

NERC Biodiversity and Ecosystem Service Sustainability (BESS) Data Management Strategy 2011-2016

1. Introduction

This document aims to provide guidance to all research projects operating under the NERC funded BESS programme regarding data management throughout the life of the project and beyond. Each project will be expected to collect and archive data in accordance with this Data Management Strategy and work with the BESS Directorate and the Environmental Information Data Centre (EIDC) to ensure that data management within the BESS programmes is carried out in accordance with best practice.

2. Data management planning

Each research project is required to prepare a data management plan (DMP) within 3 months of the project start date. Although these plans should address all key stages of data management and archiving specific to the project in the first instance, they should also be seen as living documents. In this way, these DMPs should be regularly updated throughout the course of the project, as new data management issues arise. Support and guidance on the development and revision of data management plans will be provided by the EIDC.

The DMP should identify a key individual within each project who will act as a data management officer (DMO). The DMO will be the main contact regarding data issues, not only for other researchers within the project but also for the EIDC and BESS Directorate. The DMO for each project will ultimately be responsible for checking that all data protocols have been checked and that data are successfully transferred to the EIDC for future use (see section 5).

3. Data collection, recording and validation

Standardised and consistent methods should be used to collect and record data. During data collection, researchers must ensure that the data recorded reflect the actual facts, responses, observations and events.

Quality control methods during all stages of data collection and entry are important to ensure validity of data. Certain standardised methods may be useful when data are digitised, transcribed, entered in a database or spreadsheet, or coded. These may include: using data entry screens, detailed labelling of variable and record names to avoid confusion, and/or designing a purpose-built database structure to organise data and data files.

4. Data storage, use and sharing

4.1 Co-ordination and back-up

The project DMO will be responsible for collating data produced via each project work package via the work package leaders. It is the responsibility of the DMO to ensure that the data are adequately stored and backed-up on a regular basis.

Making back-ups of files is an essential element of data management. Regular back-ups protect against accidental or malicious data loss due to:

- hardware failure
- software or media faults
- virus infection or malicious hacking
- power failure
- human errors

Backing-up involves making copies of files which can be used to restore originals if there is loss of data. Choosing a precise back-up procedure depends on local circumstances, the perceived value of the data and the levels of risk considered appropriate. Where data contain personal information, care should be taken to only create the minimal number of copies needed, e.g. a master file and one back-up copy. Back-up files can be kept on a networked hard drive or stored offline on media such as recordable CD/DVD, removable hard drive or magnetic tape.

In the case of network drives and servers it is best practice to have three mirrored copies of data – one can be on a local (to the building/department) server, one in a different physical building, and one regular backup. The rule of thumb is that if data does not exist in these three locations – then it does not exist.

Physical media are best removed to another location for safe-keeping. Virtual storage on Cloud Servers such as Drop-Box are not suitable for data BESS back-ups.

For best back-up procedures, consider:

- whether to back-up particular files or the entire computer system (complete system image)
- the frequency of back-up needed, after each change to a data file or at regular intervals
- strategies for all systems where data are held, including portable computers and devices, non-network computers and home-based computers
- organising and clearly labelling all back-up files and media

4.2 File names and version control

Once data are stored, they are likely to be used by other members of the project team that have agreed permission to the data. It is therefore important that file names contain enough information about the data and that different versions are documented. It is recommended that only one, checked, version is made available for sharing via the DMO. File names should include, where possible, project acronyms, researchers' initials, study site locations, version numbers and dates but avoid using spaces, special characters and very long names. It is good practice to always maintain a master copy of each data file. It is the responsibility of the DMO to ensure an appropriate file system is developed and used consistently throughout the project.

4.3 Documentation

Whenever data are used, sufficient contextual information is required to make sense of those data. Data documentation should commence as soon as data are planned, and updated continually to keep it current. Good documentation will include information on: the context of data collection, data collection methods, data validation and any changes made to the data over time after initial creation. At a data-level, documentation should include: labels and descriptions of variables and definitions of codes and acronyms used. All contextual information must be available separate to the data itself.

4.4 Metadata

Metadata are structured sets of information that describe a dataset and its creation. Typically, metadata will explain the origin, purpose, time, reference, geographic location, creator and access conditions of the dataset. Metadata differ from documentation in that they are structured to common international standards. The production of effective metadata throughout the life time of the project will ensure effective utilisation of the data during and after the project has finished.

The legally required minimum level of metadata for collection by the project teams is set out in standards defined by the UK Location Programme <http://www.agi.org.uk/storage/standards/uk-gemini/MetadataGuidelines2.pdf>. EIDC will provide guidance for the completion of metadata to this standard.

5. Legacy and archiving for use by other researchers

One of the requirements of the BESS programme is that all data collected should be made available for use by other researchers.

All data collected and managed (as the above guidelines recommend) should therefore be submitted to the EIDC before the formal end of the project. Data must be accompanied by documentation and metadata as described in section 4 of this document. Each dataset should also be in the form agreed with the EIDC. DMOs may request an embargo on the use of the data by other researchers for a period of up to 1 year following deposition with the EIDC,

and access to the data thereafter may be subject to reasonable conditions specified by the DMO at the time of deposition. EIDC will work with DMOs to ensure appropriate Terms & Licensing are in place for each dataset.

6. Engagement with EIDC

As the EIDC is the primary NERC Environment Data Centre for BESS, Project Managers/DMOs are expected to maintain regular contact with the EIDC over the course of the project. Each project will be assigned a single point-of-contact from EIDC who will liaise with the project to agree support as necessary. EIDC will provide support and guidance on data management planning, data conditioning and data ingestion, as set out in Annex 1, to ensure all data of long-term-value generated by projects are submitted to the data centre according to the NERC Data Policy. Where appropriate, EIDC may issue doi's (Digital object identifier) to enable ingested datasets to be widely cited.

7. Further information

For information on current research council specific data policy please refer to the NERC website at:

- nerc.ac.uk/research/sites/data/policy.asp

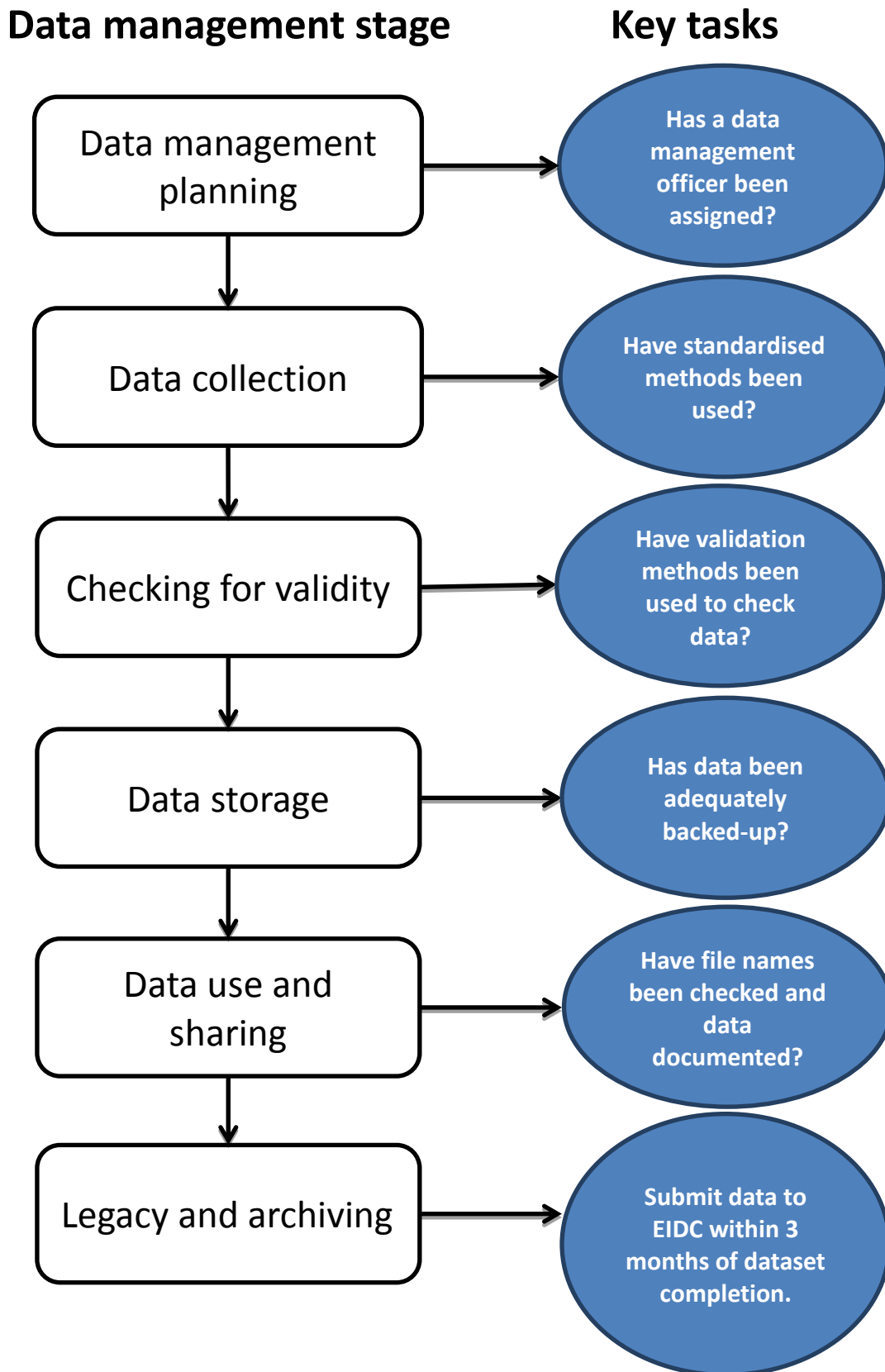
For information on best practice with regards economic and social data sets please refer to the guidance at the Economic and Social Data Services site available at:

- esds.ac.uk/support/datamanguides.asp

This includes guidance on anonymisation and confidentiality issues in relation to interview data.

Zoë Austin, Steve Cinderby and Piran White
September 2012

Figure 1. Checklist for each stage of data management



Biodiversity and ecosystem service sustainability (BESS)

Data Centre Support to Projects

Data produced by projects require a considerable amount of support in order to bring them up to the required standards for preservation and usability. The service provided by EIDC includes:

- Data management planning
 - Support and guidance on the development and revision of data management plans.
 - Establishing a data contact for each project.
 - Advising project participants on NERC Data Policy and JCOPR.
 - Helping to initially select data for long-term retention.
 - Refining the data selection over the course of the project.
 - Identifying potential view and download services for selected data
 - Determining data storage needs of the project
 - Help to identify long-term curation needs.
 - Advising on data licensing and access arrangements.
 - Specifying appropriate data standards for use in data management (e.g. INSPIRE)

- Data conditioning
 - Guidance on documenting the data products to allow reuse of the data and provide context.
 - Guidance on writing metadata records, discovery-level and contextual (e.g. for observation stations and activities).
 - Collating additional (supporting) documentation for metadata records
 - Specifying the required data formats
 - Ensuring data are described according to the appropriate data standards
 - Support and guidance on the development of necessary view and download services

- Data ingestion
 - Advising on how to format data, providing example scripts and file templates
 - Monitoring of ingested files, reporting problems
 - Providing dissemination routes for the ingested data over the long-term (e.g. populating the CEH Information Gateway (for EIDC-bound data) and/or NERC Data Discovery Service with discovery-level metadata, and associated view and download services
 - Defining appropriate citations on ingested data
 - Managing embargo periods on data, according to the NERC data policy.