

Recorder 3

Transferring your data from *Recorder 3.4 to Recorder 2002*

Start Ball

Transferring your data from *Recorder 3.4* to *Recorder 2002*

Contents

1. INTRODUCTION	3
2. CHECKLIST	3
3. PREPARING FOR TRANSFER	4
3.1. RECORDER 3.4 AND RECORDER 2002 UPDATES	4
3.2. DATASET DETAILS (NECESSARY FOR THE XML TRANSFER ONLY, NOT FOR THE BULK TRANSFER)	4
3.3. SURVEYS	5
3.4. CLEANING YOUR DATA	6
3.5. USER ADDED CODES	8
3.6. DISTRICT/PARISH ENTRIES	9
3.7. CHECKING YOUR DATABASE	9
4. EXPORTING FROM <i>RECORDER 3.4</i>	13
4.1. IDENTIFYING THE USER WHO IS PERFORMING THE EXPORT.	13
4.2. USE NBN GUIs?	13
4.3. OPTIONS FOR DEALING WITH SUB-SITES	13
4.4. NBN XML EXPORT	14
4.5. BULK TRANSFER	14
4.6. EXPORTING SPECIES POPUPS AS <i>RECORDER 2002</i> RUCKSACK FILES	15
5. IMPORTING INTO <i>RECORDER 2002</i>	16
5.1. NBN XML EXPORT	16
5.2. BULK TRANSFER	16
6. ISSUES	17
6.1. HOW LONG WILL IT TAKE?	17
6.2. HOW MUCH SPACE IS REQUIRED?	17
6.3. KNOWN PROBLEMS	17
6.4. WHAT IS AND IS NOT TRANSFERRED?	17
7. CONCEPTS	18
7.1. DOCUMENTING DATASETS	18
7.2. NBN TRANSFER FORMAT	18
7.3. NBN GLOBALLY UNIQUE IDENTIFIERS	19
7.4. DATA "OWNERSHIP"	19

1. Introduction

The update from *Recorder 3.3* to version 3.4 allows data to be exported in the xml transfer format developed for the National Biodiversity Network (NBN). A bulk transfer mechanism is also included which allows a *Recorder* database to be transferred to *Recorder 2002* as a whole. This document is concerned primarily with the bulk transfer, as this is the advised method of upgrading to *Recorder 2002*. The differences between the two processes are described in section 5 of this document but it is very probable that the best method of upgrading for you is the bulk transfer.

Moving data from one application to another is always complex and this is especially true in dealing with data in *Recorder 3*. The software has been in use for a long time and its flexibility has allowed considerable variation in how it has been used, not to mention the customisations that have been made by some users.

By working through the checklist below, it should be possible to follow the process with the minimum of difficulties.

Subject areas raised in the checklist are detailed in sections three to five of this document.

The sixth section, "Issues" covers some of the things you need to know about, including limitations and problems and also provides some recommendations about how to tackle the whole process.

Finally, some of the conceptual differences between *Recorder 3* and *Recorder 2002* are outlined in section seven along with an outline of some important elements of the new system.

2. Checklist

The following checklist outlines the steps that should be taken in order to achieve a trouble free bulk transfer. **Make sure you have at least one good, up-to-date backup of your *Recorder 3* database before you start!**

- Perform the *Recorder 3.4* update and install the import add-in on to your copy of *Recorder 2002*. (section 3.1)
 - Think carefully how you are going to organise your Surveys and the Events and Records within them. (section 3.3)
 - When you have decided this, create the Surveys and associate your Records and Events with them. (section 3.3.2)
- TIP - It is probably best to assign your Events to a separate Survey. These will appear as Survey events without any Samples in the observation tree of *Recorder 2002*.
- Check that all Records and Events have been assigned to Surveys. (by running the Select statements in 3.3.2)
 - Run the data cleaning tool with all options selected (section 3.4)
 - Check that the cleaning process has not changed your data inappropriately. (section 3.4)
 - Sort out any Codes you may have added (section 3.5)
 - Assign NBN keys to any districts & parishes that you may have added (section 3.6)
 - Use the Check tool to find any errors. (3.7.7)
 - Fix any errors that the check tool has identified. You can use the broken link removal tool to speed up this process.
 - Keep running the Check tool and fixing any problems until you don't get any more errors reported. (section 3.7.7)

You may wish to use the clean data tool again, just to be safe and its probably time to make another backup!

- Use "NBN - Bulk move" to export your data. (section 4.) **This could take some hours for a large data set.**
- TIP- It is probable that you will not want to "use GUIs" as if you do use them, you will be unable to edit your imported data in *Recorder 2002*.

Make a backup copy of the nbndata.mdb file before you start any imports.

- Attempt an import into *Recorder 2002*. (section 5) **This could take some hours for a large data set.**
- If you get "Invalid items" reported, abort the import and fix the problem in *Recorder* then export again and have another go at import.
- Check that your data in *Recorder 2002* is recreated as you would like it. If not, use the editing tools within *Recorder 2002* to tidy up the details. One particular area that you may wish to check is where the information from SEX_STAGE and ABUNDANCE in *Recorder 3* has gone to. (section 3.5)

Some users have reported problems with their transfers due to their screen save and energy save options. The process may lock itself up in certain circumstances. It is therefore suggested that you disable these functions before commencing the export.

3. Preparing for transfer

3.1. Recorder 3.4 and Recorder 2002 updates

The tools necessary to export data from *Recorder 3* to *Recorder 2002* form part of an update of *Recorder 3.3* to *Recorder 3.4*. The final version of this update is available from the NBN web site.

To perform the update:

- Go to the NBN web site's updates screen and select the *Recorder 3* Update zipfile.
- Unzip the Rec34.zip file to a temporary directory somewhere on your hard disk. (For example, c:\rec34.)
- In Recorder 3.3, go to TCL (press [F5])
- Issue the command: RUN DOS C:\REC34\UPDATE (substituting the name of your directory for "C:\REC34".)

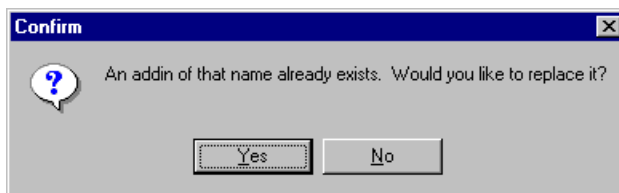
At the end of the process you will be logged out of *Recorder*. When you next use *Recorder*, it will be the latest version of *Recorder 3.4* and the transfer facilities described in this document will be available.

The tools necessary to do a bulk import into *Recorder 2002* are available in the same update folder.

- In *Recorder 2002* select Tools - Install - Add-in Module ...
- In the Add-in Configuration dialog, click the [Install] button and locate the file Recorder_3_Import.ocx in the unzipped *Recorder 3.4* update folder.
- Click the [Open] button to install the add-in.

You should see messages confirming that the add-in has been installed and the item "Recorder 3 Import" should appear in the list of installed add-ins.

If this add-in has ever been installed previously on the particular machine you will see a message:

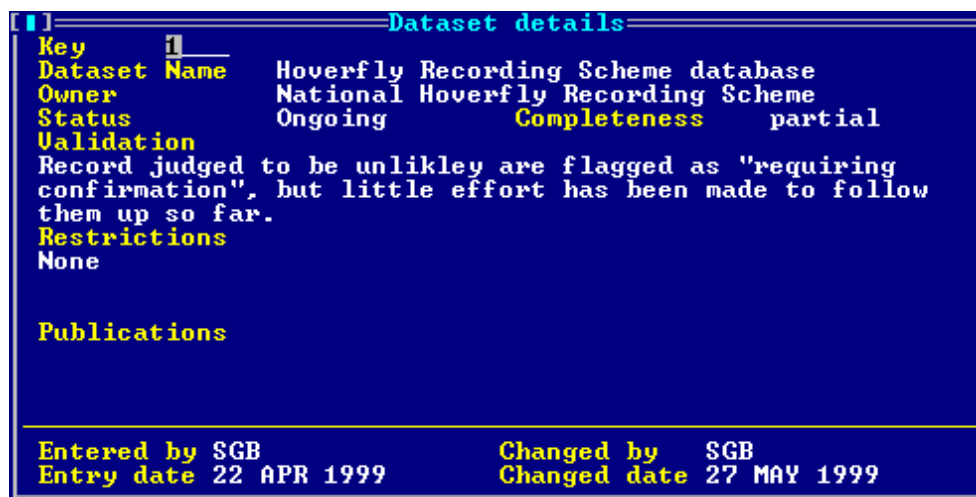


Click the "Yes" button to continue with the installation.

3.2. Dataset details (necessary for the xml transfer only, not for the bulk transfer)

The Dataset record provides meta-data about this particular database which is required in the XML transfer file. In general you will only need to set up a single record which will be used in all NBN exports.

- In *Recorder 3.4*, Select NBN - Dataset
- Fill in the Dataset window. Pressing <F1> at any of the prompts will display a help screen which gives details about the contents which are expected for each field.
- Press <F9> to save the data and <Esc> to close the window.



3.3. Surveys

Both the bulk transfer to *Recorder 2002* and the xml transfer require that all records are part of a Survey. Survey records provide information about the records such as when they were made, who organised for the surveying to be done and for what purpose. You need to decide what surveys you will need to set up, but here are a few guidelines:

- If records were collected as part of an organised survey, then the content of the survey record and the RECORDS that should be assigned to it will be obvious.
- If you have a collation of other people's records, then it is probably best to set up a survey for each person's data.
- If you are only dealing with your own data, then you may only need to set up a single survey. But if it is a very large dataset (see next point) it might be worth splitting it up (say, by time periods).
- Having a very large number of Survey events in a single survey will have performance implication in *Recorder 2002*. It will slow down the process of opening the observation hierarchy to look at your records. At the extreme, a single survey - with 81,000 Survey events - took approx. 20 minutes to open!.

Try to arrange your surveys in the most appropriate manner for your reporting needs. If however, you change your mind, it is now possible to move events and samples around between surveys within *Recorder 2002*.

3.3.1. Creating a Survey

- In *Recorder 3.4*, Select NBN - Survey
- Fill in the Survey window. Pressing <F1> at any of the prompts will display a help screen which gives details about the contents which are expected for each field.
- Press <F9> to save the data and <Esc> to close the window.

Dates: "Date started" and "Date completed" indicate the period over which the survey was operating and are for information only. They are not mandatory fields. However the "Earliest" and "Latest date" are used to indicate the period for which records are acceptable. If you assign RECORDS or EVENTS to the survey which are dated before "Earliest date" or after "Latest date" they will be rejected by *Recorder 2002*. At these two prompts, press the Options key [F2] to search for the earliest or latest dates amongst your records.

3.3.2. Assigning RECORDS and EVENTS to surveys.

- Select the records you want to assign to a particular survey. This could be done from the Filter menu by selecting records or by activating a saved select list (for example, in the case of RECORDS, saved by Reports - Record selection).
- Select NBN - Assign Survey
- You will see the following window:



- In "Survey" choose one of the survey entries you have created. The Option key, <F2>, lists the available surveys.
- In "Table to update" choose either RECORDS or EVENTS depending on the type of records you selected.
- Press <F9> to save the data and close the window. You will see a progress bar as the records are updated and a confirmation message will appear to show how many records were updated.

Both the Record and Events entry screens (*Add/Edit Data - Records or Events*) have new prompts added so that you can view or change the Survey to which individual records are assigned.

TIP:

Run the following select statements (*Filter - Select*) to find all the records or events you have not yet assigned to a Survey

```
SELECT RECORDS WITH SURVEY = ""
```

or

```
SELECT EVENTS WITH SURVEY = ""
```

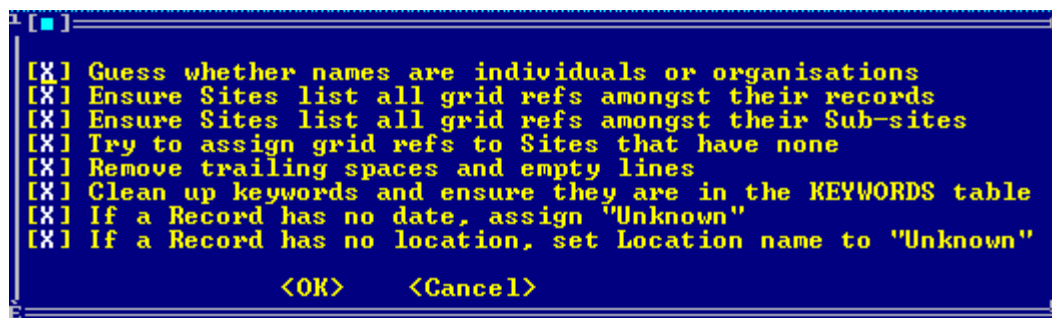
Run the following report (*Report - R/List*) to show you how many records you have assigned to each survey number (pay close attention to the punctuation in this LIST statement!):

```
LIST RECORDS BY SURVEY BREAK-ON SURVEY ""VL"" TOTAL COUNT DET-SUPP ID-SUPP
```

3.4. Cleaning your data

A tool is provided which checks for, and fixes, a variety of common problems in your data that would otherwise create difficulties during transfer.

Select NBN - Clean data to start the data cleaning tool. You will see the following screen:



You can choose whether or not to apply each of these eight options by “checking” or “un-checking” the boxes on the left. It is probable that you will want to have all options checked. Click the “<OK>” button or press [F9] to start the cleaning process.

You will be prompted to choose how much detail you would like reported to you. If you opt to produce a report, a file called “CleanUp.txt” will be created in the default output directory. I.e., if Recorder is directly on your c drive, you will find the report in c:/recorder/output.

A “Verbose” report includes a line for each item that needed action. The table, row ID and nature of the change that was made is reported. A “Summary” report only shows how many rows were updated in each table, but does not give any details of which rows or what changes were made.

Outlined below is a description of the changes performed by the clean tool. Also described are some methods for checking that the changes are appropriate.

3.4.1. Guess whether names are individuals or organisations

Recorder 2002 makes a distinction between Individuals and Organisations. They are stored differently and **an organisation cannot act as a recorder or determiner**. Records in the PERSONAL table in *Recorder 3.4* must be changed to state whether they refer to individuals or organisations by putting an entry in a new field, ORG, that has been added to the PERSONAL table for this purpose. *Recorder* will “guess” that it is not an organisation (and put “false” in the ORG field.) if the record contains a title or initials.

You can view the results by running a report (Reports - R/List):

```
LIST PERSONAL NAME ORG BY SORT_NAME
```

This will show, for each name in the table, either "Yes" (it is an organisation) or "No" in the "Org" column. Check and make any necessary changes via the Personal entry screen (Add/Edit Data - People) where the guess was wrong.

3.4.2. Ensure sites list all grid refs amongst their records

Recorder 2002 requires that the spatial reference given in a Survey event or Sample is the same as, or within one of the grid squares listed for the Location to which it is attached. This process checks the grid reference from each Record against the grid squares listed for the Site to which it is attached and, if the 1km square derived from the Record grid reference is not amongst those listed in the Site, adds it to the list.

3.4.3. Ensure sites list all grid refs amongst their Sub-sites

Recorder 2002 requires that a sub-site is within one of the grid squares listed for its parent location. The process checks the grid reference from each sub-site against those listed for its parent Site and, if the 1km square derived from the sub-site grid reference is not amongst those listed in the parent, adds it to the list.

3.4.4. Try to assign grid refs to Sites that have none

Sites in *Recorder* do not require a spatial reference. However, in *Recorder 2002* a spatial reference is required for each Location. (which is what a site is called in *Recorder 2002*). This process attempts to assign a grid reference to sites without one. The tool gets a grid reference from the Vice-county, County, District or Parish associated with the site.

NOTE: If a grid reference can be found from one of these sources, it will be assigned at 10km square precision.

3.4.5. Remove trailing spaces and empty lines

If odd blank lines or trailing spaces have crept into your data and are not removed, then they can cause referential integrity violations which are difficult to spot. The tool checks the main tables in *Recorder* for extraneous trailing spaces and empty lines and removes them.

3.4.6. Clean up keywords and ensure that they are in the KEYWORDS table

Keywords in RECORDS (for Substrate) and SITES are frequent causes of problems because;

- Keywords in the main tables may not be present in the KEYWORDS table.
- In the KEYWORDS table, keys are stored with a prefix to distinguish the keyword type (e.g. U/ANGLING is a keyword "ANGLING" of type "Use"). In the SITES table, the keyword should be stored as "ANGLING" without the prefix but is sometimes found with it.

This tool checks all keywords in the SITES and RECORDS tables and:

- Removes any prefix
- Checks that the resulting keyword exists in the KEYWORDS table and, if it does not, adds it.

Following this process you may want to review your keywords and add explanations for any that have been added.

3.4.7. If a Record has no date, assign "Unknown"

Recorder 3 does not allow a Record to be saved without a date, even if that date is "unknown". However it is not uncommon to find Records with no date information at all. This locates such records and sets the date to "Unknown".

3.4.8. If a Record has no Location, set Location name to "Unknown"

Recorder 3 does not allow a Record to be saved without some sort of locality information. However it is not uncommon to find Records with none. This locates such records and sets the vague location field to Unknown.

3.5. User added CODES

One of the more difficult aspects of moving data from *Recorder* to *Recorder 2002* is dealing with Sex/Stage/Abundance fields because these fields have been used in many different ways that mix together a number of types of information.

Recorder 2002 separates these concepts and stores them in different places, but disentangling them from *Recorder* is challenging and the methods that have been developed will undoubtedly not please everyone.

SEX_STAGE and ABUNDANCE are put into the Measurement page of the Taxon occurrence. The value from ABUNDANCE goes into the Data column and the SEX_STAGE into Qualifier with measurement type "Count". Because SEX_STAGE has been used in such a variety of ways, you may wish to review this data in *Recorder 2002*. RECORD_TYPE is particularly difficult because it embodies at least three different concepts:

- Sampling method (e.g. transect, quadrat, pitfall)
- Type of specimen (e.g. voucher retained, genitalia prep.)
- Record type in the sense of *Recorder 2002* (e.g. field observation, captured and checked, etc.)

Therefore, two new fields have been added to the Codes entry window which are only used for RECORD_TYPE codes. These are for Sampling method and Specimen type. These fields have been filled in for all the RECORD_TYPE codes supplied with *Recorder 3.3*, but you must fill them in for any RECORD_TYPE codes you have added.

Find RECORD_TYPE codes you have added by:

- Select Filters - Select from the menu and enter the following select statement

```
SELECT CODES WITH ENTRY_TYPE = "N" AND WITH SAMPLING_METHOD = " "
```

If this finds some records, you can edit them by:

- Select Filters - Save list
- Enter a name for the list when prompted and press F9 to save the list
- Go to the Code entry window (Add/Edit data - Codes)
- Press Ctrl-F10 and select "Browse selections" from the Browse Options popup
- Select "Get List" from the Browse selections popup
- Enter the name you gave for the list and press F9 to activate it
- The first record should load into the Code entry window.
- Fill in the Sample method prompt (i.e. this is required), and Specimen type if appropriate (i.e. this field is optional), and save the record by pressing F9. The next code should then display.
- You can also look through the list of codes by pressing Alt-F to go forward to the next record in the list or Alt-B to go back to the previous one.

Sample method	A brief description of the sampling method. Try and keep these generic. They will be used to define Samples (see section 4.3). An entry is REQUIRED for all RECORD_TYPE codes.
Specimen type	If there is an entry in this field, information will appear in the Specimen page of the taxon occurrence. Use this for codes which indicate that a specimen was retained: e.g. Voucher. If you leave this blank, then nothing will appear in the Specimen page of Taxon occurrences derived from records with this type.

Please note:

- Records with no entry in the "Type of record" will use the default Sampling method "Field observation"

- If a record has more than one entry in the "Type of record", and those entries have different Sampling methods, then the Sampling method from the first one will be used to define the Sample.

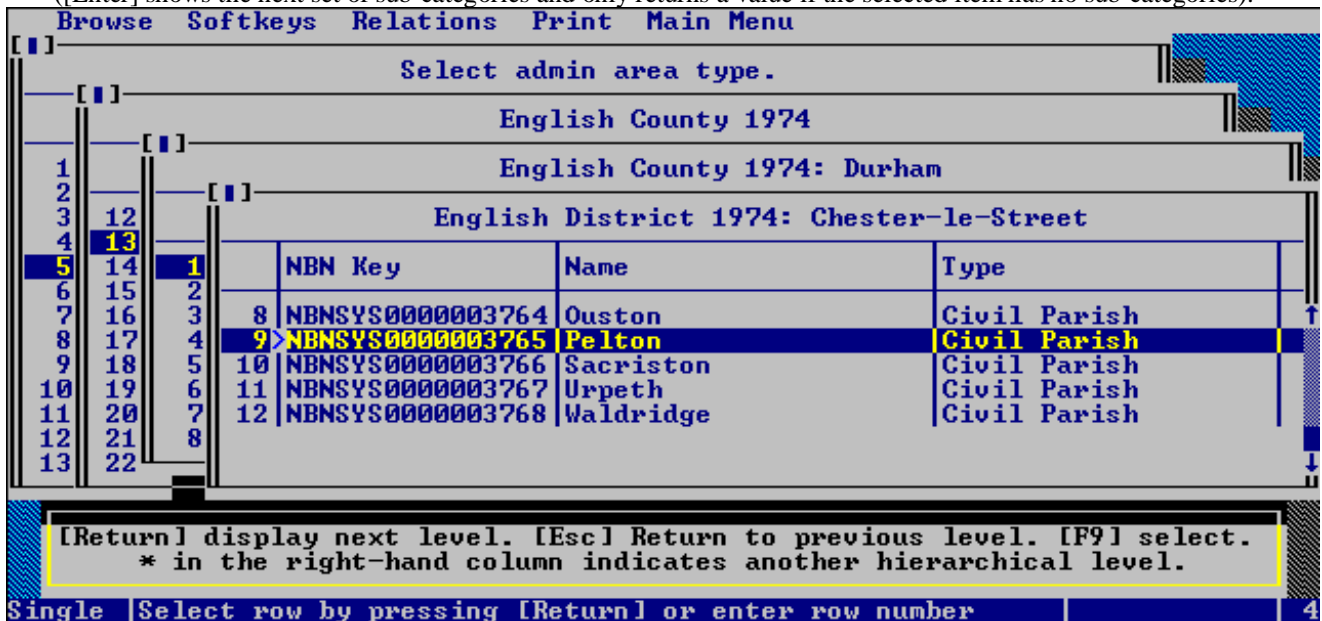
3.6. District/Parish entries

Only Post 1974 Counties and Watsonian Vice-counties were supplied as part of *Recorder 3.3*. However, users had the facility to enter and use districts and parishes themselves. The Administrative Areas dictionary in *Recorder 2002* now includes lists of districts and civil parishes for GB. However, these cannot be automatically matched with the district and parishes that user's have entered because each user may have used a different coding scheme.

If you want any district and parish entries you have used to be transferred to *Recorder 2002*, it is necessary to insert the appropriate NBN key into your entry via the District & Parish entry window (Add/Edit data - Parishes):

A new prompt "NBN Admin area" has been added. You can find the appropriate area by:

- Typing a name and pressing [Enter] to lookup the name in the indexes.
- Press the Options key [F2] to see a lists from the NBN Administrative Areas dictionary. This works just like selecting a habitat. In the first list you select the type of administrative area. You then see a list of the categories within that type, which may in turn have sub-categories. To select a particular item, highlight it and press [F9] ([Enter] shows the next set of sub-categories and only returns a value if the selected item has no sub-categories).



This screen-shot shows the selection of a parish in Durham. The "English County 1974" category was selected and then the County "Durham" from this list. Then "Chester-le-Street" was selected from the list of Districts in Durham, and finally the parish can be selected.

3.7. Checking your database

A tool is provided which checks for problems in your that would result in your records either not being exported or being rejected by *Recorder 2002*. This facility checks your database for a variety of problems:

- Referential integrity problems
- Data required by *Recorder 2002* is missing
- Site GUI clashes
- Dates that would fail *Recorder 2002* validation
- Grid references would fail *Recorder 2002* validation
- Organisations used in places where *Recorder 2002* does not allow it

(These errors are explained more fully in sections 3.7.4 to 3.7.9 of this document.)

A file called nbncheck.txt is then produced in the default output directory that can be used to correct these errors. I.e., if Recorder is directly on your c drive, you will find the report in c:/recorder/output. Use explorer to reach it.

3.7.1. Running the check

- Select NBN - Check from the menu.
- Confirm that you want to continue.
- Choose the options you are going to use for export.

These options can affect the outcome of some of the tests which the check process performs, so it is important to run a check using the same options you intend to use for export. These options are:

- "Treat sub-sites as samples" and
- "Use GUIs" and are described in sections 4.1 and 4.2 of this document.

You will see a progress display in the bottom left-hand corner of the screen which shows when each table is being worked on, a "% done" and the number of errors that have been found so far. You can interrupt the check by pressing the [Escape] key. If you do interrupt, any errors that have been found so far are saved and can be examined. At the end of the check you may see that the error count is reducing. This is because multiple errors resulting from a single cause are being removed.

At the end of the process you will see a message reporting how many errors were found. This process may take some time for a large database. Here are some example timings (on a Pentium III 500MHz under NT 4.0):

Number of rows in:					Time to Check (h:m:s)
RECORDS	SITES	PERSONAL	LITERATURE	EVENTS	
3,771	419	37	402	68	00:00:29
29,063	858	189	408	0	00:02:58
374,853	17,838	1,989	1,452	0	00:30:04

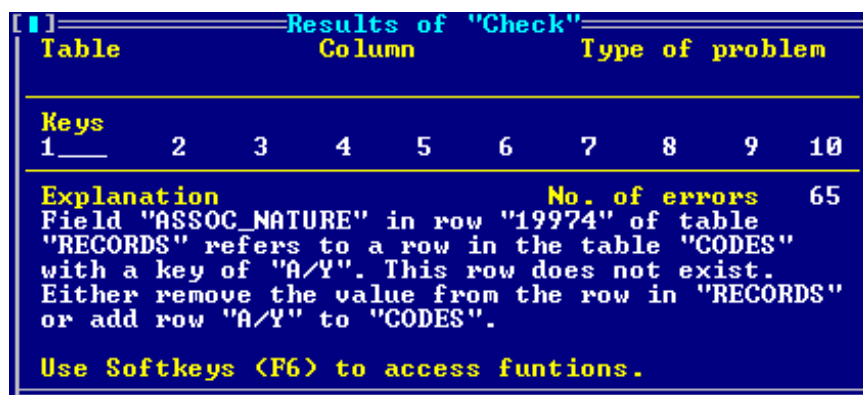
3.7.2. Examining errors

Errors in the nbncheck.txt file are in the following form:

Table: CODES
No errors found.

Table: RECORDS
RELATIONAL: The value "A/Y" in field "ASSOC_NATURE" of row 19974 is missing from table CODES
RELATIONAL: The value "W/RIVER_-_STREAM" in field "HYDROLOGY" of row 719 is missing from table KEYWORDS

Errors are also stored in the table NBN_CHECK_ERROR. They can be examined using the NBN - Show errors command from the menu.



```

Results of "Check"
Table          Column          Type of problem
-----
Keys
1_____  2    3    4    5    6    7    8    9    10
Explanation          No. of errors  65
Field "ASSOC_NATURE" in row "19974" of table
"RECORDS" refers to a row in the table "CODES"
with a key of "A/Y". This row does not exist.
Either remove the value from the row in "RECORDS"
or add row "A/Y" to "CODES".

Use Softkeys <F6> to access funtions.
  
```

Table	You can select the name of a <i>Recorder</i> table that you want to see errors for. The Options (F2) key will list the tables for which errors were found.
Column	This is only available when a Table is selected. You can select the name of a column from the selected Table. The Options key (F2) will list the columns in the selected table for which errors were found.
Type or problem	You can select one of the types of problem which the Check process detects. The Options key (F2) lists the types which are available. This can be combined with a Tables or Table and Column to refine the list of errors which is displayed.
Keys	This lists the keys of the errors that were found. These are just running numbers 1 to "No of errors". Select an error in this list and an explanation of that error is shown in the "Explanation" prompt.

Three Softkeys are available (F6)

Save selected list	The list of keys currently shown in the "Keys" prompt will be saved as a list. You can load such a list later on if you want to print a report about the records with errors.
View error	The record that is affected by the error that is currently selected in the "Keys" prompt will be displayed. You can then edit the record to correct the problem.
Remove broken links	This fixes referential integrity errors by removing the reference. It will not remove references in cases where the data is required by <i>Recorder 2002</i> so it will not fix all referential integrity problems.

3.7.3. Fixing errors

If a related record is missing, you have two options.

- Enter the missing record in the related table,
- Or remove the reference.

For example:

RELATIONAL: The value "A/Y" in field "ASSOC_NATURE" of row 19974 is missing from table CODES

You can add the record "A/Y" to the CODES table or edit the "Nature of association" in RECORD 19974 so that it uses an code that exists. You may need to do some investigation to find out what CODE "Y" means!

The "Fix broken links" softkey will resolve these problems automatically by applying the second option (remove the offending reference) unless the reference is required by *Recorder 2002*, in which case you will need to fix the problem by creating the missing record. Be careful about using this method as it will result in some loss of data.

If required data is missing, you need to enter something in the relevant field. This may require a bit of thought and investigation to come up with an appropriate value. Don't forget that dates can be "Unknown" and it is possible to set up records in the PERSONAL and LITERATURE table for "Unknown" in order to satisfy the requirement.

In the case of SITE GUI clashes, you need to change the ID of one of each pair of rows in which a clash is reported. Use Tools - Manage Recorder - Change ID - Sites from the menu.

The "Fix SITES/GRIDS grid ref mismatches" essentially runs the cleaning tools which fix problems arising because *Recorder 2002* is more strict about validation of Records within the Site to which they are attached than is *Recorder 3*.

3.7.4. Referential integrity

Recorder 3.4 and *Recorder 2002* both use a relational database model where lots of information is stored via links from one table to another. For instance, the species table contains a row of information about each species. Each row has a unique key. In a record, this key is used to reference back to the data stored in the species table. If such a key and therefore the data associated with it, does not exist in the SPECIES table then this is a *relational integrity violation*.

The check tests that all such keys have a matching row in the relevant table. If the key is missing, an error message is generated which reports the table and key which is missing along with what table and row references that missing key.

3.7.5. Required data is missing

Both *Recorder 3.4* and *Recorder 2002* require that certain fields contain valid entries before a record can be saved. However, the required fields are not identical and, in general, *Recorder 2002* is more demanding. Therefore there may be cases where *Recorder* has allowed the user to leave fields blank that will be required by *Recorder 2002*. The check tests that required fields are not blank. If such a field is empty, an error message is generated.

3.7.6. Site GUI clashes (only an issue if you answered yes to “use GUIs” when running the check)

Each record in *Recorder 2002* must have a unique 16 figure key called an NBN GUI. This can be derived for most records in *Recorder 3.4* using the copy number, the particular record’s key and by padding it out with zeros to give a 16 digit number.. For example, row “123” in the RECORDS table from the copy number 999 would have the NBN GUI:

```
0000099900000123
```

However, this does not work for sub-sites. Consider a sub-site like “123/2/4/7”.

- The “/” character is not allowed in NBN GUIs.
- If it was replaced by say “S” to make “123S2S4S7”, it would be too long (i.e. more than 8 characters).
- If the “/” characters were merely ignored (to make “00123247”), this potentially clashes with site number 123247 or with sub-site 1232/47.

To avoid these problems, a rather complex algorithm has been designed which attempts to generate an NBN GUI which is unique, legal and reproducible (i.e. a given sub-site ID always produces the same GUI). However, it is not infallible. One of the checks that is carried out is whether all sites produce unique GUIs. If this is not the case and you wish to generate NBN GUIs from your *Recorder 3.4* data, then you need to change the ID of one of the pair of sites whose IDs produce clashing GUIs.

Unfortunately, if the site with the problem ID was imported from someone else, changing the ID does not help.

- If one of the pair of problem sites was imported and one was locally produced then change the ID of the site that belongs to you.
- If both were imported then you may be dealing with the same site which you have imported more than once. If so, you need to merge the two site records (using Change ID) Otherwise, one possibility is that you have imported data from two or more sources with the same Centre number.

3.7.7. Date validation

A number of validation checks are performed to ensure that dates given in related records are consistent. For example, if you have given a start and end date for a survey, then this process checks that all events and records assigned to that survey have dates within this range.

3.7.8. Grid reference validation

A number of validation checks are performed to ensure that grid references given in related records are consistent. For example, the process checks that the grid reference which will be assigned to a sample is the same as, or within, the grid reference which will be assigned to the survey event to which the sample will belong.

3.7.9. Organisations used in places where *Recorder 2002* does not allow it

Recorder 2002 does not allow the determiner or recorder of a species (or biotope) to be an Organisation. Determiners and recorders must be individuals. This check identifies cases where the determiner is marked as an organisation in the PERSONAL table.

4. Exporting from *Recorder 3.4*

4.1. Identifying the user who is performing the export.

Recorder 2002 identifies the user who created each record it contains and the date on which it was created. In the case of a transfer from *Recorder 3.4*, the creator of the record is taken to be the person who is logged on when the export is performed. However, if this user name does not match an entry in the PERSONAL table, you will be prompted to state which of the entries in the PERSONAL table should be used as the creator of the records. <F2> lists all by name and initial and if there is not an appropriate entry, pressing <F6> will give access to the PERSONAL entry screen.

4.2. Use NBN GUIs?

GUIs are the unique, 16 figure identifiers for each record in *Recorder 2002*. They are required for the model. See also sections 4.4, 4.5 and 7.3

If you answer yes, the GUIs for the transferred database will be based on your record numbers in *Recorder 3.4*. **This means that you will not be able to edit them in *Recorder 2002*.**

If you say no, the GUIs will be created afresh upon import to *Recorder 2002*. **You will be able to edit them.**

If all the data on your copy of *Recorder 3* is your own and there is no other copy of it on another *Recorder 3* database, choose “no” and it will be a simple transfer over with all editing rights preserved and a fresh set of GUIs.

If either the data doesn't belong to you, or someone else has a copy of the data, you must be more careful. Because the new GUIs you create do not reference back to the original version, you could easily end up with multiple versions of the same data floating about and no way of knowing it. If this is the case, you must either choose “Yes” which preserves the link back to the original number and stops any duplication, or ensure that all but one copy of the data is destroyed.

4.3. Options for dealing with sub-sites

4.3.1. How Survey events and Samples are generated

Recorder does not have Survey events and Samples so they must be created during the export from information stored in RECORDS. The samples and survey events are created from the following rules:

A Survey event is a unique combination of

- location (this can be a defined site, a vague location description or a grid reference)
- date or range of dates
- person or people who did the recording

A Sample is a unique combination of sampling method and location within a Survey Event.

- The date of a sample must be the same as or within the range of dates given for the survey event.
- All people involved in a sample must be involved in the survey event as well

4.3.2. Effect of the "Treat Subsites as samples?" switch

Imagine the following scenario and let us assume for this example that all people and survey types are the same:

You have defined a site called Denby Wood. This site is subdivided into North and South. Both North and South sub-sites have records associated with them.. The switch gives you two options.

- If you answer yes to “treat subsites as samples”, you will have a single survey event for Denby Wood and two samples, one for the north sub-site and one for the south sub-site.
- If you answer no, you will have two survey events, one for north and one for south with a single survey in each.

If you have more than two levels of hierarchy in your locations, and answer yes, only the top level location will be used for the survey event. All sub-sites with records associated with them will have a separate sample.

4.4. NBN xml export

All records in the current select list are exported to NBN transfer format using NBN - NBN export. To perform an export:

- Select the records you want to export. This could be done from the Filter menu by selecting records or by activating a previously saved select list (for example, saved by Reports - Record selection).
- Select NBN - NBN export from the menus.
- You will see a window similar to the following:



- You are prompted to name the file to which output will be written and a default name of **Recorder.xml** on your default output directory is suggested. Enter any valid, DOS filename here.
- You are asked to identify the dataset. The options key (<F2>) lists all available dataset definitions (see section 3.2). If only one definition is available, it is automatically filled in.
- You are prompted whether to treat sub-sites as samples? The default is "Yes" (see section 4.3).
- You are prompted whether to use NBN GUIs? The default is "Yes" (see section 4.2)

When you have filled in all the prompts, press <F9> to close this window and the export process will commence.

4.5. Bulk transfer

This option is available using NBN - Bulk move. It differs from NBN export as follows:

- The entire database is exported in contrast to NBN export which only exports those records in the current select list.
- It is able to export a wider range of data from *Recorder 3.3* (see section 6.4),
- Instead of writing an XML file in NBN transfer format, it writes a series of comma and quote separated (CSV) files - one for each *Recorder 2002* table.
- It does not use the NBN data import mechanism built into *Recorder 2002*, but requires a special add-on (supplied via the NBN website) to be installed.
- The NBN transfer route has been found to work well with up to about 30,000 species records (on a computer running NT 4.0 with 128Mb RAM). With larger XML files, problems have been encountered with the operating system running out of system resources. The Bulk transfer mechanism has been successfully tested with much larger datasets including one with 375,000 species records.

To perform a bulk export:

- Select NBN - Bulk move from the menus.
- You will see a window similar to the following:



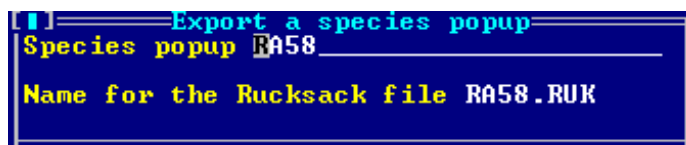
- You are asked to choose a directory on which the files will be written and your default output path is suggested. Enter any valid, DOS directory path here.
- You are asked to identify the dataset. The options key (<F2>) lists all available dataset definitions (see section 3.2). If only one definition is available, it is automatically filled in.
- You are prompted whether to treat sub-sites as samples? The default is "Yes" (see section 4.3).
- You are prompted whether to use NBN GUIs? The default is "Yes" (see section 4.2)
- When you have filled in all the prompts, press <F9> to close this window and the export process will commence.

4.6. Exporting Species popups as *Recorder 2002* Rucksack files

This tool writes the species list from a *Recorder* Species popup to the Taxon list of a *Recorder 2002* Rucksack file. However, it can only export "system supplied" species (the items that came in the *Recorder 3.3* SPECIES table as part of the original installation). Any entries that have been added by a user will not be included in the Taxon list.

To use the tool:

- Select NBN - Species popup from the menus
- You will see a window similar to the following:




- In the "Species popup" prompt, type the name of the Species popup you want to export. The Options key <F2> lists the available Species popups - just choose one from the list.
- In the "Name for the Rucksack file" prompt, type the file name you want to use to save the rucksack file. This must have a ".RUK" extension to be recognised by *Recorder 2002* as a Rucksack. You are offered a default which is the name of the Species popup (truncated to 8-characters if necessary) with the extension ".RUK" added. Note that, whilst *Recorder 2002* supports long file names, *Recorder* does not - so you are restricted to an 8-character file name which conforms to the DOS rules for naming files.
- Press <F9> to perform the export and close the window. You will see a message reporting that the file has been written and giving the full name and path of the file. The file will be written to the directory used by *Advanced Revelation* for its temporary files (by default this is **\Recorder\Output**).
- To use the exported file with *Recorder 2002*, you need to move it to the directory from which *Recorder 2002* expects to read Rucksack files. You can find this location by looking in Tools - Options - File locations in *Recorder 2002*. By default, it is a directory called **\Recorder 2002\Rucksacks**.
- It is recommended that you rename the file when you move it to the Rucksack directory. *Recorder 2002* uses the name of the file to identify the Rucksack, so it is useful to give it a longer, descriptive name. For example, rather than leaving the list of hoverflies based on the BRC card "RA33" as "RA33.RUK", rename it to "Hoverfly list from BRC card RA33.RUK" - it will then be much easier to identify when you want to use it.

5. Importing into *Recorder 2002*

5.1. NBN xml export

XML files are imported using Tools - Import Data. This command displays the Import dialog. In the Import dialog:


- Select "NBN Data" as the Import Type
- Select the XML file you generated from *Recorder* in the Source prompt. You can type the file path and name or click the  button and locate the file. By default, the file will be located in your \Recorder\Output directory.
- Click the [Analyse] button to start the import process.

See the "Import NBN Data" topic in the *Recorder 2002* on-line help for further information (press F1 whilst the Import dialog is open, search for "Import" in the index or go to Tasks - Exchanging data - Importing NBN data in the Table of contents).

If you choose to use the xml transfer you will need to have some dataset details to attach to the transfer.

5.2. Bulk transfer

5.2.1. Performing the import

- In *Recorder 2002* select Tools - Import Data ...
- This displays the Import dialog. Select "Recorder 3 Import" as the Import Type. (This import type will only appear in the list if the "Recorder 3 Import" add-in has been installed (see section **Error! Reference source not found.**)
- Select the EXP file you generated from *Recorder* in the Source prompt. You can type the file path and name or click the  button and locate the file. By default, the file will be located in your \Recorder\Output directory.
- Click the [Analyse] button to start the import process.

See the "Import NBN Data" topic in the *Recorder 2002* on-line help for further information (press F1 whilst the Import dialog is open, search for "Import" in the index or go to Tasks - Exchanging data - Importing NBN data in the Table of contents).

If Invalid items are found you need to abort the import process, correct the problem at source in *Recorder* and re-export the data to try again.

5.2.2. Error reports during import

If error occur in the process of loading data from the CSV files into the temporary database you will see a count of errors appear in red at the bottom right-hand corner of the progress dialog. In this case a file "Rec3Import_Errs.txt" will be written to the directory from which the CSV files were read. It contains the error messages reported by Access. This usually indicates an internal failure in the program and there is probably little you can do about it.

Once the "Import NBN Data" dialog has appeared, and you have pressed the "Import" button, *Recorder 2002* may reject items. If this happens, a file called "ImportRejects.txt" will be written to the \Recorder 2002 directory. This file lists the table name and key of the items that were rejected and may give some indication of why. In most cases, items are rejected because some required data in the rejected row, or in another row it relies on, is missing or incorrect. You may be able to fix these problems and then re-try the import.

6. Issues

6.1. How long will it take?

Here are some example timings for Bulk move (Pentium III, 500MHz, 64mb RAM, NT 4.0).

Number of records in table				Time to export from Recorder 3.4 (h:m:s)	Size of export files (mb)	Time to import to Recorder 2002 (h:m:s)	Size of temp mdb (mb)	Size of nbndata.mdb (mb)
RECORDS	SITES	PERSONAL	LITERATURE					
374,548	17,657	1,989	1,452	05:42:00	173	04:44:22	252	318
29,063	858	189	408	00:20:41		00:19:53		

Comparison of XML and Bulk move on a test data set with 2,472 species records:

	Time to export from Recorder 3.4 (h:m:s)	Size of export files (bytes)	Time to import to Recorder 2002 (h:m:s)
XML	00:05:03	7,094,567	00:09:30
Bulk move	00:02:24	1,723,501	00:05:13

6.2. How much space is required?

The Hoverfly Recording Scheme RECORDER\DATA directory occupies 95.8mb in Recorder. As the table above shows, the bulk move produced 173mb of export files. During import the temporary database occupied 252mb and the finally database (nbndat.mdb), 318mb. Assuming Recorder and Recorder 2002 are installed on the same machine, this gives, at maximum, 839mb of files (i.e. before the export files and temporary database are deleted) in use. Thus, getting on for 1Gb of free disk space was needed to complete a transfer of this scale.

The test dataset with 29,000 records produces a 45mb XML file. This is about as big as Recorder 2002 can handle on the particular test machine (running NT 4.0 with 64mb of RAM) without running out of system resources.

6.3. Known problems

If you do an export using NBN GUIs to an XML file and import that file, then export another set of data to XML with some records overlapping (e.g. they have recorder's names and sites in common), again using NBN GUIs, then Recorder 2002's import process will correctly determine that there are duplicates in the data and offer you the usual options for dealing with them. This will prevent duplicate Names and Addresses or Locations appearing in Recorder 2002 as you would expect. Unfortunately, you will find that the Locations, Individuals, etc involved have some items of associated data duplicated. For example, you will find that the address and contact numbers for a person are duplicated. This happens because Recorder 2002 does not just store one row for an entity like an individual, it stores associated information like addresses or contact numbers, in other tables which have their own NBN keys. Recorder can generate an NBN GUI for the main record in the INDIVIDUAL table, but it cannot supply NBN GUIs for all the associated tables and new keys are assigned by Recorder 2002 during import. Thus when the data is re-imported, only the INDIVIDUAL row is detected as a duplicate, all the associated data like ADDRESS and CONTACT_NUMBER is not detected as duplicate and is imported again. There does not seem any way round this, given the big differences in the structure of tables in Recorder and Recorder 2002. The duplicated items can however, be deleted without problems. This does not affect Recorder 2002 to Recorder 2002 data transfers because, in this case, all keys for all associated tables are transferred, so duplication of any part of the data will be correctly detected.

6.4. What is and is not transferred?

LITERATURE Keywords are not exported. These are not included covered by the NBN data model, so this cannot be fixed in future.

Pictures relating to the site in SITES are not transferred. XML export is more limited than Bulk move. It does not cope with events, site keywords, lists of habitats with areas in sites, management aims and methods or associated species. These are exported by Bulk move.

7. Concepts

7.1. Documenting datasets

If the objective of sharing biodiversity data is to be achieved, one of the changes that is necessary is for those organising and running surveys which collect such data to document what they do much more thoroughly and consistently. If someone else who had no contact with, or knowledge of a particular survey is to make effective use of data and information derived from it then they need to be able to find out a number of details such as:

- What was surveyed and over what geographical area and time period?
- What sampling methods were used?
- What quality controls checks were made to ensure the accuracy and validity of the data?
- Is the data available for use and, if so, under what terms and conditions?
- Where is it located and what format is it in?

This sort of information is termed *meta-data* (i.e. data about data) and is essential to the NBN in establishing a catalogue of data sources. Such a catalogue will be the first place where a potential user can look to find out if there is anything available in the NBN relevant to their particular query or project and where to go next to find out more.

The NBN transfer format has been designed with a number of required meta-data elements to ensure that the necessary documentation is available. It is important for the future use of the dataset that these meta-data are carefully considered and are as representative as possible of the data.

7.2. NBN transfer format

The NBN transfer format uses eXtensible Markup Language (XML) as a means of transferring data. XML is a language designed to describe the contents of a document held in an ordinary text file. XML looks like HTML (Hyper-Text Mark-up Language - the language of the world wide web), which is by now a familiar format to most computer users. Like HTML, XML uses *tags* (words bracketed by “<...>”) and *attributes* (terms like *name=“value”* inside these tags). For example:

```
<spatial_ref system="OSGB">NZ1265</spatial_ref>
```

In this case, the “system” attribute indicates that this particular spatial reference, “NZ1265”, is an OS National Grid reference for Great Britain.

However, HTML defines a limited set of tags and attributes and, in many cases, specifies what the text in the tags must be like. XML only defines the rules for defining tags and attributes. The specific tags and attributes for a particular application are entirely up to the designers of that application. A set of tags and attributes can be formally defined in a Document Type Definition (DTD) which is itself written in XML. The DTD for *Recorder 2002* transfer files can be found in **Recorder 2002\DTD\nbndata.dtd** in a standard installation.

XML is licence free and well supported. It is being increasingly supported by software packages and tools. For example, Microsoft Office 2000 uses it extensively and it is supported by the latest version of Internet Explorer and Netscape browsers. There is an extensive and rapidly growing body of literature available both via the internet and in publications.

XML files are easy to generate and read (by a computer). They are unambiguous and avoid common pitfalls like platform-dependency and lack of extensibility. XML provides support for internationalisation and use of different character sets.

XML is text, but it isn't meant to be human-readable. Because XML files are simple text files, they can be transferred between different types of computers without difficulty and can be examined and edited using any text editor if necessary. However, they are meant to be written and read by computer programmes, and are not designed to be easily read by people.

One drawback is that XML files are extremely verbose. Wrapping each datum in a tag makes the files very repetitive and large - usually much larger than the original data they represent in the internal format of the application that generated them. However, disk space is no longer expensive, and highly repetitive text files zip very efficiently for transfer by disk or across communication lines. XML files typically compress by at least 98%.

7.3. NBN Globally Unique Identifiers

When data are stored in a database table, each row typically has to be identified by a unique “key” which is used to find and retrieve that row. Keys can be constructed in many ways, but a common scheme is to use a running number. Thus, if you had 1,200 observations in a table, they would be identified by numbers 1 to 1200. The problem with this simple scheme is that if **A** wants to submit data to **B**, then what happens when **A**’s record 5 is copied into **B**’s database which already has its own row 5? Either **B**’s record will be overwritten by **A**’s and lost, or **A**’s data will be rejected and won’t be copied.

One way round this is for the import process in the application to detect that **A**’s data have keys that are already in use, and renumber the incoming records, assigning the next available running numbers in **B**’s database. This is the approach used in *Recorder 3.3*. In this scheme, **A**’s record 5 might become record 34567 in **B**’s database. The problem with this is: what happens if **A** changes record 5 and wants to update **B** (and perhaps other places to which copies have been sent)? If **B** simply imports the record again, it will be renumbered to the next available running number, and **B** will now have two copies of the record - the original and the updated versions. To avoid this, the application would need to keep track of the information that **A**’s record 5 is **B**’s record 34567 - and, if this is done for all imported data, it quickly becomes a considerable overhead in managing the data.

A better solution is to try and ensure that all data items in all copies of the software are uniquely identified by assigning a “Globally Unique Identifier”. A simple way to do this is to assign each copy of the application a unique identifier, and to form the key by combining the copy number and the running number in the table. **A**’s records might then have keys “A1” to “A1200” and **B**’s “B1” etc. Now, A5 can be copied into **B**’s database without clashing with any of **B**’s records. When **A** updates A5 and sends a new copy to **B**, it is absolutely clear which record is involved. If records are ‘date stamped’ when they are first saved, or when they are changed, **B** can even tell whether this is a new copy of A5 (because its date stamp will be later than the copy that is already present), or simply a duplicate which can be ignored.

A scheme of GUIs, consisting of a unique “Site ID” and running numbers in each table has been devised for the National Biodiversity Network. Each copy of Recorder 2002 is given a unique identifier consisting of an 8-character string of upper case letters (A..Z) and digits (0..9). Each table also has a running identifier similarly consisting of an 8-character string of upper case letters and digits. Because 36 possible characters are involved, this allows $36^8 = 2,821,109,907,456$ possible values for each of these strings. The key for each record is then formed by concatenating the Site ID and the running ID to make a single, unique, 16-character string.

Using this scheme, the key of a record might look something like:

JNCC000100002A27

7.4. Data “ownership”

Another of the changes that is necessary if the NBN objective of sharing biodiversity data is to be achieved involves proper management of data ownership and the issues it raises.

The law makes an important distinction between the ownership of an object itself and the ownership of intellectual property rights in it. The physical owner has the usual rights to sell, lend, display or even destroy the object, but the intellectual property right owner essentially controls the right to reproduce that object. Ownership of an object does not give you the right to reproduce it.

A biological record, once it is written down, is legally an "original literary work" in terms of the Copyright, Designs and Patents Act, 1988. The person who originally wrote them down is likely to be the owner of the copyright of biological records unless they were made by an employee in the course of their employment, in which case the employer will own the copyright (unless they have reached an agreement to the contrary). Copyright persists for 70 years.

If a recording scheme or record centre has collated records from a number of people, they probably own the collection of records, but the intellectual property rights of individual records will still be owned by the original recorders who, therefore, have a right to control how those records are reproduced. Authors also have the right not to have their work subject to derogatory treatment (i.e. not distorted in a way that is prejudicial to their reputation) and, if their work is commercially published, performed in public, etc., to be identified as the author.

It is therefore important to owners of collations of other people's records that they can:

- Track where the records came from and who is the original owner so that they can seek the owners permission, for commercial use for example, or attribute authorship correctly.
- Be very careful about any making any changes to somebody's intellectual property to avoid distortion.

Apart from issues of intellectual property rights, there are also common-sense issues of data management which need to be considered. If a recorder submits some records to a recording scheme which submits them to a validation process which leads to re-determination of a species and then changes the offending record to the new species names then there are now two different copies of this record: the original recorder still has the original record with the old species name and the scheme has the updated copy with the new name. Suppose a local record centre now asks both the recorder and the recording scheme for copies of records relating to its geographical area. It will receive two records, apparently of different species recorded on the same site visit.

Recorder 2002 attempts to tackle these issues.

- The Site ID of each record carries the identity of the copy of the software on which it was originally entered. All records are stamped with the creation date and the identity of the user by whom it was created. The last date on which it was changed and the identity of the user by who made the change is also stored. Note that an issue arises if the person who originally wrote down the record gives the manuscript (recording card, field notebook) to someone else to enter. The person who does the data entry will be flagged as the "owner" as far as the electronic copy of the record is concerned.
- *Recorder 2002* does not let you change a record belonging to someone else. That is, if the Site ID of a record is different from the Site ID of the record that is currently being edited, *Recorder 2002* will not allow changes to be made. This protects the intellectual property rights of the original owner. But, what if the record is erroneous in some way:
- If there is an error a part of the record (e.g. a Location description contains spelling errors, the spatial reference given for a Location is wrong), you should report the error to the original owner. They should make the correction and then export the amended version. Once this has been imported, both copies have the correct version. This way, changes can only be made to the "master" version of a record and the situation does not arise where different copies of *Recorder 2002* have versions which have had different changes made.
- However, *Recorder 2002* does allow you to add new sub-records (e.g. determinations to taxon-occurrences, Location names or designations to Locations). In the example described above, the recording scheme can validate records and re-determine species. Although they cannot change the original determination, they can add a new determination and flag it as "preferred" (the one that will be used for reporting). This creates a new determination record which is "owned" by the verifier and it is absolutely clear what has happened. Now, if a local record centre asks both the recorder and the recording scheme for copies of records relating to its geographical area it will still receive two records, but it is clear that they are the same record, one of which has been added to. They can make a reasoned decision about which determination they wish to use (presumably based on their opinion of who is the most authoritative determiner).

One consequence of this is that, if you export records from *Recorder 3.4* using NBN and import into a copy of *Recorder 2002* with a different Site ID, you would not be able to edit any of the imported records. To prevent this happening, existing *Recorder* users will be allocated a Site ID based on their existing *Recorder* copy number. For example, if your existing *Recorder* copy number is 151, then your *Recorder 2002* Site ID should be "00000151". This will be the Site ID exported by *Recorder 2002* so you will "own" all the imported data and can edit it.