16 Data processing and management

Introduction

Although an LRC's priority is to supply products and services data management is the heart of all LRC activities As a data custodian, the LRC is responsible for managing all the data it holds. Policies and procedures for acquiring, documenting, processing and securely storing data should set out this role.

As data-sets are acquired by the LRC their ownership must be clarified and agreements drawn up with the owner (normally the recorder) on how the LRC will manage and disseminate data on their behalf. The LRC must be clear about what data it holds, and their status, including factors such as ownership, age, completeness and quality, and it must be able to pass these 'metadata' to users. Following their acquisition, all data must be processed, validated and incorporated into the LRC's data management systems. Original data-sets, whether computerised or on paper, should be archived for future reference and all data must be stored safely. This requires the LRC to ensure not only that access to data is controlled but that data are provided with adequate physical security.

17 Data ownership and acquisition
Policies & Principles

- **An LRC should enter into 'data supply agreements' with all data owners whose data it manages. These agreements should define the responsibilities of both the LRC and the data owners.**
- Ownership of the original raw data should remain with the recorder.
- Data supply agreements should adhere to the Copyright, Designs and Patents Act (1988) and the Data Protection Act (1998), but should be written in plain English.

17.1 Background

In order to function at all, an LRC must have access to data. The majority of these data will be supplied by people or organisations outside the LRC. Often, these recorders are local naturalists collecting or keeping biological records on their own behalf.

Many recorders are experts in particular taxa, and as such their data are invaluable. It is therefore extremely important that the LRC gains the trust of local recorders. A clear policy and procedure on data ownership and acquisition demonstrates that the LRC has an open and honest approach to the acquisition of data. It helps individuals to understand how their data may be used, and the extent to which they retain control.

Recorders should be reassured that their data will be used responsibly and that the LRC is run professionally. This can be achieved through a clearly written data supply agreement that sets out the responsibilities of both the LRC and the recorder.

Having developed an understanding with its recorders, the LRC also needs to send out a similar message to its potential users. They must be assured that the data are reliable and are used with the owners' consent. It is also important that the LRC can identify the owner of any particular data-set precisely and reliably, in order to avoid breaching copyright law and to ensure recorders are credited when their data are used.

When considering copyright, it is important to understand the different data types that are handled by LRCs. The data supplied to an LRC by a recorder are known as 'raw data'. The copyright to data is owned by the person who made the original observation and made a physical record (ie on paper or computer). However, if that person is paid to make an observation, then the copyright belongs to the employing organisation or client. The LRC may need to process data it holds in order to produce reports for LRC users; the copyright of such publications and reports is owned by the LRC, but recorders should be credited if significant amounts of their data are used.

Because data are owned variously by individuals and organisations, the LRC needs to produce a number of different data supply agreements to be used with different types of supplier. Data collection through surveys involving the general public is dealt with in section 15 *Public information gathering*.

17.2 Policy

In writing a policy for data ownership and acquisition, the LRC must achieve a balance between the needs of its data users and of its suppliers. It must consider three main policy areas: data supply agreements, copyright, and satellite copies of data.

17.2.1 Data supply agreements

The LRC must be able to supply data to its users; otherwise it cannot fulfil its function. This means that recorders need to give permission for their data to be used; but the data owners' rights must be protected. This can be achieved through the use of written agreements setting out the obligations of the LRC and those of the recorder.

Such agreements should ensure that the data will be efficiently stored and not misused, and that both the LRC and the recorder will enjoy the benefits of access to high-quality information.

The LRC should consider how it will keep track of who owns data, and how it will credit individuals who want crediting. When acquiring data, it needs to produce agreements that gain the confidence of suppliers while minimising the restrictions placed on data use. In general, agreements should be kept simple and user-friendly to encourage recorders to sign them, and the LRC should explain clearly how the data might be used.

In practice, the LRC must consider the burden of administration that agreements may create. It must also address the issues of verification (ie checking records to ensure they are accurate and correct), and the confidentiality of recorders' personal information in accordance with the Data Protection Act.

17.2.2 Copyright

Data owners have the right to withdraw their data at any time (see section 21 *Storing and archiving records*). In certain circumstances, the LRC itself may wish to remove a data-set from its holdings, perhaps if there is some doubt as to the quality of the data. Either situation calls for the LRC to be as tactful and diplomatic as possible. The LRC should attempt to dissuade a recorder from removing their records wherever possible, or note the reason for removal. It may be possible to resolve this reason in the future. The LRC should also be aware of any potential problems which may arise with more than one recorder.

Data owners have the right to be credited for use of their data if they wish. However, the LRC cannot operate efficiently if owners also place cumbersome restrictions on the use of data, so such restrictions should be avoided and discouraged. For example, if a data owner was only happy for their data to be released to certain users, administration would be extremely difficult. The LRC would have to consider whether it was viable to hold this data-set. These issues are discussed further in section 30 *Controlling access to data*.

LRCs often release data to third parties in published form. Publications and reports are owned by the LRC, as the copyright of any collation of data belongs to the collator. However, ownership of the raw data remains with the original recorder; it is not transferred to the LRC or to any party that subsequently uses the data. Individual data owners may wish to be credited if reports use substantial amounts of their data.

Data produced under contract to the LRC are, by default, owned by the LRC unless the original contract states otherwise, since the copyright of data collected under contract belongs to the party issuing the contract.

17.2.3 Satellites

In some cases, an LRC may authorise the use of satellite copies of data-sets, for example by a recording group. The ownership of the satellite data should still remain with the original recorder(s). These data should be used only as a background to further recording, and other uses must be authorised by the LRC. Copyright to any data added to the satellite data-set remains with the recorder.

17.3 Procedures

In developing procedures to implement its policies, the LRC should consider its own needs in terms of releasing data to suppliers and maintaining good working relationships with recorders. It needs a simple, easy and cost-effective mechanism for keeping a record of data ownership, and acknowledging this ownership according to any agreement. The system for recording ownership may be either paper or computer-based.

The procedures outlined here cover data acquisition and data removal.

It is important to define the types of agreement possible:

- A 'data supply agreement' only covers supply of data to the LRC by the recorder. Its opposite is the 'data release agreement', which the LRC makes with its users (covered in section 30 *Controlling access to data).*
- A 'data exchange agreement' covers data flow in all directions; between, for example, the LRC and recorders, and the LRC and users. It also covers satellites of the LRC and recorders who require copies of data already held by the LRC (eg representatives of national recording schemes and societies).

The LRC should maintain a register of agreements either on paper or on computer, containing the appropriate metadata. The register should be checked when data are released or new data are received. If *all* data suppliers are subject to agreements, this register could also function as the LRC's record of data ownership.

In order to ease the administrative burden, an LRC may wish to offer a standard agreement to recorders when acknowledging receipt of their data. This would be especially relevant for *ad hoc* records received on an occasional basis. Specific, tailored agreements would be the exception, perhaps used only with large organisations or for specific data-sets.

Different types of agreement should therefore be made available. Procedures are needed to guide LRC staff as to which agreement is most appropriate. When a standard agreement system is to be used, the procedure should set out measures for encouraging recorders to return signed agreements (eg sending out a stamped addressed envelope with the agreement).

17.4 Process of developing the policy and procedures

The procedures should be clear and simple so that the policy is fulfilled without the need for further explanation. They should integrate as fully as possible with other policies and procedures, particularly those on *Processing newly acquired data* (see section 19), *Storing and archiving of records* (section 21) *and Controlling access to data* (section 30).

When developing the policy and procedures, the LRC should consult all its partners. It is essential that everyone using and contributing to the LRC supports its policies, and that the procedures are self-explanatory and easy for staff to follow. Recorders should be involved in developing the data supply agreements they will be asked to sign, to encourage their trust and commitment to the process.

Case study
Data ownership and acquisition
EcoRecord
Background

EcoRecord is the LRC for Birmingham and the Black Country. It developed over a number of years, beginning in 1984 when a nature conservation strategy was produced for the county of the West Midlands. In 1986, the old West Midlands County Council established a 'Wildlife Records Centre, but this was put on hold when the County Council was abolished. It was reborn as 'EcoRecord—the ecological database for the Black Country and Birmingham in 1991, as a result of a partnership between the Urban Wildlife Trust (now the Wildlife Trust for Birmingham and the Black Country), English Nature and the local authorities of Birmingham, Dudley, Sandwell, Walsall and Wolverhampton.

Initially, EcoRecord was run by the Wildlife Trust and the Joint Data Team (JDT) on behalf of these five local authorities. The JDT had been a department of the County Council, and was continued as an independent unit when the County Council was abolished. The JDT was later sold to Mott Macdonald, which now, together with the Wildlife Trust, runs EcoRecord, with the local authorities providing funding through four year Service Level Agreements (SLAs).

Discussion

EcoRecord has produced a series of policies and procedures to ensure that the rights of data owners are protected and that the owners understand the ways in which the LRC might use their data. The policy on data ownership and acquisition makes it clear that ownership of the data remains with the recorder and that the LRC will enter into data supply agreements (that define this ownership and copyright) with all data suppliers.

EcoRecord has a sub-policy covering the use of satellite copies of data-sets by, for example, recording groups to further their biological recording. Satellite copies may not be altered by the users in any way, but new data may be added to the data-sets.

The procedure for data acquisition includes checking whether a data supply agreement already exists, and establishing one when necessary. EcoRecord proposes to send out stamped addressed envelopes to data suppliers with the proposed agreement. Some LRCs may find the task of administering this particular licensing system time-consuming or expensive, and may have to consider cheaper ways of obtaining signed agreements. Using standard agreements in the majority of cases is one possibility.

LRCs may wish to include more detail in the procedure on how to decide, for particular data-sets, which type of agreement may be required, or even whether an agreement is actually necessary. Agreements may cover a specific data-set, all the records supplied by an individual or all records supplied by an organisation, including both past and future submissions. Agreements may also include or exclude the right of the LRC to use data in publications.

EcoRecord uses 'data supply agreement to refer to the list of obligations placed on the LRC and the recorder. It also uses the term 'licence to refer to the actual document signed by the recorder and the LRC manager. This is not generally recommended, as the word 'licence is 'legalese and could potentially deter recorders. Other LRCs consider it to be unnecessarily complicated to use two separate documents. LRCs may find it easier to use one document that sets out the obligations of the LRC and the recorder, with a signed copy being kept by each party.

A procedure for the removal of records from the LRC covers the two main circumstances when removal might be necessary, and calls for tact and diplomacy in both. When the data supplier requests removal, EcoRecord will attempt to dissuade the supplier and will record the reason for removal should it go ahead. LRCs should ensure that they allow a realistic length of time for removal of records from the system, depending on staff levels and resources, making it clear to suppliers that it may not be possible for the LRC to remove their data from the system immediately.

It is hoped that the entry of dubious data into the system can be avoided through a rigorous verification procedure. However, in some circumstances it may be necessary to remove data. Sometimes data are reverified and deemed inaccurate. In such cases, the LRC should rely on its good relationships with recorders to encourage improvement in the standards of recording. There can also be circumstances, mercifully rare, when records are deliberately misleading and have to be removed.

Contact

EcoRecord manager

EcoRecord, 28 Harborne Road, Edgbaston, Birmingham B15 3AA   0121 454 1808

enquiries@ecorecord.org.uk

Policy on Ownership and Acquisition of Data

1 Policy Statement

1.1 The Local Records Centre (LRC) will enter into data-exchange or supply agreements with all data owners. These will define the responsibilities of both the LRC and the data owner.

1.2 The ownership of the original raw data will remain with the recorder unless this is waived in writing by the data owner.

1.3 Data supplier agreements will adhere to the Copyright, Designs and Patents Act (1988) and the Data Protection Act (1998), but should not be written in "legalese".

2 Background to policy

2.1 The LRC needs to define the ownership of all the data it holds, including raw data and processed information. It also needs to agree with the owners how the data will be managed and used.

2.2 Data collection during public involvement surveys will be covered separately, refer to the *Policy on Public Information Gathering*.

2.3 In this context, raw data are the information supplied to the LRC prior to any processing within the LRC.

2.4 The data owner is the person who has the intellectual property rights over the data. When data are first recorded in a physical form (e.g. on paper or computer), the person making the record owns the data. This person can then give this property to any other person, if they so wish.

2.5 This policy has three distinct sections:

• Data ownership

• Data acquisition by the LRC

• Data exchange with satellites

2.6 Within this policy the term "Data Exchange" covers the transfer of data in both directions, both to and from the LRC, whilst "Data Supply" only covers the supply of data to the LRC from external data owners.

3 Policy on Data Ownership

3.1 The LRC will ensure the data owner's rights are protected through the completion of an agreement between the LRC and the data supplier, defining both the rights and obligations of both the data owner and the LRC.

3.2 The LRC will not normally enter into data supply agreements if the supplier imposes restrictions that make the data unavailable to third parties, unless agreed by the LRC's Steering Group.

3.3 Data produced under contract to the LRC will be owned by the recorder unless conditions to the contrary are included within the original contract.

3.4 Publications and analysed reports will be owned by the LRC, however, credits must be included with the products to the original data suppliers if requested in their Data Supply Agreement.

3.5 Data owners may request that the LRC remove all or any part of their data from the LRC's data management systems. The LRC will remove the data in accordance with the agreed procedure on the removal of data from the LRC.

4 Policy on Acquisition of Data

4.1 The LRC will enter into a data supply agreement with all regular data suppliers stating the terms and conditions under which:

• The LRC agrees to accept and manage records from the supplier and

• The supplier agrees to adhere to NBN standards of recording

• This will define the ownership, copyright and exceptionally, any other terms not included in the LRC's policies.

4.2 This agreement will license the LRC to release data without further consultation providing the request falls within the guideline of the policy on the Release of Ecological Information.

4.3 The LRC will discourage any restrictions being placed on data that limit their availability beyond the LRC's access terms.

4.4 The LRC will maintain a register of data supply agreements.

4.5 The LRC will normally only release information to third parties in a processed form, although the LRC must ensure any processing does not alter the original data.

5   Policy on Data Exchange with Satellites

5.1   This sub-policy covers the use of satellite copies of LRC data by recorders for the furtherance of their biological recording.

5.2   The LRC and data owner will need to complete a "Data Exchange Agreement" similar to section 4 above, but with the following additional requirements.

5.3   The ownership of the original raw data added to the satellite will remain with the recorder unless this is waived in writing by the recorder.

5.4   The ownership of data already existing in the satellite remains with the original recorder(s) and can only be used by the satellite users as background research for their own recording. Any other uses must be authorised by the LRC.

6   Links to other policies

6.1   This policy is closely associated with the following policies:

- Sources of data
- Public information gathering
- Documenting data
- Processing newly acquired data
- Physical security of data
- Storage and archiving of records
- Data services offered by LRC
- Relations with data providers (organisations and individuals)
- Controlling access to data

7   Procedures

7.1   This policy is implemented through the following procedures:

- Data acquisition
- The removal of data from the LRC

Procedure for Acquisition of Data

1   Introduction

1.1   The following document is intended to guide LRC staff through the procedures for ensuring ownership of raw data both on computer medium and paper is recognised and acknowledged, together with any processed analyses, in accordance with the accompanying "Policy on Ownership and Acquisition of Data".

Procedures for the processing of the acquisition of records

Inputs	*Procedures*	*Outputs*

Data Supply

Data Supply Agreement Exists?

No
but required or advisable

Yes
or not required

Establish Data Supply Agreement

Yes

Investigate why not and encourage agreement

No

Enter data supplier on to DS Register

Send Supplier copy of agreement

Ensure Recognition of original Copyright is enforced

Pass to: Processing of Newly Acquired Data Procedure
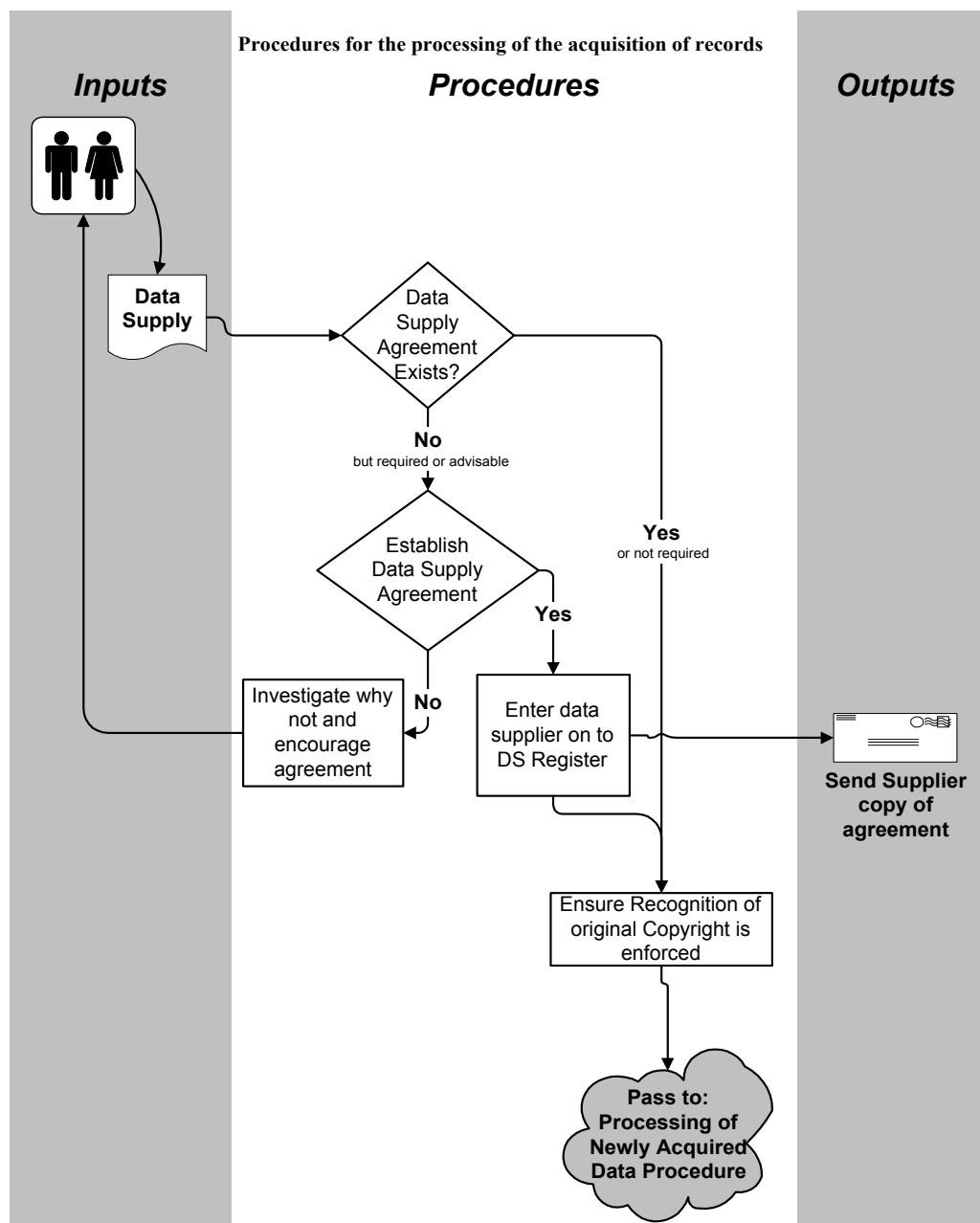
In this context, raw data is the information supplied to the LRC prior to any processing.

2 Preliminary Procedures

2.1   The LRC must maintain a register of completed data supply agreements including at least the following information:

- Name and address of supplier. A reference code would be advisable and in a computerised system, the Recorder ID for the recorder would be the most convenient.
- Taxa or biotopes covered by supplier as an agreed "expert"
- The format in which the bulk of data from the supplier will be received
- Any restrictions placed on the LRC's use of the data by the supplier
- Does the supplier wish to be credited when data is used

2.2   This register may be on computer or paper, but must include facilities for storing completed data supply agreements – refer to the procedures for storage and archiving.

3 The Licence Agreement

3.1   When new data are received by the LRC, the Data Supply Register should be consulted to check if a licence agreement exists (either data supply or data exchange), if it does, any restrictions should be noted and attached to the new data, then the data are passed to the computerisation and storage procedures.

3.2   If no data supply agreement exists, then it is necessary to decide what type of agreement is suitable, and this may involve liasing with the original supplier. There are five types of Data Supply Agreement:

- Covering all records supplied by an individual, both past and future excluding the publication by the LRC
- Covering a specific dataset supplied by an individual excluding the publication of said data by the LRC
- Covering all records supplied by an organisation both past and future, excluding publication by the LRC
- Covering all records supplied by an individual or organisation for use in any publication by the LRC
- Covering a specific dataset supplied by an individual or organisation for use in any publication by the LRC

3.3   In most cases, it is preferable to enter into a data supply agreement with an individual for the supply of records, although an agreement with an organisation can be advantageous for other aspects of biological recording.

3.4   Once it is decided which agreement is suitable, this should be filled out by the LRC and dispatched to the recorder including a postage paid or freepost return envelope. The data should be stored temporarily in a "pending" file awaiting completion of the agreement. Once a completed licence agreement is returned, the data can be passed to the computerisation and storage procedures.

3.5   If the supplier refuses to sign the agreement, then the LRC should discuss the supplier's reluctance and attempt to achieve completion. If the supplier still refuses to sign, then the dataset should be securely archived within the "No License Agreed" section and referred to the Storage and Archiving Procedure or returned to the data supplier.

4   Maintenance of Copyright

4.1   On initial receipt, the ownership of data is usually clear unless the supplier is an institution or organisation, however, once the data have entered the LRC system this clarity can easily be lost. Within the licence agreement, it is essential the LRC is able to trace data ownership.

4.2   The most effective method of linking data to their owner is to include all data supply agreements as literary references and to attach the relevant literary reference (licence) to the computer record.

4.3   If the system in 4.2 is used, the list of credits can be generated automatically by listing the literary references.

4.4   Once an analysis is produced this attains its own copyright within the LRC although the original copyright of the underlying data remains with the recorder.

4.5   If records are redetermined it should be stated that the copyright is to remain with the original recorder, although the redeterminer should also be credited. For a more detailed discussion of the aspects of verification, reference should be made to the policy and procedures dealing with the verification of records.

Procedure for Acquisition of Data:
Appendix 1 – Recorder/LRC Data Supply Agreements

The data supply agreement between the LRC and potential data-suppliers should meet the following criteria:

1   It should be easily understandable. "Legalese" should be avoided, especially considering the bulk of signatories will not have a legal background and any document they do not understand is likely to be at best viewed with caution and at worse refused!

2   It should clearly state that the LRC will manage and disseminate the data in accordance with the policy on the release of information and any further restriction placed by the owner.

3   It should clearly state the LRC's expectations from the data supplier.

Data Exchange Agreement for Individual Recorders

1   Introduction

This document outlines the responsibilities of **EcoRecord** and individual recorders within Birmingham and the Black Country in line with the policies and procedures of the National Biodiversity Network (NBN). It is not intended to be a rigorous legal document, but rather a defining of existing unwritten agreements.

2   General

2.1   An individual recorder ("The Recorder") is any naturalist collecting or keeping biological records in their own right and not holding the information on behalf of a natural history society.

2.2   The Recorder gives **EcoRecord** a licence to use the data for the benefit of natural history, where this does not conflict with any of the obligations outlined in the enclosed licence.

2.3   The ownership and copyright of all records will remain with the original Recorder. The Recorder, therefore, retains the right to publish their data without any restrictions from **EcoRecord**.

2.4   **EcoRecord** retains the right to exclude data from the computerised database which does not fulfil the NBN criteria.

3   Benefits of this Agreement

3.1   The natural heritage of Birmingham and the Black Country will benefit from the best possible information being available.

3.2   Both **EcoRecord** and the Recorder will be assured of the best available information for the taxon in question

3.3   The Recorder will be assured that his/her data will be securely stored within the **EcoRecord** and that the data will not be misused.

3.4   The Recorder will be allowed negotiable access to the facilities of **EcoRecord** and access to additional datasets where this does not conflict with other restrictions.

3.5   Support for the Recorder will be available from **EcoRecord** with regard to the procedures and methods of biological recording.

Data Exchange Agreement for Individual Recorders

Agreement Details

1.   With respect to all ecological records recorded by me past and future and passed or copied to **EcoRecord**, I agree to grant a non-exclusive licence in perpetuity to **EcoRecord** to store, process, use and pass on to other users copies of these records, within the policies agreed by the **EcoRecord** Steering Group.

Obligations of EcoRecord

2.   **EcoRecord** will ensure data are managed responsibly in line with the NBN guidelines.

3.   **EcoRecord** will validate and verify records wherever possible and will maintain a list of local and national experts.

4.   **EcoRecord** will respect a recorder's privacy, and will not release information restricted under the Data Protection Act.

5.   **EcoRecord** will respect a recorder's wishes, and will not release confidential information without the permission of the copyright holder.

6.   **EcoRecord** will keep all data supplied in their original form if possible and will not infer information without consulting the recorder.

7.   **EcoRecord** will not publish information including significant amounts of the recorder's data without prior agreement.

8.   **EcoRecord** will supply an annual audit of the Recorder's data if required.

9.   **EcoRecord** will acknowledge the use of significant datasets in any report, unless otherwise agreed.

Obligations of the Recorder

10. The Recorder will ensure he/she is the copyright holder of the supplied records or is authorised to act on behalf of the copyright holder.

11. The Recorder will ensure the data are correct to the best of their ability and supplied to NBN standards wherever possible.

12. The Recorder will notify **EcoRecord** of any redeterminations subsequent to the data supply.

13. The Recorder will notify **EcoRecord** of any change of address to ensure **EcoRecord** complies with the Data Protection Act.

Acceptable Uses of Data Covered by this Licence

14. **EcoRecord** will distribute copies of supplied raw data to the national Biological Records Centre (Monks Wood) and to recognised national recording schemes on request, on the understanding these schemes will credit the source of information in any publication produced.

15. **EcoRecord** will supply analysed data for inclusion in published Environmental Impact Assessments, as long as it does not conflict with any of the above obligations.

16. **EcoRecord** will supply analysed data for use by *bone fide* enquirers, on the understanding credit will be given to the source of the data in any publication or report and as long as it does not conflict with any of the above obligations.

Definitions

17. **Significant Dataset** shall be considered 10% of the report or 5% of the records supplied, whichever is the lesser.

18. **Copyright Holder** shall be recognised as either

    the original producer of the ecological record or

    in the case of records produced during employment, the employer

19. **Raw Data** shall be recognised as the original data supplied by the recorder, including computerised copies where all possible original information is included.

20. **Analysed Data** shall be recognised as either collated data including additional information or any species lists excluding aspects of the raw data *i.e.* personal details or confidential information.

Data Exchange Licence for Individual Recorders

With respect to all ecological records recorded by me past and future and passed or copied to **EcoRecord** I agree to grant a non-exclusive licence in perpetuity to **EcoRecord** to store, process, use and pass on to other users copies of these records, within the policies agreed by the **EcoRecord** Steering Group.

Reference Number

Name

Address

If you wish to be credited when a report uses a significant amount of your information please tick here

Are you prepared to act as a verifier for records supplied to **EcoRecord**?If so, please indicate below which taxa you are prepared to examine:

Taxa for examination

Please indicate any strengths or weaknesses

Do you allow **EcoRecord** to include you name in its list of "experts" for release to other naturalists?

EcoRecord retains the right to request details of expertise before including any name in the "experts" list.

Signed on behalf of the Recorder:

Name Date

Signed on behalf of EcoRecord

Name Position　　　Date

The above information will be held on a computer database, but will not be released without prior consent. This excludes information held in the "experts" list.

Procedure for the Removal of Data

Background to procedure

There will be unfortunate cases when it will be necessary for the LRC to remove data from its databases.

This procedure outlines the steps necessary to the removal of data from the system.

Reasons for removal

There are two major reasons for the removal of data from the database of the LRC:

2.1.1   The data supplier requests the removal of all or part of their supplied data.

2.1.2   The data are discovered to be of dubious quality, which may prejudice the quality of reports produced by the LRC.

Data supplier requested removal

The data supplier must submit the request for removal in writing.

The LRC must attempt to dissuade the data supplier, and if unsuccessful, note the data supplier's reason. If these prove to be due to the policies or actions of the LRC, then steps must be taken to improve the LRC operations to avoid any reoccurrence.

The LRC will remove all the records specified in 3.1 from the computerised database within one month of receipt of the request.

If the supplier also requests the removal of paper records, these shall be returned to the data supplier also within one month unless the LRC requests an extension, *e.g.* for very large datasets.

Once all the required removals have been accomplished, the LRC will notify the data supplier and the data supply agreement will be terminated. Unless otherwise specified, the LRC will retain the data suppliers details on file for reference.

The Removal of Dubious Data

Dubious data are probably the most awkward data to deal with. A distinction must be made between:

- poor identifications, whereby the data supplier should be tactfully advised about improving their identification skills; and
- intentionally fabricated records.

It is only the latter records which should be deleted, the former need to be verified correctly and/or assigned unconfirmed or "wrong" status.

Where false records are suspected, the LRC should request supporting evidence from the data supplier. If the data supplier cannot, or is unwilling to produce evidence, then they must be notified in writing that their records do not reach the necessary standard required by the LRC and are to be removed from the database.

Once removed from the system, the data should be retained in the LRC's archive. A note detailing the problem should be appended to the supplier's data supply agreement.

The supplier should be notified that data of similar quality will only be accepted with supplementary independent confirmation.

The LRC must be very tactful in dealing with this type of problem. It is important that the over-enthusiastic recorder be educated, but not alienated. It should be the very last resort, when datasets are repeatedly shown to have impossible or highly unlikely records, without supporting evidence, which could seriously jeopardise the credibility of the LRC.

18 Documenting data

Policy & Principles

- The LRC should document all its data holdings and manage the associated metadata using an electronic database.

18.1 Background

An LRC should document its data holdings to help it manage them and to enable it to provide information to users about these data holdings.

Many of the raw data that LRCs hold have been collected as part of specific surveys. A group of similar data collected in a similar way is called a data-set. Almost any raw data can be grouped into a data-set of some description (eg casual species records in 1991 received directly from members of the public by Borsetshire BRC or Phase 1 survey of Jim Beam's garden May 2000).

A data-set can be described and catalogued much as a book, film or painting can be. The technical term for the description of a data-set is *metadata* (data about data). In the context of biological recording, metadata can include: the name of the data owner; the dates between which the data were collected; the geographical area from which the data were collected; the terms under which the data are available to third parties.

An LRC needs metadata:

- to give a complete picture of its data holdings so that it can identify gaps in data holdings

- to enable partners and potential users to have a better understanding of what data it holds and how the data can be used, thereby making it more likely that they will use the LRC

- to track ownership and availability of its data holdings and to manage relations with owners (eg the metadata can be used to ensure compliance with data supply agreements)

- to track the status and location of data-sets within the LRC

- to ensure it knows the nature of its data (eg what species, habitats, sites, geographical areas and time periods they cover and how the data were collected)

- to understand how the data can and cannot be used (eg to deliver a product or service) and to ensure users understand any limitations of the data

- so that it can contribute that metadata to the NBN Index

18.2 The NBN Index

Users need to know what data are available, where they are held and how suitable they are for their purpose. The NBN is developing a web-based index to address this. The long-term objective is for the NBN Index to be based on a series of catalogues maintained by organisations for their own purposes. The NBN Index will provide access to these catalogues so that users can find biodiversity information held by a variety of organisations across the UK. More LRCs will be adding their data to the Index in the future. Access to some of the data-sets referred to in the Index will also be available through the NBN Gateway. The Gateway will provide direct internet access to data at varying levels according to the user's access rights.

Metadata are vital to the NBN Index and Gateway. The Index uses a compilation of metadata from many sources (eg LRCs, statutory conservation agencies, national recording societies) and provides a keyword index to them. LRCs can use the Index to publicise their data holdings to a wide audience. As the information in the NBN Index grows, it will become the foremost place in the UK to find out about UK wildlife data-sets. LRCs accredited by the NBN will be expected to contribute metadata to the NBN Index.

18.3 Policy

Every LRC should have a policy on documenting its data holdings. The policy should address the LRC's data management needs, and should ensure that metadata created conform to the NBN Index's standards. This policy should also state that the LRC will contribute metadata to the NBN Index.

Each data-set that the LRC holds should have a metadata record. The definition of what constitutes a data-set needs consideration; some records may need to be grouped to form a single data-set (see example section 18.1).

The LRC may document data on behalf of others if it does not have direct access (eg data held by a local group), but normally it should encourage the data holder to generate the metadata even if the LRC then holds the metadata on their behalf. However, to avoid duplication, the LRC should only do this if the metadata are not otherwise available, either via the organisation which holds them or on the NBN Index.

18.4 Procedures

Procedures should set out how staff will go about documenting data-sets. A data-set should be documented as soon it is acquired by the LRC, regardless of its status (ie whether it has been completely verified and validated) or completeness (ie whether it is complete or still being added to). Information on these matters should be part of the metadata record, and should be updated as the data-set's status changes.

The procedures should also give guidelines to staff on what does and does not constitute a data-set and so merit a metadata record. A discrete data-set is generally formed from data collected as part of a particular survey using a standard methodology. The methods used for validation and data management should also be standard for all data in the data-set. Data collected in an ad hoc fashion (eg observations by different individuals on a range of species groups) may be grouped together as long as the data are managed in a similar manner. Similarly, data collected by a specific recording group (eg local naturalist society or bat group) should usually be considered a separate data-set. Where a data-set has been used to derive a new data-set (eg a new habitat map), the new data-set should be considered entirely separate.

The LRC should use a computerised database to manage metadata. An LRC may create its own database or use standard software (eg Metatagger). Recorder 2000 stores some metadata about the data-sets it holds, although the fields available are currently limited. Consideration is being given to a further module to address this.

The NBN recommends that each metadata entry should ideally include:

- the name of the resource
- a summary description of the contents
- a brief summary of why it was created and/or what its purpose is
- its geographical coverage
- key words to show which species and habitats it covers
- details of how to access the resource
- access conditions
- information on its collection (eg date or period of data collection)

The most important fields, in many respects, may be those which describe the contents of the data-set and the quality of the data. One of the main uses of metadata is to help the LRC and other users of a data-set assess how the data may and may not be used without having to actually look at the data. It is therefore important that the compiler of metadata is honest about these issues.

18.5 Process of developing the policy and procedures

Policies on documenting data should normally be agreed by the LRC's management, especially where they relate to making metadata available through the NBN Index. However, the procedures to implement the policy should normally be developed by LRC staff and agreed by the LRC manager without further consultation.

In setting procedures, the LRC needs first to understand how it will benefit from a metadatabase, by describing how each of the needs listed above relates to its work. It should then choose or design a metadatabase to meet the needs it has identified, while also conforming to the NBN metadata standards.

The procedures should be developed so that they integrate as fully as possible with those which implement the policies on *Data needs*, *Data ownership and acquisition*, *Processing newly acquired data*, and *Storing and archiving of records* (see sections 2, 17, 19 and 21).

18.6 Sources of further information

NBN Index and Gateway:

www.nbn.org.uk

MetaTagger:

www.blue-bag.co.uk

Case study

Generating and managing metadata

Gloucestershire Environmental Data Unit

Background

Gloucestershire Environmental Data Unit (GEDU) was established in 1991 by Gloucestershire Wildlife Trust (GWT). GEDU provides services to GWT, the County Council and some of the district councils, and has a steering group with representatives from GWT, English Nature and the local authorities.

Discussion

The policy states that GEDU will keep a register of all the data it holds and that it will abide by NBN Index standards where appropriate. It also states why GEDU thinks that it is important to maintain metadata.

The procedures state that a metadata record will be created as soon as GEDU acquires a data-set. The information that will be recorded in the metadata record is detailed in the attached annex. Whilst the procedures state that the metadata will be computerised, how this will be done is not described; GEDU is in the process of identifying the most suitable mechanism for electronically storing its metadata information.

Contact

Ian Carle, manager

Gloucestershire Environmental Data Unit, Church House, Standish, Stonehouse, Gloucester GL10 3EU

01453 791338

ianc@gloswild.cix.co.uk

Gloucestershire Environmental Data Unit
Metadata Policy and Procedure

1. Policy Statement

We will keep a register of all data held by the Data Unit. The register will include information necessary for the proper management of the data and for the provision of entries to the National Biodiversity Network Index.

2. Background

Metadata[1] are necessary in order to manage data effectively within the Data Unit and to ensure that confidentiality, copyright and ownership constraints are complied with.

Biodiversity data should be easily accessible (NBN biodiversity data exchange – A framework of principles. Principle 1) and therefore needs to be adequately indexed.

Biodiversity data should be accompanied by metadata that describe the survey methods and quality control used in data collection. Potential users need to know if particular datasets are suitable for their purposes. They also need information related to ownership and constraints on use of the data.

The National Biodiversity Network will rely on the National Biodiversity Network Index to link users with appropriate data. Metadata recorded within the Data Unit should comply with the NBN metadata standard to ensure that information can be passed to the national index.

3. Aims

To manage the data in our care effectively and efficiently

To comply with copyright and ownership constraints attached to data that are managed by the Data Unit

To comply with confidentiality clauses in data transfer agreements

To be able to quickly respond to requests for data or information

To be able to supply enquirers with details of the scope, content and availability of our data holdings

To supply index records to the National Biodiversity Network Index

To be able to audit our data holdings

4. Guidelines

A basic metadata record should be created at the time of data acquisition and more complete index information recorded as soon as possible thereafter. [Ref policies and procedures on processing newly acquired data]

The index information should link to data transfer agreements and copyright agreements [Ref policies and procedures on ownership and acquisition of data]

Version 3.0   Page 1 of 2

Signed   Date:

The metadata record will include the information listed in Annex 1[2] to this policy

The index record will comply with the National Metadata Standard but will include further information intended for internal use and not passed on to the NBN [Ref. www.nbn.org.uk/projects/index]

The index information should be in computerised form

The quality and accuracy of index entries will be assured by appropriate checking and validation

The metadata record should be updated to reflect changes in dataset content and processing

This policy will be reviewed in the light of developing NBN standards

5.   Links to other Policies and Procedures

5.1   This policy is closely associated with:

•   Data Collection Policy and Procedure

•   Data Management Policy

•   Data Update and Maintenance Policy

•   Data Ownership and Confidentiality

•   Controlling Access to Data Policy and Procedure

•   Quality Assurance Policy

6.   Definition of Terms

Metadata: Information about the content of a dataset. This information includes:

•   A simple description of the content of the dataset

•   Information about the origin, ownership and management of the data

•   Reference to the terminological standards used

•   What validation and quality control procedures are used

•   Indexing information

National Biodiversity Network Index: The NBN Index is a searchable catalogue of information sources that are available from NBN partners. Details of the NBN and NBN Index are available from the NBN website.

National Metadata Standard: A controlled set of descriptive terms which describe sources of biodiversity information. See Annex 1. Section 2.

Version 3.0   Page 2 of 2

Signed   Date:

Annex 1
Metadata Record
1. Information for internal use

Register of data suppliers - Link to data transfer and copyright licence agreements

Where data are stored [physical records, archived data, electronic format]

Unique or copy - source

For paper records whether entered onto Recorder

Has the computer data been checked?

Has the metarecord been checked?

Version data

Data exchange – record of data supply [May include flags for provision of data to county recorder]

2. Index information

Required information for NBN Index (from MetaTagger v2.02 program)

- Type of resource (Choices are: Metadata initiative, Dataset, Document, Catalogue, Service, Project, Event, Software, Model, Conference, Discussion Forum)

- Title of resource

- Our Organisation

- Originating Organisation

- Contact at originating organisation

- Abstract

- Inclusion in NBN catalogue (Choices are: Pending, Unknown, Include, Exclude, NBN Only)

- Is it Confidential?

In addition to the above-required information the Data Unit will record the following information (Using standard terms selected from the NBN Dictionary supplied with the MetaTagger program):

- Area covered by the dataset

- Summary information on the Taxa/Biotopes covered

- Coverage

- Availability

(Footnotes)

[1] Metadata: Information relating to data

– see section 6 (definition of terms)

[2] Metadata content derived from April 2000 version of NBN Indexing software.

19 Processing newly acquired data
Policy & Principles

- **Every LRC must have procedures for processing incoming data that enable it to make those data available as quickly as possible.**

- Procedures for processing data must ensure that policies on confidentiality and data ownership are complied with.

- Procedures for processing data must ensure that all possible checks are made on incoming data to find and resolve simple recording errors and errors in transcription.

19.1 Background

The effective flow of data into an LRC is an essential element of its operation. Data enter an LRC from a wide variety of sources and in a range of formats; these data need to be integrated into the LRC's data management systems. Although the LRC should seek to computerise all of its data, this may not always be practicable, and the LRC may have to prioritise its data entry. It is essential that all data are transcribed accurately from their source into whatever data management system(s) the LRC uses.

All data should be tracked so that the LRC knows what stage they are at in the process of being entered into the databases. An LRC should use a metadatabase to maintain information about the status of its data-sets (see section 18 *Documenting data)*.

The LRC should use its data processing procedures to implement as many of its policies as possible—for example, the processes might incorporate stages which ensure that recorders receive acknowledgement that their data have been received, or that confidential records are identified as such. Having effective systems helps give confidence to data owners.

Clearly, one of the key objectives of an LRC should be to process data quickly so that they become available for use as soon as possible.

19.2 Policy

This is a slightly unusual area, as the policy is very closely aligned with its implementation. In general, these policies need little discussion with users and recorders, although the procedures for implementing them must be clear.

The LRC should have policies to ensure that:

- all records are transcribed accurately from the format in which they are received into the LRC's data management systems

- there is a record of every data-set received, and that its status is known at all times

- all possible checks for simple recording errors are made on data on their receipt

- incoming data are processed in a way that enables the LRC to make them available as quickly as possible

- the processing of incoming data-sets is prioritised according to the LRC's need for them

This last policy should take into account both the needs of users, as identified through continual review (see section 2 *Data needs)*, and the current data holdings of the LRC.

19.3 Procedures

The procedure of processing newly acquired data is one of the major day-to-day routines in any LRC. The work is likely to be systematic and repetitive, and much depends on making careful checks at each stage of the process. The checks should be clearly documented, as there may be a number of different people involved in the process. The LRC would benefit from having detailed guidelines for each stage of the process, specific to the LRC's situation (eg specifying where newly acquired data should be stored before they are processed, which acknowledgement letters or slips should be used in what circumstances, where outstanding queries should be stored). This guidance does not address such LRC-specific issues, but looks at the general principles that should inform the different stages of the process.

Procedures should be put in place to implement the following policy areas.

19.3.1 Logging data

All data must be logged as soon as they are received by the LRC. This should happen before the data are assessed or prioritised. Even data which are sent to the LRC but are not required should be logged.

The information recorded at this point should include:

- a summary of the data-set (including, for example, taxonomic group or biotope, geographical and time coverage, number of records)
- the source of the data
- the date the data were received
- the format of the information
- a reference identifying the source of the data

This should be the start of the metadata record for each data-set (see section 18 *Documenting data).*

Systems should be in place to track data back to the original records, particularly when data are received in paper formats such as recording cards or survey reports. This can be achieved through a simple referencing system, in which the same reference is given to all data derived from the same source.

The metadatabase should also be used to track the data through the various stages of processing (eg checking, computerisation, confirmation of agreements with recorders).

Keeping track of data is covered in more detail in section 18 *Documenting data*.

19.3.2 Acknowledging data

The LRC should acknowledge all data-sets it receives. As well as being a good method of thanking recorders, acknowledging receipt also helps avoid confusion or problems when data are misplaced. Recorders should know that they can expect an acknowledgement of data to be sent, and when they can expect to receive it.

For practical purposes, it may not be feasible to acknowledge every record sent in by recorders when these are received in bits and pieces (such as individual records). In such instances, it might be preferable to send an acknowledgement once a certain number of records have been received from the same source. However, when original data sources (such as field notebooks or record cards) are sent and the LRC cannot be completely confident that the recorder has kept copies, an acknowledgement must be sent to reassure the owner that the data have been received.

Acknowledgements can take many forms and may differ for different types of data source. All acknowledgements should confirm that the data have been received and what those data are, and thank the source for sending them in. For recorders who have ongoing agreements with the LRC, this may be all that is required. For recorders who are not familiar with the LRC, it can also provide a means of informing them what will happen to their data (eg how they will be processed and what terms they will be managed under).

19.3.3 Processing data

It is assumed that any LRC needs the data it uses in its day-to-day business to be computerised. Most data that LRCs receive arrive in paper format, and thus need computerising This may include digitising site boundaries or habitat parcels, or computerising species records. Even where data are received in a digital format, they need to be integrated with the LRC's existing data-sets.

Checking for accuracy of transcription. Whether data are received in paper or computerised format, checks should be made to ensure that they are transferred to the LRC's main computer system effectively and accurately. This is normally achieved through a system of random checks for accuracy—which might include printing out batches of records (eg ten per cent of the total) which are then checked against source to ensure that the number and types of errors are within acceptable limits, or checking digitised polygons by having a different operator re-digitising polygons, and then comparing the two sets. The temptation to take short-cuts in this area of work must be resisted, because accuracy checks are crucial to ensuring the quality of the LRC's data.

Checking data for simple recording errors. Data received by the LRC should be checked for obvious recording errors. Checks should include ensuring that: grid references relate to all other locality information (eg sites, vice county information); ensuring that the originator of the data is clearly identified (and addresses given where needed); ensuring that dates are valid; and ensuring where possible that the data are 'probable' (this requires a degree of understanding of the data and may include, for example, looking for records made at unlikely times of the year or for records of species not previously recorded in the area). These checks should be carried out as part of the data entry process. Some checks can be made automatically by computer software such as Recorder 2000. Any discrepancies or mistakes should be referred back to the original source of the data before they are changed by the LRC. Resolving queries made as part of this checking process need not delay computerisation, as records can be flagged as being thought to be incorrect and awaiting confirmation.

Data verification. Data verification is a process of checking that a record is 'correct'—that the species or biotope identification is accurate and from a reputable source. Processes for verifying data are often complex, and may be carried out by the data supplier prior to data being sent to the LRC. Where appropriate, data may be entered into the LRC system before they are verified; in some instances verification processes may be integrated with the main data processing. Further information relating to verification processes for habitat, site and species data are given in sections 4 *Definition of sites*, 5 *Habitat classification* and 6 *Species identification—verification*.

19.3.4 Prioritising incoming data-sets

The LRC should prioritise the processing of new data. It may also identify data at this stage which it will not normally process. This might include historical data-sets. Priorities for processing data-sets may vary between LRCs, and should be driven by core users' needs. Priority should normally be given to more recent data (eg collected within the last three to five years), habitat data and data relating to key taxonomic groups. Individual LRCs should also take into account current projects being run by users, and known gaps in data holdings (whether geographical or taxonomic), when setting priorities.

Criteria should be set to guide the prioritisation process. They should allow the LRC to adjust priorities according to demand, whilst preventing the LRC from becoming diverted from meeting core users' needs when demand from occasional users is high. Criteria might be agreed annually, in collaboration with a users' forum, and should be made widely available, in particular to recorders and recording groups.

19.3.5 Removing data from the system

An LRC must have a process for removing data from its system, so that if the owner of a data-set wishes to withdraw their data from the LRC for any reason, they can do so. This process should include removing any computerised records and returning all paper copies of data to the owner. If the owner does not already have the data in a computerised format, they should be offered a copy of the computerised data in a format chosen by the LRC. A record of the data-sets should be retained by the LRC to record what data the LRC did hold and any means by which the data can still be accessed (see section 17 *Data ownership and acquisition*).

19.4 Linked policies

These policies and procedures are generally uncontroversial, and require little consultation in their development. However, they do implement a range of other policies which should be agreed beforehand; in particular:

- *Data ownership and acquisition* (see section 17)
- *Documenting data* (see section 18)
- *Physical security of data* (see section 20)
- *Storage and archiving of records* (see section 21)
- Relations with *Data providers* (organisations and individuals) (see sections 13 and 14)
- *Minimum record standards* (see section 10)
- Process for identifying data to meet needs (see section 2, *Data needs*)

Case study

Processing data

EcoRecord

Background

EcoRecord is a joint venture between the local authorities and the voluntary sector. Established in 1991, EcoRecord is now wholly financed by local authorities through Service Level Agreements (SLAs). The LRC aims to bring together, in a highly accessible computerised database, all the available information on sites, habitats and species within the Black Country and Birmingham, and to provide analysis and interpretation of that information.

EcoRecord has recently been documenting its policies and procedures.

Discussion

Attached are EcoRecords policy and procedures for processing newly acquired data. These give clear guidance on the different types of format it will accept data in and how they will be processed. Primarily, these are operational procedures for the benefit of the LRC staff, which will ensure consistency in processing data and in validation, and will also ensure that recorders and other data sources are dealt with appropriately.

There is a separate procedure for validating data.

Contact

EcoRecord manager

EcoRecord, 28 Harborne Road, Edgbaston, Birmingham B15 3AA

0121 454 1808

enquiries@ecorecord.org.uk

Policy on Processing Newly Acquired Data
Policy Statement

The LRC will accept ecological data in a variety of formats and process this information efficiently and accurately using agreed procedures.

1 Background to policy

1.1   The Local Records Centre (LRC) needs to convert information acquired from internal or external sources into a form ready for computer entry so that data becomes available for use as quickly as possible.

Initial Receipt of Data

1.2   The LRC is prepared to accept records in the following formats:

- Recorder Import/Export format
- Other computer-readable form including only formats listed in Appendix 1
- Recognised printed recording cards
- Published documents
- Hand-written or typed species lists and reports
- Personal communications by word of mouth or telephone

1.3   The suppliers must adhere to the agreed procedures for ownership, storage of data, access and security (refer to relevant policies and procedures).

1.4   The LRC will maintain a metadatabase of data-sets received from suppliers, each data-set being assigned a unique reference.

1.5   Where possible, the LRC will check records for validity (refer to relevant procedure) prior to acknowledgement.

1.6   The LRC will enter into licensing agreements with all major data suppliers[1] and maintain a register (as per the policy *Data ownership and acquisition*).

1.7   Records should be assessed for their value to the LRC following the agreed procedures with reference to the "Data Entry Priorities" listed in the relevant procedure.

1.8   Receipt of records from external sources should be acknowledged within 10 working days of receipt

1.9   Wherever possible, records will be computerised immediately on receipt, providing they reach the minimum requirements.

1.10   Alternatively records should be stored in a "pending system" based on their priority awaiting data entry.

Data Requirements – Recorder Import Format

1.11   Recorder Import Format data will only be accepted from recognised satellite users and Recorder users registered with the LRC (See Appendix 2).

1.12   All material must be checked using a recognised Virus Checker before being accepted into the LRC computer system.

1.13   The Virus Checking software must be maintained and its virus dictionary updated regularly (at least bi-annually).

1.14   If possible the computer files should be saved to secure storage e.g. CD-ROM

1.15   The supplied data must be checked for validity[2] (refer to relevant procedure), duplication of ID's prior to entry into the LRC's primary database.

Recording Requirements – Other computer-readable

1.16   All material must be checked using a recognised Virus Checker before being accepted into the LRC computer system.

1.17   The Virus Checking software must be maintained and its virus dictionary updated regularly (at least bi-annually).

1.18   If possible the computer files should be saved to secure storage e.g. CD-ROM

1.19   The data should then be analysed in the same way as paper records (refer to the section – *Recording Requirements – On paper*)

Recording Requirements – On paper

1.20   Records should be assessed for their value to the LRC following the policy *Process for identifying data to meet needs.*

1.21   Wherever possible, records will be computerised immediately on receipt, providing they reach the minimum requirements.

1.22   Alternatively records should be stored in a "pending system" based on their priority awaiting data entry.

1.23   Rejected records should be stored in a relevant location (refer to policy on Storage and Archiving) stating reasons for omission.

Recording Requirements – Personal Communications

1.24   All personal communications with data suppliers will be committed to paper and the LRC will consider transcribed telephone information to be owned by the contributor. LRC Staff must ensure sufficient information is received to produce a valid biological record.

Completion of Accessioning

1.25   The metadatabase should be updated to reflect the processing of the data-set

1.26   Records should be stored following the procedures of the Storage and Archiving Policy.

2   Links to other policies

2.1   This policy is closely associated with the following policies:

- Documenting data
- Ownership and data acquisition
- Physical security of data
- Storage and archiving of records
- Relations with data providers
- Minimum data standards
- Process for identifying data to meet needs

3   Procedures

3.1   This policy is implemented through the following procedures:

- Processing newly acquired data
- Validation of newly acquired data

Appendix 1: Formats for computer-readable data acquisition

The LRC will accept data in the following formats on the agreement that the LRC retains the right to reject information or media which might damage the LRC's computer system, e.g. damaged disks or regular occurrences of viruses.

Acceptable Media:

CD-ROM
DVD Disk
Iomega Zip100 Disk
3½ Floppy Disk
5¼ Floppy Disk (only by prior arrangement)
Electronic mail (e-mail) attachment (only by prior arrangement)

Acceptable Formats:

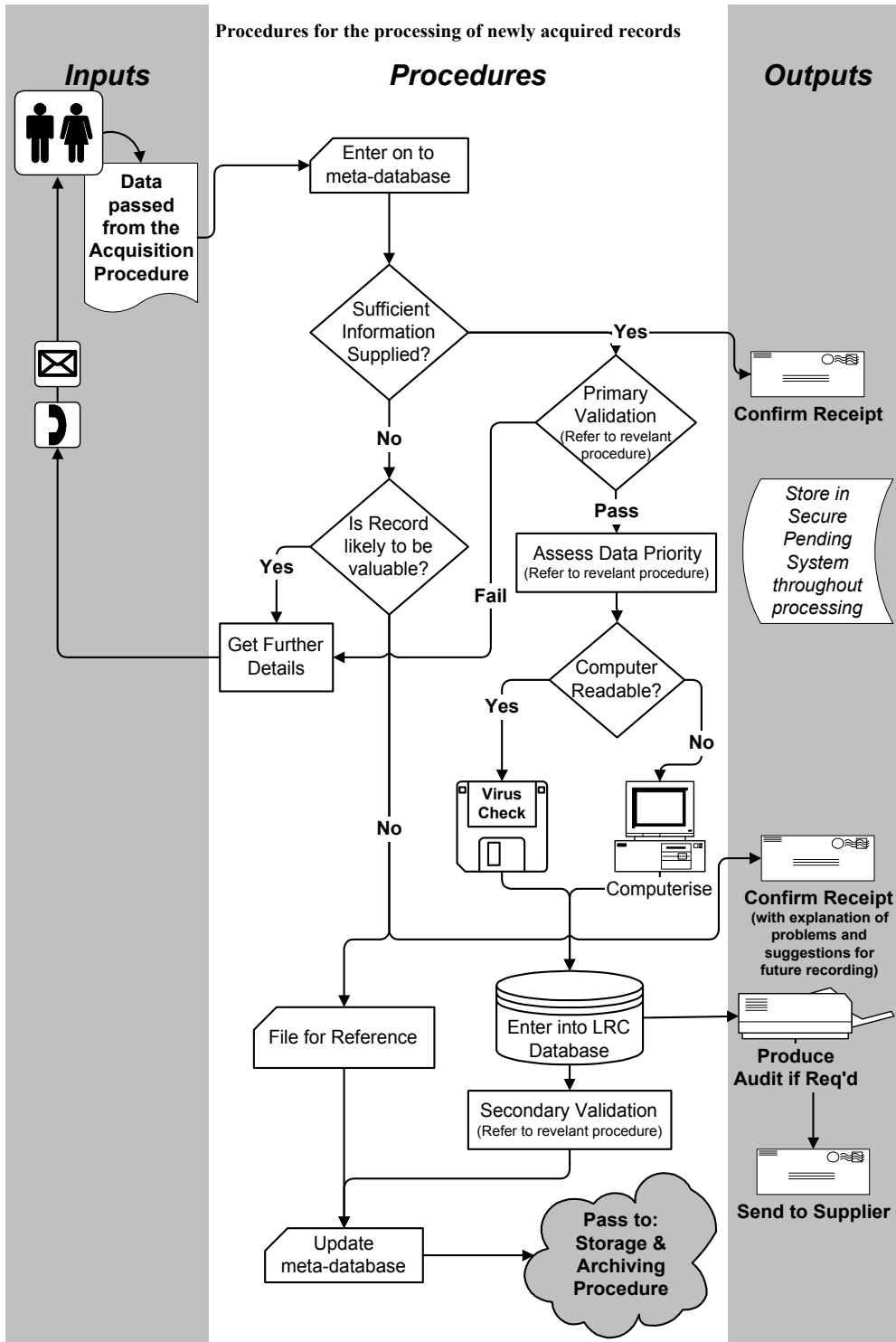|	
GIS:	**MapInfo v4 table** together with any formats supported by this program
Databases:	**Microsoft Access 97** together with any formats supported by this program
Spreadsheets:	**Microsoft Excel 97** together with any formats supported by this program
Text:	Standard **ASCII text**, TAB delimited **Microsoft Word 97** together with any formats supported by this program

NOTE: Information not conforming with any of the above formats will be retained by the LRC, but may not be incorporated into the LRC system unless translation routines are available.

Appendix 2: Registered Recorder Users

A list of Recorder users who have agreed to supply data to a specified level of validation – refer to Ownership Policy

Procedure for Processing Newly Acquired Data

1 Introduction

Procedures for the processing of newly acquired records

Inputs	Procedures	Outputs

- **Data passed from the Acquisition Procedure**
- Enter on to meta-database
- Sufficient Information Supplied?
  - **Yes** → Primary Validation (Refer to revelant procedure)
    - **Pass** → Assess Data Priority (Refer to revelant procedure)
    - **Fail** → Get Further Details
  - **No** → Is Record likely to be valuable?
    - **Yes** → Get Further Details
    - **No** → File for Reference
- **Confirm Receipt**
- *Store in Secure Pending System throughout processing*
- Computer Readable?
  - **Yes** → Virus Check
  - **No** → Computerise
- **Confirm Receipt** (with explanation of problems and suggestions for future recording)
- Enter into LRC Database
- Secondary Validation (Refer to revelant procedure)
- **Produce Audit if Req'd**
- **Send to Supplier**
- File for Reference
- Update meta-database
- **Pass to: Storage & Archiving Procedure**

The following document is intended to guide LRC staff through the procedures for processing newly acquired data in accordance with the accompanying "Policy on Processing Newly Acquired Data" and follows on from the Procedure for Acquisition of Data.

2   Initial Receipt of Data

2.1   Enter the details of the received data-set into the LRC's catalogue of received data-sets (metadatabase, for suggested contents refer to Appendix 2).

2.2   The LRC is prepared to accept records in the following formats:

2.2.1   Recorder Import/Export format

> **Section 3**

2.2.2   Other computer-readable form including only formats listed in Appendix 3

> **Section 4**

2.2.3   Computer readable form excluding formats listed in Appendix 3

> **Section 5**

2.2.4   Recognised printed recording cards

> **Section 6**

2.2.5   Published documents

> **Section 6**

2.2.6   Hand-written or typed species lists and reports

> **Section 6**

2.2.7   Personal communications by word of mouth or telephone

> **Section 7**

2.3   The data should be validated, following the ***Procedure on Validating Newly Acquired Data***

2.4   The receipt of records from external sources should be acknowledged within 10 working days of receipt. This can be either via a personal letter or using a pre-printed card (see example in Appendix 6), dependent upon the volume of acquisitions.

2.5   If insufficient information for a valid biological record has been supplied or there is a problem with the data following validation, then the supplier must be contacted by letter rather than the pre-printed card to request clarification or to educate on the requirements of biological recording. This must be done in a tactful manner, so as not to alienate a potentially valuable data supplier.

3   Recording Requirements – Recorder Import Format

3.1 Prior to any processing the supply medium (disk, CD-ROM or e-mail) should be checked for any known viruses and once clear backed-up to at least one additional location, preferably to be stored securely off-site.

3.2 If the supply medium is found to be infected with a virus, the media should be disinfected if possible or securely stored with a warning to avoid infection of the LRC's computers. In either case the supplier must be immediately contacted to inform them of the infection on their computer. Further data-sets will not be accepted from the supplier until they confirm elimination of the virus from their computer.

3.3   Recorder2000 or compatible XML data

3.3.1   Is the supplier listed on the list of recognised Recorder Suppliers. If not, the LRC must liase with the recorder to ensure they comply with the LRC's standards on the usage of Recorder.

3.3.2   The rest of this section of the procedure cannot be produced until further details of Recorder2000/XML are available.

3.4   Recorder based on ARev

3.4.1   Is the supplier listed on the list of recognised Recorder Suppliers (see Appendix 2 of the Policy)? If not, the LRC must liase with the recorder to ensure they comply with the LRC's standards on the usage of Recorder.

3.4.2   Load the data on to the slave copy of Recorder (refer to Appendix 4) to check for record validity, ID duplication and any other formatting problems with the data. If there are serious problems, the LRC must liase with the supplier to rectify the problem, the data remaining as pending until the LRC's record standards are achieved.

3.4.3   Once the recordset's validity is confirmed, the primary database should be securely backed-up and the new data imported.

3.4.4   The primary database should then be checked for any outstanding problems with the newly imported data.

3.4.5   Once the data-set is validated, the Received Data Catalogue should be updated to confirm computerisation, the database should again be backed-up (but not destroying the back-up produced at 3.4.3) and the raw data passed to the procedure for storage and archiving.

4   Recording Requirements – Other computer-readable

4.1   Prior to any processing the supply media (disk, CD-ROM or e-mail) should be checked for any known viruses and once clear backed-up to at least one additional location, preferably to be stored securely off-site.

4.2   If the supply medium is found to be infected with a virus, the media should be disinfected if possible or securely stored with a warning to avoid infection of the LRC's computers. In either case, the supplier must be immediately contacted to inform them of the infection on their computer. Further data-sets will not be accepted from the supplier until they confirm elimination of the virus from their computer.

4.3   If the LRC has the expertise to convert the data into a form compatible with Recorder and the data-set is large enough to make the additional processing worthwhile this should be undertaken, then the procedure should continue at section 3.4.2

4.4   The data should be printed out ready for re-keying. If reformatting is necessary the integrity of the included data is not compromised. The computerised data should then be passed to the Storage and Archiving Procedure, whilst the printout is handled as "records supplied on paper" – refer to section 6.

5   Recording Requirements – Incompatible Computer Format

5.1   If data are supplied in a format incompatible with the equipment at the LRC, it will first be necessary to contact the supplier to discuss the following:

5.1.1   Can the supplier produce the data-set in an alternative (compatible format), if not …

5.1.2   What information is held within the data-set, *i.e.* is it worth the effort to get the data converted by external contract or is the format one the LRC should support and therefore requiring additional hardware or software within the LRC.

5.1.3   If neither of the above is true, the data must be archived in the "incompatible data" section and referred to the Storage and Archiving Procedure.

5.2   Discussion should be held with the supplier to agree formats to be used in future transactions.

6   Recording Requirements – On paper

6.1   The data-set should be assessed for validity, *i.e.* do the records reach the NBN minimum standard of a biological record.

6.1.1   If not, are the records valuable enough to make further discussion with the supplier worthwhile?

6.1.2   The LRC should discuss with the supplier its minimum requirements as defined in the Licence Agreement.

6.2   Records should be assessed for their value to the LRC following the agreed procedures with reference to the "Data Entry Priorities" listed in Appendix 5. This will categorise records into three types:

6.2.1   High Priority

6.2.2   Medium Priority

6.2.3   Low Priority

6.3   Dependent on the LRC's current workload, records will be computerised immediately on receipt, providing they reach the minimum requirements following 6.1

6.4   However, the likelihood is that this will not be possible. In this case, the prioritised data-sets should be stored in a pending system based upon the priority and high priority data should be entered first.

6.5   Rejected records should be stored in the "Rejected Records" section explaining the reason for omission and referred to the Storage and Archiving Procedure.

7   Recording Requirements – Personal Communications

7.1   LRC Staff must ensure sufficient information is received to produce a valid biological record. The most efficient method is to have pre-printed forms (see Appendix 6) next to the telephone which prompt the LRC staff to request sufficient information from the caller.

7.2   In most cases the caller will not be a registered recorder and it is unlikely that a licence will be necessary, however, it is essential that LRC staff inform the caller of the proposed use if their record(s).

7.3   If the caller wishes to become a licensed recorder, refer to Appendix 1 and then the record(s) should be handled as "records supplied on paper" and processing should continue at section 6.

Appendix 1: The Licensing of Recorders with the LRC

It is important that the LRC complies with both the Copyright Act and the Data Protection Act and to this end, it is advisable for the LRC to enter into licensing agreements with prospective suppliers of biological information. This is covered in detail within the Policy and Procedures for Ownership and Acquisition of Records.

A sample licence agreement is shown overleaf.

Data Exchange Licence for Individual Recorders

With respect to all ecological records recorded by me past and future and passed or copied to EcoRecord I agree to grant a non-exclusive licence in perpetuity to EcoRecord to store, process, use and pass on to other users copies of these records, within the policies agreed by the EcoRecord Steering Group.

Reference Number

Name

Address

If you wish to be credited when a report uses a significant amount of your information please tick here

Are you prepared to act as a verifier for records supplied to EcoRecord?If so, please indicate below which taxa you are prepared to examine:

Taxa for examination

Please indicate any strengths or weaknesses

Do you allow EcoRecord to include you name in its list of "experts" for release to other naturalists?

EcoRecord retains the right to request details of expertise before including any name in the "experts" list.

Signed on behalf of the Recorder:

Name Date

Signed on behalf of EcoRecord

Name Position   Date

The above information will be held on a computer database, but will not be released without prior consent. This excludes information held in the "experts" list.

Recorder's Copy *Please retain pages 1-3 and return page 4 in the enclosed freepost envelope*

Appendix 2: Contents for the Data Received Metadatabase

The metadatabase for registering the receipt of data-sets should include the following information, but can be operated as either a spreadsheet or true database:

Unique Reference	An agreed system for uniquely identifying each received data-set
Date of receipt	A date format field
Supplier address	Either a link to an address book, list of licensed recorders or full name and
Acknowledgement sent	Either Yes/No or date sent
Taxa covered	Free text list of taxa
Quantity of data	Free text estimate of number of records
Sites covered	Free text description of sites covered
Format supplied	Standard list of formats, *e.g.* P (paper) R (Recorder) A (MSAccess) X (MSExcel), E (E-mail), etc.
Location within LRC	Standard list of locations within LRC, may be free text
Species entered & date	At least a Yes/No, but could also include date of entry on to computer and first and last ID number on the computer database
Site boundary entered	At least a Yes/No, but could include date of digitising, precision, etc.
LRC staff involved	Could include any member of staff involved in processing the data, although original receiver, data-enterer and final archivist are the most important

Appendix 3: Formats for computer-readable data acquisition

The LRC will accept data in the following formats on the agreement that the LRC retains the right to reject information or media which might damage the LRC's computer system, e.g. damaged disks or regular occurrences of viruses.

Acceptable Media:

CD-ROM
DVD Disk
Iomega Zip100 Disk
3½ Floppy Disk
5¼ Floppy Disk (only by prior arrangement)
Electronic mail (e-mail) attachment (only by prior arrangement)

Acceptable Formats:

GIS:	**MapInfo v4 table** together with any formats supported by this program
Databases:	**Microsoft Access 97** together with any formats supported by this program
Spreadsheets:	**Microsoft Excel 97** together with any formats supported by this program
Text:	Standard **ASCII text**, TAB delimited **Microsoft Word 97** together with any formats supported by this program

NOTE: Information not conforming with any of the above formats will be retained by the LRC, but may not be incorporated into the LRC system unless translation routines are available.

Appendix 4: The process for Validation of Recorder Data

1.   When data are supplied in Recorder Import/Export format, these data cannot be viewed prior to import. As a result it is difficult to validate these data prior to entry into the database. The following procedure is a method which enables the validation of data prior to its incorporation into the LRC's primary data-set.

2.   The exact process is dependant upon the number of Recorder Licences held by the LRC:

One licence only:

The LRC should run two copies of Recorder **on the same computer**. One as the primary LRC database and one as an empty slave copy.

The simplest method is to install twice on to separate partitions or hard-drives within the computer, although this produces duplication of certain directories and more efficient systems are possible – contact Recorder Support for further details.

Greater than one licence but no free copy

The procedure in 4.2.3 should be followed on one computer in the office.

Greater than one licence and a copy not currently in use

The vacant copy of Recorder should be installed on a computer which does not hold the primary LRC database as a slave copy.

3.   The acquired data should be imported on to the slave copy of Recorder. This data can then be checked for validity (refer to the relevant procedure) and duplication. This data-set should not be altered, but rather notes should be made regarding any problems.

4.   If the problems are too great or numerous, the supplier should be notified and requested to resubmit the data after suitable editing.

5.   Once the data has been validated, the original data-set can be imported into the LRC's primary database (after backing-up procedures have been completed). The copy from the slave copy must not be imported because the Recorder User Number will be incorrect and would report the data was supplied by the LRC rather than the external recorder.

6.   Once the data-set is successfully imported into the primary database, the contents of the slave copy should be deleted.

Appendix 5: Efficient incorporation of telephone records

To facilitate efficient transcribing of records received by telephone, a "Single Species Record" card should be used. A proposed minimum amount of information is included on the example below:

SINGLE SPECIES RECORD CARD (CASUAL RECORDS)									
Date of Sighting									
Name of Recorder									
Address									
Telephone:									
Name of location									
Grid Reference (six figure if possible)	S								
Species									
Number seen									
Observations / other information									
LRC Staff Inits & date									

Based on the recording card of Somerset Environmental Records Centre (SERC)

Appendix 6: An example pre-printed Acknowledgement Card

(Footnotes)

[1] "Major Data Supplier" covers all data-sets other than occasional and "one-off" records.

[2] Validity in this context means the data should contain sufficient information for a biological record (What, Where, When and by Whom), it does not include ensuring the species is correctly identified, this is covered by verification in a separate policy.

20 Physical security of data
Policy & Principles

- The LRC should protect the data in its care from alteration, damage, loss and unauthorised use.

20.1 Background

Ideally, data should be managed at one agreed location and be made available to a wide range of users. It is the data custodian who is responsible for managing the data and who acts as the definitive source of data; only the custodian may make changes to data. LRCs should act as the custodian for much of the data they hold, and should ensure that data are held and supplied in accordance with NBN standards and other relevant agreements. Being a data custodian brings with it the responsibility for managing data properly and ensuring that they are physically protected.

The replacement of much of the data held by an LRC is likely to be extremely costly or even impossible (eg in the case of original records); so the data must be protected against loss or damage. Data owners and users also need to be sure that the LRC is protecting the data in its care; indeed the LRC has an obligation, to data owners in particular, to do this.

The LRC should be aware of a number of risks to data:

- alterations or deletions of original data (paper records and computerised data)
- physical damage (eg by fire, flood, wear and tear, and vandalism) to original data (paper records and computerised data)
- corruption of computerised data
- theft, or unauthorised use or copying, of paper and computerised records

Note that this policy deals with the physical aspects of data security. The unauthorised use of information supplied to users (ie users failing to comply with the terms under which information is supplied) is covered in section 30 *Controlling access to data*.

Policy on providing for the physical security of data should say why and how the LRC will look after the data in its care.

20.2 Policy

Original data are irreplaceable, whereas the ease and expense of replacing other records vary widely, depending on a number of factors such as: whether they have been computerised and their transcription validated; whether their accuracy has been verified; and what further processing is required to make them useful.

In order to achieve the general aim of protecting its data holdings, the LRC needs to prevent: alterations or deletions of data; physical damage to data; loss of original paper records; and corruption of computerised data.

The data supply agreements that the LRC has with its suppliers should guarantee that their data are well cared for. Suppliers need to be assured that the LRC will prevent any access to their data outside the terms of their agreement with the LRC. They will also want to know that they can trust the LRC to look after their original records, and not to lose copies of their records and have to come back to them for the same information.

Policy on the physical security of data should set out how the risk of unauthorised access to data will be minimised, specifically addressing risks of unauthorised use or copying of paper records and computerised data within the LRC. Achieving this aim means addressing all of the risks listed in section 20.1 *Background*. This policy should make clear links to the LRC's policy on *Storing and archiving of records* (section 21) and on *Controlling access to data* (section 30). The latter policy should focus on how data will be passed out by the LRC in a way that limits any potential for damage to wildlife.

20.3 Procedures

The procedures should be based on the LRC's assessment of the risks to its data holdings. The LRC should review the list of risks given in 20.1 *Background,* above, and describe them in more detail in relation to its own particular situation. It should then undertake a risk analysis: for each identified risk it should assess:

- the likelihood of the risk occurring
- the likely result, should the risk occur
- the overall significance of the risk to the LRC, based on the two points above

The analysis should list the risks in order of their overall significance and should be included in the LRC's policies and procedures manual.

There are several mechanisms that the LRC can use to minimise these risks:

20.3.1 Physical mechanisms

- locks on windows, doors, filing cabinets
- fireproof map and filing cabinets
- security cables on computers
- fire and burglar alarms
- sprinkler systems

20.3.2 Procedural mechanisms

- designating and training holders of keys and codes to alarm systems
- managing access to IT systems through the use of passwords (including screensavers with passwords)
- managing visitor and volunteer access to offices
- managing movement of original records and management data (for example, agreements and records of enquiries)
- establishing and disseminating emergency procedures
- protecting IT systems from infection by computer viruses, whether via the internet or via storage media

Focusing on the most significant risks, the LRC should identify mechanisms to minimise the risks, and describe how these mechanisms will be implemented through its procedures. The procedures should explain what changes the LRC will make to its offices and information management systems, and what training it will give its staff. They should also explain what day-to-day working practices it will adopt to reduce risks to its data holdings.

20.4 Process of developing the policy and procedures

Because the policy and procedures relating to physical security are largely practical in nature, the staff of the LRC should develop them, particularly those who will be implementing them on a day-to-day basis. They will be easy to draw together, as long as they are not overly prescriptive.

Case study

Data security

Gloucestershire Environmental Data Unit

Background

Gloucestershire Environmental Data Unit (GEDU) was established in 1991 by Gloucestershire Wildlife Trust (GWT). GEDU provides services to GWT, the County Council and some of the district councils, and has a steering group with representatives from GWT, English Nature, the Environment Agency and the local authorities.

The LRC is housed in a rural location within a well-secured building, which also hosts GWTs consultancy and some other GWT staff. The main GWT building is found elsewhere—because of this, and because of its location, GEDU tends not to get many visits from members of the public.

Discussion

The policy statement commits GEDU to ensuring that data are protected from physical damage and unauthorised use. This is followed by a set of aims which the policy and procedures should secure. The procedures set out to address a range of risks to data such as theft, flood, inappropriate access to confidential data, and corruption of data files. Future security measures are also described; for instance, GEDU is considering the use of fireproof filing cabinets, although extra resources will need to be obtained before these can be purchased.

Contact

Ian Carle, manager

Gloucestershire Environmental Data Unit, Church House, Standish, Stonehouse, Gloucester GL10 3EU

01453 791338

ianc@gloswild.cix.co.uk

GEDU PHYSICAL SECURITY OF DATA POLICY AND PROCEDURE

1 Policy Statement

GEDU will ensure that the data (both paper and digital) in its care are safeguarded from physical loss or deterioration and from unauthorised use or alteration.

2 Background

The information held by GEDU represents an irreplaceable resource and must be protected against loss or damage.

In addition GEDU has a moral duty to those individuals and organisations that have supplied information to GEDU to manage the information in a responsible way.

3 Aims

To protect the long term integrity of the information in the care of GEDU

To prevent physical damage to data held by the Data Unit

To prevent theft of data

To prevent unauthorised use of data

To comply with confidentiality clauses in data transfer agreements

To ensure that information referring to vulnerable species, habitats, geological or geomorphological features is properly controlled and not open to misuse

To comply with the requirements of the Data Protection Act

4 Guidelines/Procedures

GEDU will take measures to ensure that the data in its care are not subject to physical damage or loss. The Data Unit will take precautions against the following identified threats - theft, flood, fire and from unauthorised use or alteration.

- The GEDU office has a number of measures designed to stop theft from the premises – the building has an alarm that is set every night, the windows have lockable security grills that are locked and the inner front door has a keypad to prevent unauthorised access during office hours. In addition to reducing the risk of theft these measures severely reduce the risk of unauthorised use or alteration of data.

- To reduce the risk of flood damage to data, Data Unit staff do not keep important documents stored in boxes on the floor.

- To reduce the risk of loss of paper data in a fire or flood event the Data Unit will undertake a data archiving programme where all unique paper data sets will be copied and stored off site. In addition the Data Unit will investigate the feasibility of using fireproof cabinets to store unique data sets.

- All site/species folders will include a list of contents for document control purposes. To prevent inappropriate information being given out by Data Unit staff, folders containing confidential information will be marked as such on the outside. Confidential records are also flagged on the Recorder database. Original documents should not be removed from the Data Unit. However where appropriate, copies may be made.

GEDU takes the following measures to protect its computerised data:

- Direct access to data held on the Recorder database is only available through authorised users whose accounts are protected by passwords. There is no direct public access to the Recorder database.

- A back up copy of Recorder is made to the file server after every input session. A further copy is made every night as part of the Data Units tape back up routine. To ensure that the back up tapes are not lost during a catastrophic event at the GEDU offices they are stored off the premises.

- All incoming computer data (including e-mail, CD's, Floppy Disks etc) is checked for computer viruses by an up to date anti virus package.

- Access to the internet is at present only available on one GEDU PC. The internet is accessed via Dial Up Networking for short periods of time reducing the risk of hacking. If GEDUs use of the internet changes then GEDU will re examine the potential threat from hacking and take appropriate measures to minimise the risk.

Version 2.0

signed: ……………………………………………..   date: ……………………………….

21 Storing and archiving of records

Policy & Principles

- The permanent and secure storage of original records by an LRC is a service to the owner of the records that also provides advantages to the LRC.

- Archived records must be properly protected against loss, damage and alteration.

- An LRC may need access to original records in order to validate data that have been entered into computerised databases, or to check for additional information contained in the original records, or to support further verification of data.

21.1 Background

The permanent storing and archiving of original biological records is a service that most LRCs should offer. It provides advantages to the owner of the records and to the LRC, and protects the records themselves.

An LRC is often better placed to provide secure facilities for the permanent archiving of original records than the owner of the records, particularly when the owner is an individual recorder. All recording media degrade over time, and may be subject to loss, damage or alteration. As professional bodies, LRCs are well placed to provide the infrastructure, facilities, policies and procedures to archive recorders' original records securely.

Storing original records has a number of advantages for LRCs, which may often need to access original records at short notice. Data validation and verification processes often take place over considerable periods of time, and errors may be discovered years in the future. Having easy access to original records makes validating data simpler and allows inputting of new data to be scheduled in advance. Data may need to be 'redetermined'—that is, to go through a process of further verification as a result of doubts being cast on the original verification process or on the accuracy of the recorder's records, or as a result of changes to a species' taxonomy. In such cases, further verification of data may require inspection of the original records.

Original records, particularly those on paper, often contain additional information that may be of interest to an LRC's users; such information may include notes of a recorder's opinions or supplementary observations that were not transferred to the LRC's database during data input.

21.2 Policy

An LRC should set its policy on:

- the service the LRC will provide to data owners through protecting their records

- the provision it will make for storing and permanently archiving the original records submitted to it

- the types of media and format of computerised data for which it will provide secure storage

- the conditions of access to archived material

Providing access to a recorder's original data should be controlled in much the same way as providing access to data that have been entered into the LRC's data management systems, in order to protect 'sensitive' information (see section 30 *Controlling access to data*). However, the issue of preventing loss, damage and alteration to these records should be considered.

21.3 Procedures

21.3.1 Data input and validation

All LRCs need procedures to deal with the management of original records during data inputting and validation and prior to archiving and storing them. It is essential that filing and storage systems permit easy access while data are being entered onto the LRC's database. Following validation, original records may have the data inputter, data validator and relevant database references appended to them. The process of permanently archiving the original records should commence after the initial validation of data.

21.3.2 Physical security of records

It is important that archived records are physically secured (ie that they are protected from loss, damage and alteration).

Archived original records must be protected, just as data that are being used and actively managed by an LRC must, through procedures to ensure their physical security (see section 20 *Physical security of data*). However, there are additional factors to be considered when designing procedures to permanently archive original records. Archived data may be accessed infrequently, and any damage or degradation may therefore go unnoticed for a considerable time.

21.3.3 Original records

It is important that original records are stored on recording media (for example, paper, floppy disks, magnetic tapes, hard drives, Zip disks) that are reliable and well maintained, and that storage is monitored. Individual procedures are needed to deal with the archiving and cataloguing of records held on each of the different media that the LRC will use for archiving.

Records may arrive at an LRC from different sources and in different forms. For example, data may arrive on paper record cards, as hand-written lists, as Recorder Export files, in Excel spreadsheets or via telephone messages. It is important that the LRC is able to store these original data permanently and securely whilst also enabling them to be accessed.

Procedures are needed for handling any original records submitted in a form the LRC does not archive. Options include: returning records to the owner, transferring records to a new recording medium, or translating records into a new computer format (see section 21.3.4).

21.3.4 New formats and media

Sometimes the archiving and provision for retrieval of data contained in original records is best achieved through storing these data on a new medium or in a different computer format. For example, if original paper records must be stored off-site due to space restrictions, they may most easily be retrieved from images stored on CD-ROMs; or, if data have been supplied on floppy disks, it may be best to copy these to CD-ROM or onto a hard drive.

Original records supplied in databases or spreadsheets created in unusual software formats may best be reformatted in more standard software packages prior to archiving.

In all cases, the LRC should also retain records in their original form and take steps to ensure that no translation errors occur when copying to the new medium or translating to a new computer format.

21.3.5 Physical security of data

Copies of original records, or of the data they contain, must be made as insurance against accidental loss or damage to originals. Losses may result from fire or theft, so at least one copy must be reasonably mobile so that duplicates can be stored separately (see section 20 *Physical security of data)*.

All recording media are subject to eventual degradation over time. Some are particularly sensitive to damage from heat and humidity. In the case of original records recorded on particularly degradable media, LRCs must ensure that copies are made onto new, more durable media. Even supposedly permanent media do not last for ever, and an LRC must therefore ensure that copies are made regularly enough to prevent any loss or damage to data. All storage environments must be maintained according to the advice given by the manufacturers of storage media with respect to heat, humidity, light and radiation.

21.3.6 Providing access to archived records

It is important that archived data are available for consultation. A catalogue should be maintained of all original data-sets that are stored and archived by the LRC. Since the only purpose of the catalogue is to manage information on original records, it should be separate from other catalogues of metadata covering actively managed data-sets (see section 18 *Documenting data)*.

Procedures to control the release of original records that the LRC has archived should be put in place. These should be based on procedures on controlling access to data (see section 30) and should also take into account any measures needed to prevent their loss, damage and alteration. LRCs may wish to prevent original records from being removed from the premises.

The catalogue of archived records should list the file locations of original, duplicate and transcribed records (those copied to new media or alternative computer formats). The most common catalogue references are likely to be to archived paper and computer files. If a file is removed from storage for consultation, an appropriate 'document removed' slip should be completed and left in its place until it is returned.

It is best to enter the archived record's file reference in every data entry in the LRC's database, to allow for easy reference to original records. It is also useful to append information on validation and verification, and any other relevant data input references, to the original record prior to archiving.

21.4 Process of developing the policy and procedures

Securely storing and referencing all original records requires significant resources, particularly staff time. Setting the LRC's policy on storing and archiving original records requires the agreement and support of the LRC's partners.

The LRC should consult data suppliers through their representatives or through a forum when developing policy and procedures as storing and archiving are services to them as well as having advantages for the LRC.

These procedures are technical in nature, and developing them requires knowledge of data storage media, software formats, the LRC's cataloguing and indexing systems and other data management procedures. They should therefore be developed and implemented by LRC staff, particularly those with IT responsibility.

Case study
Storing and archiving of records
EcoRecord
Background

EcoRecord is the LRC for Birmingham and the Black Country. It developed over a number of years, beginning in 1984 when a nature conservation strategy was produced for the county of the West Midlands. In 1986, the old West Midlands County Council established a 'Wildlife Records Centre, but this was put on hold when the County Council was abolished. It was reborn as 'EcoRecord—the ecological database for the Black Country and Birmingham in 1991, as a result of a partnership between the Urban Wildlife Trust (now the Wildlife Trust for Birmingham and the Black Country), English Nature and the local authorities of Birmingham, Dudley, Sandwell, Walsall and Wolverhampton.

Initially, EcoRecord was run by the Wildlife Trust and the Joint Data Team (JDT) on behalf of these five local authorities. The JDT had been a department of the County Council, and continued as an independent unit when the County Council was abolished. The JDT was later sold to Mott Macdonald, which now, together with the Wildlife Trust, runs EcoRecord, with the local authorities providing funding through four year Service Level Agreements (SLAs).

EcoRecord has recently been documenting its policies and procedures on storing and archiving records and these follow the discussion.

Discussion

A single broad policy statement sets out EcoRecords decision to store and archive all 'raw data it receives. Here 'raw data are defined as information received prior to processing—corresponding to the term 'original records as used in the guidance. LRCs may need an additional short policy on its conditions of access to original records, or this issue may be considered in separate LRC policies on access to data (see also section 30 *Controlling access to data*).

Most of the issues covered in the guidance are dealt with in EcoRecords policy under the 'Background to the policy and 'Procedures sections. The 'Background section adds detail to the policy statement, including a list of media and computer formats that EcoRecord undertakes to archive. The range of acceptable formats is fairly comprehensive, although some LRCs may wish to extend the range of computer-readable records they will archive.

Procedures are also introduced in the 'Background section and are elaborated under the 'Procedures section.

This discussion will follow the headings used by EcoRecord, which refer to the different media on which original records are received.

Original records supplied in computer-readable formats

EcoRecord undertakes to archive at least one version of each original record in the computer format in which it is received, although duplicates may be made in a more convenient format (format refers to software application).

Original records supplied on paper

In addition to archiving and cataloguing original paper records, EcoRecord plans to scan them onto digital media, but this requires improvements in technology and investment in a high-quality automated scanner. As an interim measure, EcoRecord usually photocopies to generate duplicates. LRCs may view photocopies as sufficient for the purpose of generating duplicates of paper records and, depending on the circumstances, the advantages of an automated digital system may not justify the cost. Photocopies of paper records have the advantage of requiring the same catalogue entry (other than for location) as the originals.

EcoRecord maintains a separate 'library catalogue of references to printed material that relate to records held. This is an important class of records for many LRCs, although it is only relevant to policies on storing and archiving records when the original printed source material is held by the LRC.

Personal communications

EcoRecord does not produce a separate record of personal communications. Other LRCs may wish to do so and to catalogue biological data supplied in this way, as is done for data received in other forms.

Physical security

EcoRecord undertakes to ensure that the physical conditions of storage are suitable (referred to as 'fit for purpose), noting that confidential records are afforded additional protection. Confidential records are probably at more risk of inappropriate release or misuse when held as actively managed data within an LRCs database, although LRCs may wish to afford additional security to all particularly valuable original records. In addition to confidential records, valuable records may include collections of a recorders lifetimes output or records containing considerable amounts of additional information not captured by the LRCs database.

Contact

EcoRecord manager

EcoRecord, 28 Harborne Road, Edgbaston, Birmingham B15 3AA

0121 454 1808

enquiries@ecorecord.org.uk

Policy on storage and archiving of data

1   Policy statement

The LRC will store all raw data received, both digital and on paper, in a secure and accessible system.

2   Background to policy

2.1   The LRC needs to store and archive raw data.

2.2   In this context, raw data is the information supplied to the LRC prior to any processing within the LRC.

Recording formats

2.3   The LRC must be capable of archiving records in all of the following formats:

- Recorder Import/Export format

- Other computer-readable form including only those formats listed in Appendix 1

- Recognised printed recording cards

- Published documents

- Hand-written or typed species lists and reports

- Personal communications by word of mouth or telephone

Storage requirements − computer-readable formats

2.4   If facilities are available these data should be saved to CD-ROM or DVD for secure archival storage; otherwise they should be saved to magnetic media and a paper copy should be produced and stored as in the section *Storage requirements – on paper*.

2.5   A metadatabase cataloguing the contents of archived CD-ROMs or DVDs should be maintained. This should be either part of a corporate metadatabase of all data-sets, or an automatic computer-generated catalogue.

Storage requirements − on paper

2.6   All paper records will be stored in a secure filing system, allowing simple retrieval of said data if required.

2.7   If facilities are available, all data sheets should be scanned into computer and archived to CD-ROM, and a metadatabase cataloguing the contents of archived CD-ROMs or DVDs should be maintained.

2.8   References to published documents, articles or books will be maintained in a library catalogue, either on computer or on paper, including the book's location. If the LRC owns a copy of the document, it will be included in the LRC's library.

Storage requirements − personal communications

2.9   The storage of these is not applicable – only the paper copy produced by LRC staff can be archived and should be processed as a paper record.

Storage requirements − rejected records

2.10   If records fail to meet the minimum requirements for a biological record (refer to the detailed procedure) these data will be stored within the archiving system including the reasons for omission from the primary data-set.

Storage requirements − general

2.11   The LRC will ensure the storage location is 'fit for purpose' and will avoid locations which could seriously degrade the stored information.

2.12   The LRC will ensure additional security is afforded to the storage of confidential records.

3   Procedures

3.1   This policy is implemented through the following procedure:

- Storage and archiving

Procedure for storage and archiving of data

1 Introduction

1.1 The following document is intended to guide LRC staff through the procedures for storage and archiving of raw data both on computer media and paper in accordance with the accompanying 'Policy on storage and archiving of data'.

1.2 In this context, raw data is the information supplied to the LRC prior to any processing within the LRC.

2 Recording Formats

2.1 The LRC must be capable of archiving records in any of the following formats:

2.1.1 Recorder Import/Export format

> Section 3

2.1.2 Other computer-readable form including all formats listed in the Appendix

> Section 3

2.1.3 Computer-readable data not included in the Appendix

> Section 3.8

2.1.4 Recognised printed recording cards

> Section 4

2.1.5 Published documents

> Section 5

2.1.6 Hand-written or typed species lists and reports

> Section 4

2.1.7 Personal communications by word of mouth or telephone

> Section 6

3 Storage requirements − computer-readable formats

3.1 If facilities are available, these data should be saved to CD-ROM or DVD for secure archival storage. At least two copies should be produced and one copy stored off-site.

3.2 There should also be a procedure for duplicating disks at regular intervals to ensure against degradation of the media. At present the life of a CD-ROM is about ten years, so the duplications should be made once every five years.

3.3 If the CD-writer is not available the data should be saved to magnetic media and a paper copy produced and stored as in section 4. Again, at least two copies should be produced and one stored off-site.

3.4 At least one copy of the data must be stored in its original format; however, copies may be stored in a more convenient format if required, ensuring the included information is neither corrupted nor amended.

3.5 Amended data is no longer covered by the policy on 'Storage and archiving of data' and should be considered as analysis or LRC 'value added' information, and becomes subject to the relevant policy?.

3.6 The Received Data Catalogue should be updated to indicate the location of the archived data.

3.7 A metadatabase cataloguing the contents of archived CD-ROMs or DVDs should be maintained. These are available as either commercial software or shareware.

3.8 If the data format is incompatible, the data, although copying to CD-ROM, etc. will probably not be easily accessible. The original source media must be stored separately allowing for translation at a later date if necessary.

4 Storage requirements − on paper

4.1 If the LRC has both the facilities and the staff time, all raw data should be scanned into the computer and stored as images, preferably on CD-ROM or DVD, following the procedures under section 3.

4.2 The original paper copy should still be retained, but there is no need for photocopying. It should be permanently archived as in section 8.

5 Storage requirements − published documents

5.1 If funds allow, the LRC should acquire a copy of the original document, to be included within the LRC's library.

5.2   A library catalogue should be maintained, including summary details of local records contained within the documents.

5.3   References to published documents, articles or books will be maintained in a library catalogue, either on computer or on paper, including the book's location. If the LRC owns a copy of the document, it will be included in the LRC's library.

5.4   If copyright allows, a copy should be made of the relevant section(s) of the document, either by scanning to computer or by photocopying and then processing according to the procedures given in sections 3 and 4 respectively.

6   Storage requirements − personal communications

6.1   The storage of these is not applicable – only the paper copy produced by LRC staff can be archived, and should be processed as a paper record following the procedures under section 4.

7   Storage requirements − rejected records

7.1   If records fail to meet the minimum requirements for a biological record (refer to the procedure for processing newly acquired data), these data will be stored within the archiving system, including the reasons for omission from the primary data-set.

8   Storage requirements − general

8.1   The location of the primary data-store must be "fit for purpose" and should ideally be protected from extremes of temperature (including fire), humidity and light.

8.2   If the data have been computerised, there should be no need to photocopy source material, copies and research material can made from the computer copy. However, it is essential to use an agreed storage system, either based on grid reference (10km sq), supplier or date of receipt.

8.3   The storage system must also include sections for rejected records and confidential records

8.4   Confidential records will be afforded greater security and should at least be stored under lock and key.

8.5   The archive system must be documented and the location of all data attached to the record in the Received Data Catalogue.

9   Storage requirements − organisation

9.1   If all data are computerised, the original data can be archived without the need for easy access – access should only be necessary for validation.

9.2   If data are either not computerised or only partially computerised, it will be necessary to be able to access the data, possibly on a regular basis.

9.3   If a full manual filing system is to be developed it will need at least two copies of all data plus an additional copy for certain data, in addition to an off-site archive. This requires a high level of administration and paper and should be avoided if possible. It is, however, essential for a non-computerised LRC.

9.4   The required filing systems are:

9.4.1   A copy of all data stored by grid reference – in most LRC regions 10 km sq should be the major divisions of this system.

9.4.2   A copy of all data stored by taxonomy – the major divisions should be by phylum for most groups with reduction to class for vertebrates and order for insects.

9.4.3   A copy of any data falling within a recognised conservation site as defined by English Nature (eg SSSIs), local government (eg LNRs) or the Wildlife Trust (eg Wildlife Sites or their equivalent).

9.4.4   An unaltered copy of all data should be archived, by date of receipt or by supplier.

9.4.5   Note: depending on the information supplied, data in sections 9.4.1–9.4.3 may need to be split up between divisions of the filing system; hence the need for at least one unaltered copy (9.4.4).

9.5   If species information is computerised, the taxonomic copy (9.4.2) can be omitted.

9.6   To reduce administration and duplication, the special sites copy (9.4.3) can be included within the 10 km sq system (9.4.1) as sub-divisions. This will only cause problems with sites straddling 10 km squares, large sites covering more than one 10km sq or long linear features crossing several squares (eg rivers).

Appendix: Formats for computer-readable data archiving

The LRC will archive data in the following formats on the agreement that the LRC retains the right to reject information or media which might damage the LRC's computer system, e.g. damaged disks or regular occurrences of viruses.

Acceptable Media:

CD-ROM
DVD Disk
Iomega Zip100 Disk
3½ Floppy Disk
5¼ Floppy Disk
Electronic mail (e-mail) attachment

Acceptable Formats:

GIS: **MapInfo v4 table** together with any formats supported by this program

Databases: **Microsoft Access 97** together with any formats supported by this program

Spreadsheets: **Microsoft Excel 97** together with any formats supported by this program

Text: Standard **ASCII text**, TAB delimited **Microsoft Word 97** together with any formats supported by this program

Note: information not conforming to any of the above formats will be retained by the LRC, but may not be incorporated into the LRC system unless translation routines are available.