before. If a taxon is in the existing database as a synonym (S) and it appears in the incoming data as an accepted name (A) and you have said NO to reversing this AND the A name has synonyms attached to it, then these synonyms cannot be left as they are (unless they are orthographic variants, SYNONYM\_TYPE = 'O') because they will be synonyms of synonyms. You now get a message

Synonymy of [accepted name] may need to be reversed

and these synonyms are put into CONFLICTS.