

## **Cambridgeshire and Peterborough Environmental Records Centre (CPERC)**

### **Data Management Policy**

July 2016

#### **Contents**

- 1) NBN Trust guidance on data management
- 2) Definitions
- 3) Software used in the data management process
- 4) Data selection
- 5) Data formats
- 6) Digitisation
- 7) Validation
  - 7.1) Summary
  - 7.2) Validation procedures
    - 7.2.1) Visual inspection
    - 7.2.2) Automated procedures
    - 7.2.3) More details on the validation of who, what, where and when
    - 7.2.4) GIS data validation and storage
- 8) Verification
  - 8.1) Summary
  - 8.2) Guidance for verifiers
  - 8.3) How this relates to the Recorder database
  - 8.4) Selection of verifiers
- 9) Archiving, backing-up, and updating data

## 1) NBN Trust guidance on data management

Taken from:

Improving wildlife data quality: Guidance on data verification, validation and their application in biological recording

Section 3: What makes a good wildlife record?

If we are making a wildlife record, there is not much point in doing so unless it is as correct and complete as possible. It becomes increasingly important for wildlife records to be 'correct' the more these are used by others in understanding or making crucial decisions about biodiversity. The creation of a wildlife record is therefore a means of creating a 'true' statement about the occurrence (or even the absence) of a species at a particular locality at a particular time. However, the number of variables involved is often considerable.

Check Box 5: Compiling data - quality control checks and procedures

- Aim to acquire raw data in standard formats (e.g. standard recording forms or through using data loggers, the latter enabling direct download of data).
- Ensure necessary verification procedures have been carried out, preferably before collation of data into databases.
- Consider using quality-control checks on data entry (e.g. double-entry).
- Use standardised data entry systems (e.g. purpose-built databases or adapted spreadsheets, with in-built taxon checklists, habitat codes etc.).
- Use recognised standard term lists, taxon checklists, habitat codes etc. wherever possible (e.g. NBN Species Dictionary).
- Ensure all relevant parts of records are retained during data capture, including details of determinations, locations of vouchers, sources of records etc. Arrange for original records to be archived as a back-up.
- Aim for standardised data formats (e.g. dates, place-names, uniform formats of locality details, personal names).
- Carry out data validation routines on data entry (grid refs, dates, sources).
- Remember it is easier to correct a record at the start than it is to expunge a faulty record once it has been disseminated.

## 2) Definitions

Reference is made through these documents to 'records', 'data' and 'datasets'.

### **Record**

A record is assumed to be a single item of biodiversity information. For example, this could refer to the sighting of a particular species; or the incidence of a given habitat type. The record is an un-processed piece of information received by CPERC to be shared with other users.

### **Data**

Data are a collection of records, though usually taken to be digitised, and therefore removed from their original context. Data are records that have been processed in some way. Beyond the basic requirements for records (see below) data are unstructured collections.

### **Dataset**

A dataset is a structured collection of data which is managed as a discrete unit. Individual records (pieces of data) can be in several datasets at the same time (e.g. all records for VC31, and all plant records); and can be added or removed from datasets over time (for instance, after verification a record would no longer be in a dataset of unverified records).

For the purposes of these documents, data is often used as a catch all phrase to mean collections of processed and un-processed, structured and un-structured records. For most purposes 'data' and 'records' can largely be thought of as interchangeable. Datasets should always be taken as a structured collection of data. Ideally all data within a set should have been verified and validated to a common standard.

## 3) Software used in the data management process (as at July 2016)

- Recorder 6.24 - Central CPERC database used to store species records
- MapInfo v12 - Main GIS platform used to store and query site information, habitat data and species data.
- Microsoft Excel 2010 - Used to store records prior to importation to Recorder and manipulate records after export for end users.
- Microsoft Access 2010 - Used to query large datasets and access the Recorder database tables for queries or update
- SQL Server Management Studio Express - Also used to access the Recorder database tables for queries or update

## 4) Data selection

CPERC gives priority to information relating to the biodiversity of the Cambridgeshire and Peterborough area. This is most commonly in the form of species records, habitat survey information and information relating to statutory and non-statutory designated nature conservation sites. Since 2010 CPERC (formerly CPBRC) became an environmental records centre and as such is aiming to increase its data holdings of natural environment data above and beyond that relating to biodiversity, such as information on local geological sites of interest and flood plain boundaries. However, CPERC's data holdings are still primarily biological.

More recent information is deemed to be of more value to the records centre and is therefore a higher priority for data capture. With regards to species records, CPERC encourages all recorders of all abilities to submit records and does not explicitly favour or value records of one species over another. However, some species records (such as those of protected and UKBAP species) tend to be requested from enquirers more frequently than others, and so the work of the records centre is somewhat guided by the demands placed upon it by the needs of its users.

When a dataset is acquired, the records centre will process it in its entirety, rather than selectively processing key records, leaving the other records unused.

## 5) Data formats

Data comes into the records centre in various formats:

### Paper data

Ideally the original survey notes or recording card - though a transcribed version can help

### Digital (non-GIS) data

- MS Word
- MS Excel (preferable to Word)
- MapMate
- Recorder export - although older versions may cause problems during importing/exporting and there may be issues over custodianship
- Records exported from online recording systems, such as iRecord are usually exported as a csv file and opened in MS Excel

### GIS data

Vector: MapInfo.tab files are preferable  
Interchange .mif and .shp files are also usable

Raster: Uncompressed .tif files are preferable. If compressed .jpg or .ecw  
If geo-registered, needs be as a .tab file, or a GeoTIFF

GPS device data can be downloaded onto GIS systems

Accompanying metadata may be important for digital datasets

## 6) Digitisation

The method of digitisation will vary depending on the data, and what it has been collected or will be used for. Wherever possible, the original raw data is digitised, as this will ensure more data attached to the record is captured, and will enable more effective validation of the record. All relevant parts of records are retained during data capture, including details of determinations, locations of vouchers, sources of records etc.

Standardising data content and format helps validate the data. Once data is standardised it is also easier to combine with previously standardised data, which is then easier to manipulate and query. Be aware that standard formats should not be imposed if it leads to any loss of valuable information.

Paper species lists are usually digitised into MS Excel and ad-hoc species records can be collated into an Excel list, or entered directly into Recorder. Map data is digitised in MapInfo, usually against OS MasterMap. Date, method, and name of digitiser(s) are noted in the metadata.

When large amounts of standardised records are to be digitised (e.g. taken from a survey form), a record card can be used or developed in Excel. Record cards make data entry easier, and help to standardise the format of the data, which in turn aids validation (e.g. dates, place names, uniform formats of location details, recorder names).

It is an NBN principle that original data should never be altered, the original and any subsequent determinations associated with records should be preserved, and that species name changes (even changing an old name to a current one) should be regarded as re-determinations.

Original source data will be kept, either the original, copies, or through access to the recorder's notes. With the permission of the owner of the records, source material for unverified data will be made available to verifiers if requested.

## 7) Validation

### 7.1) Summary

Records should be validated to ensure the structure of the record is correct. All data coming into the records centre should be validated regardless of the expertise and experience of the recorder and/or digitiser involved.

Validation should not be confused with verification. Verification is checking the likelihood of the record being true, i.e. is it likely that species A, was seen on date B, in locality C. Validation includes, for example, checking that A has been spelt correctly; B is not in the future; that C exists, and that the record has the correct grid reference.

Validation:

- will check the record is complete, and will cover all attributes of the record
- is best carried out by comparing the digitised and un-digitised record against the original record notes (as data is cleaned and processed, it becomes increasingly difficult to spot errors - this is also true for verification)
- will (wherever possible) be undertaken by someone other than the original digitiser(s)
- is made easier by compiling records to an agreed standard
- happens right through the process of managing data, from when the records are first viewed through to release of the data to users

All records must possess, as a basic minimum, the following attributes to be considered valid:

- **Who** made the record (information may also be collected on subsequent determiners)
- **What** - the organism (animal, plant etc) recorded. It can be a common or scientific name, and can be at any point within the taxonomic hierarchy; though obviously more precisely defined records will be more useful.
- **When** it was recorded
- **Where** it was recorded. The location where the recording was made. The description of the location should be sufficient so that it can be identified by others at a future date.

Records which do not meet this minimum standard are considered invalid. Where possible invalid records should be returned to the data provider for clarification. If more information cannot be provided to make it valid, it should be excluded from the records centre's main data holdings.

Initial validation checks can sometimes flag records as not being valid where only simple corrections are needed, particularly when automated procedures are used. Data can be validated using visual inspection or by automated procedures. Where automated procedures exist, they will be used to supplement visual checking. Automated procedures alone are not sufficient and may lead to further errors without extra care being taken. The methods used in validation should be noted in the dataset metadata.

Ideally whole datasets (or at the minimum single surveys) would have all records processed together. Consequently, where data is found to be invalid, where possible work processing the data should stop whilst the records are clarified.

Extra information given as part of the record, such as further comments on the record, will also be digitised and stored as part of the CPERC data holdings.

Beyond this, there are no formal prescriptions on how data should be structured, although use of Recorder 6 standardises the data format to a certain extent and enables it to be in a form suitable for sharing on the NBN Gateway along with other data providers.

It is important that all information in a record is retained in digitised data. If the information does not fit the way the data is stored, then the way the data is stored may need to be adjusted to accommodate the information.

## 7.2) Validation Procedures

General validation checks:

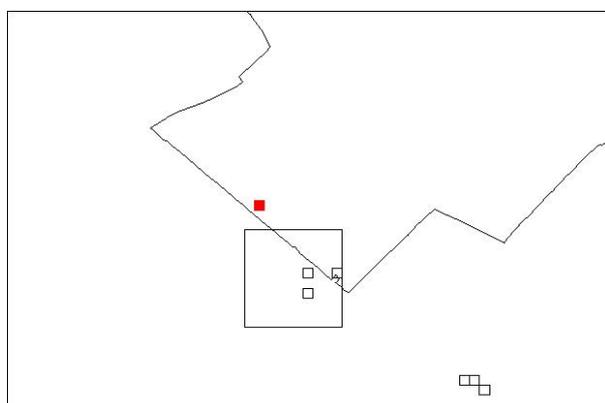
- Ensure that all mandatory columns are present (i.e. who, what, where, when)
- Each row of data has the correct number of fields
- Dates are valid in the calendar sense
- Grid references are in the correct format and match the given location description
- Values in the 'Projection' field are correct (for GIS data) (i.e. OS BNG)
- Values in the 'Precision' field are correct (for GIS data) for the grid reference precision given
- Fields that should contain numbers or dates just have number or date values in them
- Values in a field are no longer than the maximum length allowed (for GIS Data)

### 7.2.1) Visual inspection

Visual inspection can include:

- Checking names of recorders
- Checking recorded taxa
- Checking locations - through checking over digitised lists or mapping spatial data
- Checking dates
- Checking the record is complete - is all data from the original record captured by the digitising
- Checking of record sources - e.g. data derived as part of a named (series of) survey(s), or collected for an Environmental Statement. A reference to the data source should be stored with the final digitised data.

Mapping data for visual inspection is useful, even for data which would not ordinarily be mapped:



A record (in red) lies outside the county boundary. It may indicate a grid reference error. However, this may also be a valid record and needs checking against the original source. Data from outside of Cambridge and Peterborough can be preserved in Recorder, but it should be passed to the adjoining records centre for their use (with the permission of the owner of the data).

### 7.2.2) Automated procedures

Automated procedures can ensure records comply with a standard format, or can check records against look-up tables (such as a table of observer names). The use of drop-down lists, and pre-defined recording cards can help avoid problems with data validation later in the process.

Automated procedures can include:

- An automated check in an Excel spreadsheet. This can check the format of records (grid reference, date, site name), or can check the validity of records against look-up tables (e.g. dates for a survey period, observer/compiler/determiner names, location name)
- Mapping the data and running a simple tool to highlight records outside of Cambridgeshire and Peterborough
- Standard automation as part of entering or importing to the Recorder database can help to check the following:
  - Appropriate use of taxonomic names and authorities
  - Identifications validated against checklists
  - Statuses of taxa correct
  - Format of grid references correct
  - Grid references checked against counties/vice-counties or other defined geographic areas
  - Site names checked against standard gazetteers
  - Formats and contents of dates correct
  - Dates checked against survey periods
  - Observer/compiler/determiner names checked against standard lists
  - Validity of record sources checked
- The 'NBN Record Cleaner' tool can be used to aid validation, but should not be used without other validation procedures, particularly visual inspection from records centre staff

#### More details on Validation in Recorder:

Records should be validated before importing to Recorder. Recorder provides additional methods of validation, but should not be used as the only source of validation.

#### Dictionary

Species names should be imported and validated against dictionaries in Recorder. Recorder's name server tables aid validation by linking synonyms. In Recorder the 'preferred' dictionary lists should be used preferentially as these are the ones which are kept most up to date. Species names should, however, be imported into Recorder with the name as it was originally recorded, although this may not always be possible if the exact name does not exist in the Recorder dictionaries. Where this is the case only definite one-to-one relationship synonyms should be used, as one species may have been split into two since the original record was made for example.

### Sites and Sub-sites

Sites should represent reporting units i.e. locations that are commonly queried and reported on. Sub-sites (within those sites) should represent recording units within the larger site.

Records relating to existing sites within Recorder should be attached to those sites. Where the existing site is not the same as the location of the records to be imported, a new site should be created. This is particularly important when dealing with records which occur at locations where different designation boundaries may not match or have changed over time.

### Verified records

Records which have been verified will be given an appropriate Determination Type (see section 8 - Verification). Unverified records are given the Determination Type 'Requires Confirmation'.

### Confidential records

Records highlighted as confidential by recorders will be checked as confidential in Recorder or not imported into Recorder, as appropriate. These records will not be released in the ordinary course of work by the records centre. However, sometimes only part of the record is confidential, such as detailed comments, and in this case the confidential information can be removed from the parts of the record which are released to users.

## **7.2.3) More details on the validation of who, what , where and when**

### Who - recorder names

Recorders should (ideally) have one name or format of their name given for all their records. Records should be ideally associated with one individual and not a group or organisation, although individual recorders themselves can obviously be associated with such.

### What - species / habitat

If the name is validated by Recorder, it is acceptable. Recording cards should be based on those provided by the recorder or based on preferred lists if created for cards with new records - though other Recorder standard cards are acceptable if requested. Scientific names are more precise than common names and are always preferred. Common names can lead to ambiguity.

### Where - location

The location should be described by a site name and a grid reference. Having both a site name (or description) and grid reference makes it easier to validate data. The precision of the grid reference will vary depending on the record and its location. A location name and grid reference will make it easier for records to be assigned to a parish, administrative area and vice-county.

Records can come into the records centre without grid references. Where this is the case the description of the location given should be sufficient to assign an appropriate grid reference. All records stored in digital format shall be assigned grid references (OS BNG). Therefore where it is not possible to assign an appropriate grid reference a record shall be considered invalid. Grid references are essential to the work of the records centre as they enable querying of the data by geographical parameters.

Further notes on grid references and locations:

- Grid references are always preferable to post codes - six figure grid references (100m precision) are usually more precise than a post code (and easier to use), four figure grid references (1km) can also be more precise in some cases.
- Grid references from GPS data are more precise than those from paper maps. However, this can sometimes give the record a degree of precision the record does not warrant, or the appearance of a precision it does not really have.
  - GPS grid references give the location of the observer and not necessarily of the species being observed. For highly mobile species ten figure (1m) and possibly eight figure (10m) references may be inappropriate.
  - The standard deviation of the given grid reference should always be noted when using a GPS. SDs of 10m are not uncommon making ten figure (1m) precision grid references unreliable, and eight figure (10m) borderline for acceptability.
- A convenient way of assigning grid references to a site is to use the site centroid. Grid reference centroids for linear sites may be misleading (for long, thin sites the centroid may lie outside the site). For large sites use of a centroid would also not usually be appropriate.
- Appropriate site centroids should be stored in Recorder for reference. The appropriate precision of the site centroid is directly related to the size and shape of a site - it may be a 1km grid reference for a site which covers one or more 1km grid squares and a 100m grid reference for smaller sites.
- Where a site crosses grid lines (particularly 10km lines) records taken in one square may be given a grid reference in an adjacent square - this may unfortunately be unavoidable when accurate grid references are not given by the original recorder. Lower precision grid references should always be used in these circumstances.

Boundaries of sites can change over time and CPERC staff should always be aware of this when validating records and assigning site centroids. An ideal would be a series of unchanging boundaries for defined recording units. This is possible using Watsonian Vice-Counties and some parish based schemas, other standard site boundaries/locations can be developed over time by the records centre in conjunction with recorders - though even these standard location references may need to be modified over time as boundaries of key sites change, and habitat alters.

The imposition of artificial recording units can create problems when managing data in the future, and so should only be applied in a few clearly defined cases. As creating defined units is not likely to be possible for all recording sites, so it is important to retain old CWS, SSSI and nature reserve boundaries and units in particular to quantify how the boundaries of sites and units have changed over time. In this way older records can be put into context against newer ones.

Where possible, use of vague location names should be discouraged. If records can only be identified at a parish or higher level then that should be done. The over-riding principle is to maintain the integrity of the record. Different names for the same sites should also be stored in Recorder with the appropriate preferred name checked.

## When - date

**MS Excel:** The default country/region should be set to the UK allowing access to UK date formats. The format of the date should be suitable for its required end use (e.g. importing into Recorder/MapInfo). Vague dates (e.g. YYYY, SEASON-YYYY, MM-YYYY) should be stored in Excel as "General" or "Text" format as appropriate.

**Recorder:** Recorder accepts full dates (day, month, year) and vague dates including month and year, season and year and year only. Recorder can also accept date ranges from one date to another, although full dates are always preferred.

**MapInfo:** The "Date" data type in MapInfo consists of a day, a month and a year, DD/MM/YYYY. The "Text" data type should be used in datasets where full dates are not known for every record (i.e. year only or month and year) - in practice this is more likely.

## **7.2.4) GIS Data - Validation and storage**

GIS layers used in CPERC data searches should only contain data which meet the standard criteria for release (access agreement, metadata, verified and validated). Certain GIS layers (most notably the CWS layer) will be periodically distributed to users. Frequency of update and limitations on use and release should also be noted in metadata.

### Species layers

Species GIS layers can be created from exports of records from Recorder. Species records are represented in GIS by CPERC as squares, with the location of the squares based on the grid reference and the size of the squares based on its precision. This is based on the assumption that the species could have been recorded at any point within that grid square.

Wherever possible, the attribute tables of species GIS layers should follow a standard format. This allows records to be easily moved between layers to create new datasets.

The current CPERC standard species GIS data structure includes the following columns (although not every column has to have an entry for every record):

- Record ID (ID for this GIS dataset)
- Observation Key (ID derived from Recorder)
- Taxon Group
- Common Name
- Scientific Name
- Location
- Location Details
- Grid Reference
- Easting
- Northing
- Precision
- Date
- Abundance
- Record Type
- Comments
- Designations
- Recorder

- Source

### Site layers

Site layers will either be statutory site layers (e.g. SSSIs) obtained from Natural England, or non-statutory site layers obtained from the Wildlife Trust BCN and/or local authorities.

Site layers should contain the following as a minimum:

- Site name
- Site grid reference
- Site area

Additional columns may also be needed in site GIS layers:

- Site ID
- Polygon number
- Polygon grid reference
- Polygon area
- Site designation
- District authority
- Parish
- Reasons for selection
- Date of most recent survey
- Habitat types present

### Habitat layers

CPERC currently holds Phase 1 habitat information in GIS format. The current Phase 1 habitat GIS data structure includes the following information:

- Phase 1 Code
- Habitat Description
- Associated Species
- Target Note
- Comments
- Type (polygon or line)
- Digitisation Date
- Digitiser

Additional metadata regarding the original source of the habitat data that is digitised should be held with the relevant GIS layers.

Habitat areas are stored as polygons and target notes can be stored as points. Some features such as streams or rivers can only be digitised as lines depending on how they were originally mapped. However, polygons should be used wherever possible. Polygons should not overlap and each habitat polygon or parcel is mapped with only one Phase 1 habitat code. Habitat polygons are always mapped to OS Mastermap where possible.

Where habitat polygons have been classified into another habitat category (such as BAP Priority or NVC) other columns can be added to store this information. It is the aim of CPERC to translate all GIS habitat information into BAP habitat classifications where

possible. The Integrated Habitat System (IHS) developed by Somerset Environmental Records Centre can be used to help this process.

CPERC is currently developing its own broad habitat classification which will link with BAP broad and priority habitat types wherever possible. This is to be used when digitising new habitat data using up to date aerial photography in combination with previous Phase 1 data and other information sources. CPERC has already used this on a small scale for parish habitat mapping projects.

CPERC currently holds copies of the latest BAP habitat inventories from Natural England relevant to the Cambridgeshire and Peterborough area. CPERC has contributed towards producing some of these inventories since the inception of the records centre in 2005.

#### Aerial Photography Interpretation (API)

API can be used to map habitat or to interpret the boundaries of features that are not obvious from OS base mapping.

GIS layers should not normally be digitised from aerial photography alone. As OS mapping (MasterMap) is more accurate, that should ordinarily be used as the primary digitisation framework. Where habitat is being mapped and it does not correspond to features mapped on the OS layer, aerial photography can be used where appropriate.

#### General GIS layer creation notes

What a layer has been digitised against (e.g. MasterMap) should be noted in the metadata. GIS layers should usually be checked by someone other than the digitiser before use/release.

## **8) Verification**

### **8.1) Summary**

The ability to guarantee the quality of data, and to demonstrate this through metadata is essential to demonstrating the quality of the work of the records centre. Verification, along with validation is intended to ensure the accuracy of a record.

The records centre assumes the following in relation to verification:

- Where data can be verified it should be, and it is the aim of CPERC to have all of its records verified
- If it can be verified but hasn't been, it should not be released, as the quality of the record can be improved (and it may be incorrect)
- If it cannot be verified, then the quality of the record cannot be improved and in certain circumstances it can be released, but its unverified status should be clear to users from metadata

Data that has been both verified and validated is considered by CPERC to be as accurate as possible, and is fit for release without any further notes (other than metadata).

The following CPERC notes for verifiers explain CPERC's current policy on verification. These notes have been written so that the categories of verification status correspond with determination types in the CPERC Recorder database.

Records coming directly from a county recorder or appointed CPERC verifier are considered to be verified by that individual and therefore do not undergo any further verification procedures. However, data coming into the records centre this way will still need to undergo validation checks.

### **8.2) Guidance for verifiers**

#### **Cambridgeshire and Peterborough Environmental Records Centre Guidance for Verifiers**

These guidance notes are intended for those verifying records for CPERC. Species records come into CPERC from a variety of sources and from a variety of different recorders. Therefore it is important that records go through a quality control stage, where the likelihood of these records being correct is determined by an independent verifier (who has expertise in the relevant taxonomic group) before these records are released to data users.

Important questions to be considered by the verifier are -

- 1) Is it possible that this species was found at this location?
- 2) Is it possible that this species was found on this date?
- 3) Does this species need particular knowledge and expertise in order to identify it?
- 4) Are there any obvious errors in the record (such as grid reference and location not matching) CPERC will, however, aim to ensure that such errors are removed before the data is passed to the verifier.

CPERC would like the verifier to score every record with a score from 1 to 5. This explanation of this scoring system is outlined in the table below.

Score	Category	Explanation of Category
1	Correct	This is where the verifier was the actual observer or has positively identified a specimen.
2	Considered Correct	This is where there above are not true but the verifier has no reason to believe that it was not possible for that particular species to be recorded by that recorder at that location on that particular date. The majority of records will fall into this category.
3	Considered Incorrect	The verifier should score records in this category where they consider that it would be highly unlikely or impossible for the species involved to be recorded by that recorder at that location on that particular date.
4	Incorrect	This is where the verifier is absolutely certain that the record is incorrect because there are obvious unredeemable errors in the record or the verifier has examined a specimen or supporting evidence and determined the original identification to be incorrect. In this case a new correct determination may be given by the verifier.
5	Cannot Confirm	This is where there is insufficient information to verify a record and it is unlikely that any more information can be obtained to consider that record correct.  Additional evidence which could be provided by the recorder for the verifier to consider the record correct may include a photo or a sample taken at the recording event. CPERC may hold such additional evidence which can be provided to the verifier in such circumstances.

Due to the fact that the records centre obtains many records from members of the public and those within unknown expertise, to avoid large numbers of records being collected but not used, CPERC would like the verifier to mark all records as correct or considered correct unless there is a strong and justifiable reason to say otherwise. That is to say that all records are presumed to be correct unless a sufficient reason is given. Such reasons should be noted down by the verifier next to the relevant record and attempts may be made to provide additional evidence for records to be considered correct where possible.

### 8.3) How this relates to the Recorder database

Score	Determination Type in Recorder	Verification Flag Status in Recorder
1	Correct	Passed verification
2	Considered Correct	Passed verification
3	Considered Incorrect	Failed/pending verification
4	Incorrect	Failed/pending verification
5	Cannot Confirm	Failed/pending verification

Records awaiting verification in Recorder are given the determination type of 'Requires Confirmation' which corresponds to a verification flag status of 'Failed/pending verification'. 'Cannot Confirm' is a non-standard determination type added by CPERC to describe situations where there is significant doubt over a record (but not enough to consider it incorrect) which could be resolved by further information, but that information is not available as explained above.

## 8.4) Selection of verifiers

A verifier should be:

- An expert with significant experience and knowledge relating to the particular taxonomic group they are verifying
- Have local knowledge of the area that they are verifying records for, particularly with regards to the distributions of species within that particular taxonomic group

Many records centres are affiliated with their county naturalist's society who have appointed county recorders for particular taxonomic groups. In such circumstances these county recorders are often also used by the records centre as verifiers - as they should immediately fulfil both of the above requirements.

In Cambridgeshire and Peterborough there is no naturalist's society which covers the entire records centre area and there are three vice-county recording units within the area (one of which, the Soke of Peterborough part of VC32, is only a small part of the total area of that vice-county). Therefore the distribution of recorders for some taxonomic groups is patchy and in some areas there are none for the particular group of interest.

Therefore the records centre has to appoint verifiers based on the recording structure that already exists - for some taxonomic groups this will be vice-county or county verifiers and for other taxonomic groups it may be an individual affiliated to a county based recording group (such as the Cambridgeshire Bat Group) or an interested amateur with no affiliation to a particular society or group.

Obviously the competence of the verifier has to be assessed by the records centre before they are appointed and the verifier has to be willing and able to do the task within a reasonable time frame.

A verifier may consider that they do not need to verify every single record individually for some common and easy to identify species. This is left up to the verifier's discretion, although the verifier should inform the records centre if this is part of their method of verification.

A verifier may also choose to verify all records from a particular individual that they know and trust with a lower level of scrutiny. This is also left to the verifier's discretion but again the verifier shall inform the records centre of this and name these selected recorders.

### Habitat and Site data

Habitat and site data when digitised by the records centre should be passed back to the original data provider (which may be the original surveyor) for validation and verification, and any errors noted should be corrected. Ideally a separate competent surveyor should also check the data to look for any obvious anomalies.

## **9) Archiving, backing-up, and updating data**

Data is likely to be stored in one or more forms in the records centre, different archive and back-up procedures will exist for the differing forms.

### **1. Paper records**

Original source data will be kept, either the original, copies, or through access to the recorders notes. All data sharing agreements past and present will be also stored. Paper records to be suitably filed in filing cabinets.

### **2. Recorder database**

Stored on a networked SQL Server. Back-ups of the database are stored on the server, and also copied onto external hard drives. An automatic back-up has been set up on the Recorder server to back up the database overnight, every night, and keep the recent copies. Older back-ups are also stored separately and dated. All versions on the server are backed-up as part of the server back-up procedure.

### **3. MapInfo GIS layers**

Stored on the CPERC server, and backed-up as part of the server back-up procedure. All layers have a clear name and are dated. Older layers are archived on the server.

### **4. Other digital data**

Stored on the CPERC server, and backed-up as part of the server back-up procedure. Files of particular importance to individual staff-members can also be backed up to portable USB/hard drives.

All species, habitat and site file names should be dated, and archived when superseded. This allows comparison between an older and newer datasets.

## **Server back-up procedure**

A back-up to a removable data cartridge (Dell RD1000 500GB) is automated to occur on the server every night. CPERC has two of these which are swapped over every day, with one being stored off site out of office hours.