

The Merging Software – A Progress Report

On the 24th October 2001, Richard Pankhurst sent 3 PANDORA databases and a new version of the PANDORA software to the Secretariat. These databases were:

- FE2 (an updated version of the old ESFEDS database, including the expansion of 'non' authors in authority strings, e.g. *Fictitius vulgaris* Sm. non L. was replaced with *Fictitius vulgaris* Sm. and *Fictius vulgaris* L. (taxon added if not already present)
- Flora Macaronesia (1993, Edition 4)
- FE2 + Macaronesia merged together. The merge has resulted in an additional 3590 name records (1850 of which are accepted, and of these 186 are new accepted genera and 18 are new accepted families) and 23809 distribution records being added to FE2.

Included with the PANDORA software was the version of the Merge Software used to merge Macaronesia to FE2.

In relation to testing the Merge Software there are 3 separate issues to be addressed:

1. Is the data from the incoming database transferred correctly to the target database?
2. Is the Merge software reporting conflicts between names correctly?
3. Is the output from the database in a suitable format for use by Checklist Reviewers?

1. Is the data from the incoming database transferred correctly to the target database?

1.1 Merging a Family Block with FE2

A test merge was carried out with a Family Block dataset from one of the Data Input Centres. However this merge could not be run through to completion. The merge software needs to know if the information in the incoming database is already sourced or whether a source reference needs to be provided for the incoming data. In the case of the Family Block datasets all information is sourced to the Standard Flora from which the information was taken, so the former option was selected. However, when the merge software came to merge the PTAXON table (the table containing information on names), the source data (from the CITATIONS table) was not recognised by the software, so none of the PTAXON information was merged.

1.2 Merging the Macaronesia dataset with FE2

Because the information in the Macaronesia database is not sourced we were able to run a successful merge using Macaronesia as the incoming database and FE2 as the target database.

For this test merge we selected the genus *Bryophyllum*, as it is a small genus and has several species which are new to FE2. *Bryophyllum* also has the additional problem of being an accepted genus in Macaronesia, but a synonym in FE2 (see section 2.2 below).

There are no Bibliographic or Citation records in Macaronesia, so we cannot comment on whether the merging of these datatypes is being carried out correctly.

The merging of Ranks, Persons, Committee and Kewauthind were successful.

The merging of PTAXON creates various conflict messages, which are viewable by the operator of the software only, i.e. they are not available to the Checklist Reviewers. With each message the operator is asked to make a decision. It was previously agreed that the operator of the merging software will not make *any* taxonomic decisions, this is the role of the Checklist Reviewers. So in all cases the name statuses of taxa in the target database, in this case FE2, remain unchanged by the operator.

The PTAXON data is transferred correctly, but there are still areas to test. For example, in circumstances where a name is a synonym in FE2 [the target database], but an accepted name in Macaronesia [the incoming database]: we need to test if data associated with the name in Macaronesia (e.g. Type Citation or Basionym), which do not occur in FE2, are carried across to the synonym record in FE2 even though a conflict is generated.

1.3 Distribution information

Distribution information appears to have been appended correctly. If a distribution record is attached to an accepted name in Macaronesia, and this name occurs as a synonym in FE2, then the distribution record is added to the accepted name in FE2, and the 'AS_TAXON' field is filled in with the name of the synonym. For example, *Bryophyllum pinnatum* is accepted in Macaronesia and a synonym of *Kalanchoe pinnata* in FE2. *Bryophyllum pinnatum* has a distribution record for Madeira in Macaronesia. When the record for Madeira is merged the record in FE2 reads:

Taxon: *Kalanchoe pinnata*
As_Taxon: *Bryophyllum pinnatum*
Area: Madeira
Source: Flora Macaronesia

There was only one case in the test data set where a Distribution record was duplicated between FE2 and Macaronesia. *Kalanchoe pinnata*, accepted in FE2, has the DIST record Macaronesia. *Bryophyllum pinnatum*, accepted in Macaronesia, has the DIST record Macaronesia

Bryophyllum pinnatum occurs as a synonym of *Kalanchoe pinnata* in FE2. During the merging of Distribution a source record is added to the Distribution record for *Kalanchoe pinnata* occurring in Macaronesia so the DIST record for Macaronesia in FE2 now has two sources: Flora Europaea and Flora Macaronesia. However the As_Taxon field is left blank.

2. Is the Merge software reporting conflicts between names correctly?

2.1 An example of conflict reporting

Merging the small test data set for *Bryophyllum* resulted in the reporting of 6 conflicts in the merged database. The conflicts are recorded in a table called CONFLICTS. Each conflict has an individual record, so there are 6 records in the CONFLICTS table of the merged dataset. The record number of the conflict is also displayed in the 'Conflicts' field in the appropriate PTAXON record. In the Ptaxon record for *Bryophyllum* the number 1 appears in the 'Conflicts' field.

For Conflict 1 the information in the CONFLICTS table is:

Number: 1
Taxon: *Bryophyllum* Salisb.
Exist Type: S
Genus:
Source: Flora Macaronesia
Name Type: A
Syn Of:
In genus:

This means that *Bryophyllum* exists as a synonym in FE2, but as an accepted name in Flora Macaronesia. In the PTAXON record for *Bryophyllum* in the merged database the number 1 is displayed in the 'Conflicts' field.

A more complicated example is *Bryophyllum pinnatum* (Lam.) Oken. In the PTAXON window there are 4 conflict numbers displayed in the 'Conflicts' field: 3, 4, 5 & 6. In order to see what these numbers mean the user needs to go to the CONFLICTS table and look at records 3, 4, 5 & 6. Conflict number 4 also appears in the PTAXON window for *Bryophyllum calycinum* Salisb. and Conflict numbers 5 & 6 appear in the PTAXON window for *Kalanchoe pinnata* (Lam.) Pers.

Number: 3
Taxon: *Bryophyllum pinnatum* (Lam.) Oken
Exist Type: S
Genus: *Bryophyllum* Salisb.
Source: Flora Macaronesia
Name Type: A
Syn Of:
In genus:

Number: 4
Taxon: *Bryophyllum calycinum* Salisb.
Exist Type: S
Genus: *Bryophyllum* Salisb.
Source: Flora Macaronesia
Name Type: S
Syn Of: *Bryophyllum pinnatum* (Lam.) Oken
In genus: *Bryophyllum* Salisb.

Number: 5

Taxon: *Kalanchoe pinnata* (Lam.) Pers.

Exist Type: A

Genus: *Kalanchoe* Adans.

Source: Flora Macaronesia

Name Type: S

Syn Of: *Bryophyllum pinnatum* (Lam.) Oken

In genus: *Bryophyllum* Salisb.

Number: 6

Taxon: *Kalanchoe pinnata* (Lam.) Pers.

Exist Type: A

Genus: *Kalanchoe* Adans.

Source: Flora Macaronesia

Name Type: S

Syn Of: *Bryophyllum pinnatum* (Lam.) Oken

In genus: *Bryophyllum* Salisb.

The situation here is that *Bryophyllum calycinum* Salisb. and *Kalanchoe pinnata* (Lam.) Pers. are both synonyms of *Bryophyllum pinnatum* (Lam.) Oken in Macaronesia. All three of these names occurred in FE2 prior to merging so a conflict message is generated for each name. The problem here is knowing what status and relationship each of these names has in FE2. The names have to be looked up in FE2 individually. *Bryophyllum calycinum* Salisb. and *Bryophyllum pinnatum* (Lam.) Oken are both synonyms of *Kalanchoe pinnata*.

2.2 Synonym reversal and unreported conflicts

Synonym reversal occurs during merging when a genus in the *incoming* database has a different status to that which it has in the *target* database, and as a result, the status of all species of the incoming genus are in conflict with the target database.

e.g. *Bryophyllum* is a synonym of *Kalanchoe* in FE2 [target db], but *Bryophyllum* is an accepted name in Flora Macaronesia [incoming db], thus there is a conflict between **all** the accepted taxa of *Bryophyllum* in the incoming database and the target database.

However, in the example merged database sent by RICHARD, conflicts of this type are only recognised if incoming names are already in the target database. In Flora Macaronesia 7 species of *Bryophyllum* are listed: two of these (*B. pinnatum* and *B. calycinum*) are synonyms in FE2, the other 5 are not in FE2.

e.g. 1: when the incoming name is in the target database

- In Flora Macaronesia *Bryophyllum pinnatum* (Lam.) Oken is **accepted**
- In FE2 *Bryophyllum pinnatum* (Lam.) Oken is a **synonym** of *Kalanchoe pinnata* (Lam.) Pers.
- In the merged database *Bryophyllum pinnatum* (Lam.) Oken is a **synonym** of *Kalanchoe pinnata* (Lam.) Pers., and two conflict messages are reported,

indicating that in Flora Macaronesia *B. pinnatum* is an accepted name, and *K. pinnata* is a synonym of *B. pinnatum*.

In this situation the conflict information works.

e.g. 2: when the incoming name is not in the target database

- In Flora Macaronesia *Bryophyllum proliferum* Bowie & Curtis is **accepted**
- *Bryophyllum proliferum* Bowie & Curtis is not in FE2
- In the merged database *Bryophyllum proliferum* Bowie & Curtis is **accepted**, and no conflicts are reported, even though the genus *Bryophyllum* is treated as a **synonym** in this database.

In this situation, the data in PTAXON in the incoming database are carried directly across to the target database, despite the conflict at the generic level being identified by the MERGE software.

A similar situation is also found for all other accepted species of *Bryophyllum* in Flora Macaronesia which are not already listed in the target database (FE2) at the time of merging, i.e.

Bryophyllum daigremontianum (Raym.-Hamet & Perr.) A. Berger

Bryophyllum delagoense (Eckl. & Zeyh.) Schinz.

Bryophyllum fedtschenkoi (Raym.-Hamet & Perr.) Lauz.-March.

Bryophyllum tubiflorum Harv.

3. Is the output from the database in a suitable format for use by Checklist Reviewers?

After PTAXON has been merged, conflicts can be viewed in two ways:

1. A Conflicts Report can be generated directly from the CONFLICTS table (see Appendix 1 and 3.1. below)
2. Descriptions of individual conflicts can be viewed in the CONFLICTS table (see 3.2)

3.1 *The Conflicts Report*

An example of the Conflicts Report generated is shown in Appendix 1. Richard has suggested that this is suitable, along with a printout from the database, to send to the Checklist Evaluators. However, the raw data as it appears in the Conflicts report is not user-friendly, and is tabulated in a more readable way below..

Conflict No.	Name Status in merged database	Name	Name Status in Macaronesia
1	S	Bryophyllum Salisb.	A
2	A	Kalanchoe Adans.	P
3	S	Bryophyllum pinnatum (Lam.)Oken	A
4	S	Bryophyllum calycinum Salisb.	S Bryophyllum pinnatum (Lam.)Oken
5	A	Kalanchoe pinnata (Lam.)Pers.	S Bryophyllum pinnatum (Lam.)Oken
6	A	Kalanchoe pinnata (Lam.) Pers.	S Bryophyllum pinnatum (Lam.)Oken

3.2 *The Conflicts Table*

The information within the CONFLICTS table in PANDORA is directly linked to the PTAXON window, although it is not visible in PTAXON. Richard has informed us that separate files within PANDORA can be combined in one report, so we should be able to create a report for Checklist Reviewers which shows both the information in PTAXON (names) and the information in CONFLICTS together. [we have not yet established with RICHARD exactly how to do this]

If the information contained within the CONFLICTS Table is to be put in a report used by the Checklist Reviewers, it MUST be clear and user friendly. The information displayed in the CONFLICTS table is informative, but some more user friendly field names are suggested below (together with a description of each field).

Current field name	Contents of field	Change field name to:
NUMBER	Gives number of conflict message	CONFLICT NUMBER
TAXON	Gives taxon number and full name plus authority of the species for which the conflict message was generated	✓
EXIST TYPE	Gives name status (e.g. S or A) of the name in the existing <i>merged</i> database (ultimately this will be the Euro+Med database)	NAME STATUS IN E+M DB
GENUS	Gives the name and authority of the genus of the conflicting name [if the conflicting name <i>is</i> a genus, this field is left blank]	✓
SOURCE	Gives the bibliography reference number and the title of the source of the conflicting data	SOURCE OF CONFLICT
NAME TYPE	Gives the status (e.g. S or A), in the <i>incoming</i> database, of the name which is causing the conflict	STATUS IN INCOMING DB
SYN OF	Where applicable, gives the name which the conflicting taxon is a synonym of in the <i>incoming</i> database	[SYN OF IN INCOMING DB]
IN GENUS	Gives the name and authority of the genus of the above name	✓

3.3 Synonyms of synonyms

In the case of synonyms of synonyms, it would also be helpful to have additional fields in the CONFLICTS Table, so that the taxonomic conflict can be clearly identified. For example:

In the Flora Macaronesia (incoming) database:

Rumex acetosella subsp. *angiocarpus* (Murb.) Murb. **Accepted name**

Acetosella angiocarpa (Murb)Á.Löve & D.Löve **Synonym** (of *Rumex acetosella* subsp. *angiocarpus*)

In the FE2 database:

Rumex acetosella subsp. *multifidus* (L.) Arcang. **Accepted name**

Rumex acetosella subsp. *angiocarpus* (Murb.) Murb. **Synonym** (of *Rumex acetosella* subsp. *multifidus*)

Acetosella angiocarpa (Murb)Á.Löve & D.Löve does not have an entry in the FE2 (target) database.

In the Merged database:

Rumex acetosella subsp. *multifidus* (L.) Arcang. **Accepted name**

Rumex acetosella subsp. *angiocarpus* (Murb.) Murb. **Synonym** (of *Rumex acetosella* subsp. *multifidus*)

Acetosella angiocarpa (Murb)Á.Löve & D.Löve **Synonym** (of *Rumex acetosella* subsp. *angiocarpus*)

How does the CONFLICTS Table indicate the problems associated with the incoming name, *Acetosella angiocarpa*, which is a synonym of a name that is itself a synonym in the target database?

One conflict message is reported in the 'Conflicts' field of the PTAXON window for *Acetosella angiocarpa*. This corresponds to the following record in the CONFLICTS Table:

NUMBER 387

TAXON *Acetosella angiocarpa* (Murb.) Á.Löve & D.Löve

EXIST TYPE S

GENUS *Acetosella* (Meisn.) Fourr.

SOURCE Flora of Macaronesia

NAME TYPE S

SYN OF *Rumex acetosella* subsp. *angiocarpus* (Murb.)Murb.

IN GENUS *Rumex* L.

The conflict message for *Acetosella angiocarpa* [387] does not indicate that it is a synonym of a synonym.

The conflict message would be clearer for reporting to Checklist Reviewers if two additional fields were added in the CONFLICTS table. These would indicate (where relevant) the name that the TAXON is a synonym of and the status of the name that the TAXON is a synonym of in the Target database. These two additional fields could be called "SYN_OF_IN_E+M_DB" and "STATUS_IN_E+M-DB". With the old field names changed to those suggested in the table in section 3.2, Conflict 387 would thus read:

CONFLICT NUMBER 387

TAXON *Acetosella angiocarpa* (Murb.) Á.Löve & D.Löve

NAME STATUS IN E+M DB S

GENUS *Acetosella* (Meisn.) Fourr.

SYN OF IN E+M DB *Rumex acetosella* subsp. *angiocarpus*
(Murb.)Murb.

STATUS IN E+M DB S

SOURCE OF CONFLICT Flora of Macaronesia

NAME TYPE S

SYN OF IN INCOMING DB *Rumex acetosella* subsp. *angiocarpus*
(Murb.)Murb.

IN GENUS *Rumex* L.

In the merged database *Rumex acetosella* subsp. *angiocarpus* (Murb.) Murb. is treated as a **synonym** of *Rumex acetosella* subsp. *multifidus* (L.) Arcang. (i.e. exactly how it is treated in FE2), and two conflicts messages are generated. One of these is the same message that is described above [387], relating to *Acetosella angiocarpa*, the second [345] is as follows:

NUMBER 345

TAXON *Rumex acetosella* subsp. *angiocarpus* (Murb.)Murb.

EXIST TYPE S

GENUS *Rumex* L.
SOURCE Flora of Macaronesia
NAME TYPE A
SYN OF
IN GENUS

This message would also be clearer with the new field names and the two additional fields added.

CONFLICT NUMBER 345
TAXON *Rumex acetosella* subsp. *angiocarpus* (Murb.)Murb.
NAME STATUS IN E+M DB S
GENUS *Rumex* L.
SYN OF IN E+M DB *Rumex acetosella* subsp. *multifidus* (L.)
Arcang
STATUS IN E+M DB A
SOURCE OF CONFLICT Flora of Macaronesia
NAME TYPE IN INCOMING DB A
SYN OF IN INCOMING DB
IN GENUS

The new field names and additional fields would also be helpful in interpreting the conflicts described in section 2.1 above.

Judith Heald & Liz Caddick (23rd November 2001)

Appendix 1

The Conflicts data as it appears in the output file from Pandora Key to field names:

Key – Record number in the Conflicts Table

TN – The name of the Taxon that has a Conflict

Estatus – The status of the name in FE2

Genus – The Genus the name belongs to

XNTCIT – The Name of the Source from which the conflict arose

Name Type – The name type in the incoming data

NS3 – The Accepted Name in Flora Macaronesia if the incoming name is a synonym

```

Key.....
TN.....
... ESTATUS GENUS.....
XNTCIT.....
... NAME TYPE
NS3.....
1          Bryophyllum Salisb.
S          Flora Macaronesia
A
2          Kalanchoe Adans.
A          Flora Macaronesia
P
3          Bryophyllum pinnatum (Lam.) Oken
S          18107      Flora Macaronesia
A
4          Bryophyllum calycinum Salisb.
S          18107      Flora Macaronesia
S          Bryophyllum pinnatum (Lam.) Oken
5          Kalanchoe pinnata (Lam.) Pers.
A          18095      Flora Macaronesia
S          Bryophyllum pinnatum (Lam.) Oken
6          Kalanchoe pinnata (Lam.) Pers.
A          18095      Flora Macaronesia
S          Bryophyllum pinnatum (Lam.) Oken

```

6 Rows Processed