Formatting and submitting datasets for the National Moth Recording Scheme (Revised May 2012)

Introduction

This document is intended to provide the County Moth Recorder (CMR) with informed choices and best practice solutions for managing moth datasets, in order to maximise compatibility with the National Moth Recording Scheme (NMRS) and save data validation and processing time.

This document replaces all previous versions sent to CMRs since 2007.

Although this document describes guidance and recommendations on formatting datasets, by not adopting some of the guidelines, for some datasets, our ability to process datasets efficiently may be significantly impacted. This in turn considerably delays the publication of refreshed data on the NBN Gateway.

Change to what and how much data to submit to the NMRS

The following only affects datasets in excess of about 200,000 records for a given VC. However, the same process can be applied to datasets with any number of records at the CMR's discretion.

Since the beginning of the scheme, CMR's were requested to submit complete datasets every time in order to capture all new, amended and deleted data since the previous refresh. This process has now become unsustainable by virtue of the amount of data being submitted and the amount of data that can physically import in records in a given time.

CMR's are generally aware of any recent data added to, and changes made to their datasets. New data, for example the previous year's end-of-season data entry or additional historical data, are relatively easy to capture and forward. These data are easily appended to the existing VC's survey.

Amended and deleted data are more difficult and there is no ideal way to handle these data without dealing with complicated synchronisation processes. *As the NMRS deals with data in all manner of formats and quality, individual 'sync' agreements regretfully cannot be entered into.*

CMR's are respectfully requested to consider the data that needs to be refreshed each year when a submission is made. A full refresh of VC data may not be needed every year and CMR's should be in a position to decide what needs to be forwarded to maintain the quality of their datasets in the NMRS.

How do I deal with amended and deleted data?

As previously mentioned, handling amended and deleted data poses processing difficulties, depending on the number of amendments and deletions performed on a given dataset since the previous NMRS submission. If a relatively small amount of data are amended and deleted, these same changes can be made by sending an email to the NMRS Data Manager requesting amendment or deletion of specific records. Amendments and deletion involving larger numbers of records will have to be captured by virtue of a full dataset refresh at the discretion of the CMR.

When is the data submission deadline?

From the publication of this document, there will be only one submission deadline per year, 31 March. This date should enable CMR's to capture, verify and submit the previous year's data in a reasonable timeframe. An NBN Gateway upload will be performed as-and-when all data received before 31 March, are imported into the NMRS.

Any data received after 31 March each year will be dealt with in the normal manner and imported into the NMRS; however, these data will not be guaranteed to make an upload to the NBN Gateway until the following year.

Data fields required

These are the fields or columns in your exported data. The headings described here are taken from the *MapMate Import Specification*, the most popular software choice of moth recorders and used by the overwhelming majority of CMRs; however, the same headings are relevant whatever the choice of electronic storage by the moth recorder.

All datasets, irrespective of their source and format, received by the NMRS are initially imported into MapMate to provide a common focal point for data validation and feedback as necessary to CMRs. Once validated, data are exported from MapMate in the correct format for the NMRS database.

Taxon	Site	Gridref	VC	Recorder	Determiner	Date	Quantity	Method	Sex	Stage	Status	Comment

MapMate Users who 'sync' data to the NMRS: If at all possible, please ensure no carriage-return/line-feed or Tab characters are inserted into any field. This can happen especially during the MapMate import process when Excel worksheets are exported to tab-delimited text files prior to importing into MapMate. Excel does have a habit of exporting blank rows and columns!

The best way of checking for blank rows and columns is to press *Ctrl-End* in your worksheet(s) to check what Excel believes to be the bottom-right-hand cell of your data i.e. the last record's last column e.g. commonly the Comment field. If the cell highlight isn't either on the last row or the expected right-hand column, you need to highlight and *Shift-Delete* all rows below your last record and/or all columns to the right of your last headed column.

Always check thoroughly all rows and columns to be deleted don't hold any valid data!

The following SQL WHERE clause when inserted into a MapMate-supplied "Browse..." SQL report will check for carriage-return/line-feed characters:

WHERE (([Sites\Default].OSGridRef) Is Not Null) AND (([Sites\Default].Name Like "*" & CHR(10) & "*") OR ([Sites\Default].Name Like "*" & CHR(13) & "*") OR (Recorders.Name Like "*" & CHR(10) & "*") OR (Recorders.Name Like "*" & CHR(13) & "*") OR (Methods.Method Like "*" & CHR(10) & "*") OR (Methods.Method Like "*" & CHR(13) & "*") OR (Records.Comment Like "*" & CHR(10) & "*") OR (Records.Comment Like "*" & CHR(13) & "*"))

The following SQL WHERE clause when inserted into a MapMate-supplied "Browse..." SQL report will check for Tab characters:

WHERE (([Sites\Default].OSGridRef) Is Not Null) AND (([Sites\Default].Name Like "*" & CHR(9) & "*") OR (Recorders.Name Like "*" & CHR(9) & "*") OR (Methods.Method Like "*" & CHR(9) & "*") OR (Records.Comment Like "*" & CHR(9) & "*"))

Code / Taxon / Vernacular

- *Code* - The Bradley number is useful as a cross-referencing tool when used in conjunction with *Taxon* and/or *Vernacular*.

- *Taxon* - Please use the names used in Bradley J.D., 2000. Checklist of Lepidoptera recorded from the British Isles. Second edition (revised) and ensure the spelling is correct.

- *Vernacular* - The English name is useful to further assist with data confirmation along with *Code* and *Taxon* to help eliminate nomenclature confusion for example, *Lomographa bimaculata* (Fabr.) **White-pinion Spotted** and *Cosmia diffinis* (L.) **White-spotted Pinion** and where species with multiple sub-species are allocated different English names i.e. *Eupithecia denotata denotata* (Hb.) **Campanula Pug** and *Eupithecia denotata jasioneata* Crewe **Jasione Pug**.

Site / Gridref

The NMRS database supports four spatial reference systems (in order of preference):

- OSGB (Ordnance Survey Great Britain)
- OSNI (Ordnance Survey Northern Ireland)
- UTM (Universal Transverse Mercator)
- Latitude and longitude

Best practice is to ensure all spatial references have unique site names to avoid confusion, although it is recognised in practice this isn't always easy to achieve.

OSGB and OSNI references

Please use at least a six-figure (100m) grid reference precision for records as this detail is vital for many recording and conservation activities; eight-figure and 10-figure grid references are most welcome where possible; four-figure (1km) or tetrad (2km) grid references for larger sites are acceptable.

Two-figure (10km) grid references are also acceptable, especially for vague sites where site confidentiality is required by landowners.

For species of conservation interest, more precise eight-figure grid references (10m) are useful for the exact locations of the sighting or any breeding colony.

Irrespective of grid reference resolution recorded, by default the highest resolution the NMRS will display records via online distribution maps is 2km.

Regretfully, records submitted with no spatial reference are not included in the NMRS database.

Please do not set up site names called "Unknown" unless you have to, it is much better to have at least a vague site name for each grid reference; the main element of a site name should at least be recognisable from an Ordnance Survey map, preferably the 1:25,000 scale series.

Important: *If a site is confidential, please use an ambiguous site name so not to identify that site. Please also do not use a full postal address as a site name. When matched with a recorder's name this constitutes personal data which is covered by the Data Protection Act 1984 (Revised 1998) and requires special considerations under that Act. Best endeavours are maintained to ensure personal data doesn't get into the public domain.*

Recorder

The best formats for names are:

- R.B.Powell
- Jackie Lee

The following formats are also satisfactory:

- Robert B.Powell
- Mr J.C.Jones
- Dr Jim Hopkirk

You can combine multiple recorders' names, separated with a semi-colon (;) - preferred, comma (,), forward-slash (/), backward-slash (\), plus sign (+) and either 'and' or '&' before the last name, using any of the above formats.

e.g. "R.B.Powell;Jackie Lee;Dr Jim Hopkirk"

Col / Dr / Miss / Mr / Mrs / Ms / Prof / Rev are titles that can be accommodated.

Please note: Recorders' names will not be used for any other purpose than correlating a species record to a recorder i.e. defining the ownership of that record for verification purposes or acknowledgement.

Determiner

The 'determiner' is the person who actually identified the species. Usually this will be the same as the recorder.

If there is a separate determiner, for example if a specimen has been given to someone else to identify, then their name must be added just as for the recorder. Some people also like to record that a record has been 'confirmed', i.e. R.B. Powell recorded and identified a species, but he got someone else to give a second opinion and confirm it. In this case R.B. Powell would go down as recorder and as determiner, but the name of the confirmer could be put in the *Comment* field to provide additional evidence of the species identification.

In all instances, there should only be a single determiner in the Determiner field, formatted as per Recorder names; however, there can be any number of other determiners listed in the Comments field as required.

Please use the determiner name "Undetermined" to represent unverified records.

Date

Date formats for reasons previously stated follow the *MapMate Import Specification*.

- Specific dates must be in dd/mm/yyyy format e.g. 12/07/2007.
- For any date range or vague-date, use dd/mm/yyyy-dd/mm/yyyy e.g. 03/04/2011-05/04/2011.
- For a given month, use the format 00/mm/yyyy or dd/mm/yyyy-dd/mm/yyyy e.g. 00/05/1997 or 01/05/1997-31/05/1997.
- If the date is a year only, use yyyy or 01/01/yyyy-31/12/yyyy e.g. 2012 or 01/01/2012-31/12/2012.
- If your date is a year range use yyyy->yyyy e.g. 2011->2012.

If exporting from MapMate, please ensure you export using the following SQL SELECT clause which will format the date field correctly in all cases:

IIf([Records].[Date] = [Records].[DateTo], Format([Records].[Date],'dd/mm/yyyy'),
(Format([Records].[Date],'dd/mm/yyyy') & '-' & Format([Records].[DateTo],'dd/mm/yyyy'))) AS [Date]

The following date formats are not suitable and should be avoided where possible for the reasons given above. The examples below show some of the common errors and isn't exhaustive:

- *dd-mmm-yyyy* e.g. 19-Mar-2012.

- *dd mmm yyyy* e.g. 19 Mar 2012.

- *mmm-yyyy* e.g. Mar-2012.

- Any date format with a 2-digit year *dd/mm/yy* e.g. 01/05/00 – *This ambiguous date could potentially mean either the year 1700, 1800, 1900 or 2000.*

Quantity

Abundance will be utilised for data analysis; therefore, please where possible provide this data.

If a species is present but no abundance is recorded please use "0" (zero) or "Present". Please do not just use "1" to show that the species is present unless it really was only "1" of that particular species recorded.

Zero-abundance records are not required.

Method

Use this field to note the method of recording for example, light trap, daytime observation, dusking etc. If a light trap, please specify the bulb type and where known for example, 125W MV.

Sex

If gender is recorded, please annotate accordingly and create separate records for each (if recorded together).

Stage

Adult, Larval, Pupal, Exuviae, Not recorded etc.

The NMRS will produce charts and atlases based on mature and immature stages. Knowledge of locations of immature stages is important as it equates to breeding locations, whereas adult moth sightings enable phenology or flight charts.

Status

Not required.

Comment

Any relevant comment can go here, record whether you have a photo, retained voucher specimen or seen the genitalia, or if anyone else saw and confirmed your identification - this is extremely useful supporting evidence, especially for rare or difficult species, and the more you can put in here the better!

IMPORTANT: Submitting Unverified Data

The NMRS has always respectfully requested and reasonably assumed all datasets forwarded are validated and verified. However, experienced has demonstrated this isn't the case and there may be instances when it is desirable to have unverified data in a verified dataset; however, these data must be recognisable as such by the NMRS. Flag a record as unverified in the following manner:

- Use the determiner name "Undetermined", and

- Add the phrase "Dubious record" anywhere in the Comment field.

This enables flagging of these data as "Failed/Pending Verification" so these data are not exported or used for reporting/ mapping.

In all instances, if there are any queries whatsoever, however big or small, please do not hesitate to contact me to discuss.

Les Hill
Data Manager, National Moth Recording Scheme
Butterfly Conservation
Manor Yard
East Lulworth
Wareham
Dorset
BH20 5QP

T: 01929 406008

E: lhill@butterfly-conservation.org

MapMate is a Registered Trade Mark of MapMate Ltd.