



Guidance note for contributors of metadata to the National Biodiversity Data Centre



August 2008

(Version 1)

Pavel Janda & Liam Lysaght

1. Background

Metadata (or *meta information*) is the term used to describe the information about the dataset that should accompany any dataset. The production of good quality metadata is an essential element of any effective data management system, as it facilitates the understanding, characteristics and management usage of data. Any dataset that is included in the National Biodiversity Data Centre's database will require associated metadata. This will aid in determining the 'fitness of use' of data by users.

2. Generating metadata

When a dataset is ready for inclusion on the Data Centre's database and loading onto the mapping system, an associated metadata file is generated using Microsoft InfoPath. This will normally be done by staff of the Data Centre, but using the information and text supplied by the data provider. The information should be supplied by the data provider in Word document or Excel worksheet, and an Excel template is attached to guide you. Please ensure that all the necessary information is supplied with the dataset.

3. Metadata attributes

Metadata is organised under six heading;

1. Dataset Provider Information
2. Dataset Information
3. Level of Access
4. Survey Information
5. Attribute Information
6. Dataset Constraints

The dataset information necessary is set out in the **Table 1** below. This includes the data item, a description of each item and a short explanatory note.

4. Contributing data to the National Biodiversity Data Centre

If you have data you wish to submit please contact the National Biodiversity Data Centre.

Beechfield House, Carriganore, West Campus
Waterford Institute of Technology
Ireland

Tel +353 51 306240

Email: info@biodiversityireland.ie

Website: www.biodiversityireland.ie

Table 1: Metadata attributes and explanatory notes

Data item	Description
Section 1	Dataset Provider Information Details of the organisation or group providing dataset to the National Biodiversity Data Centre
Data provider	The name of the person(s) or organisation that provided the dataset. This should be the person who has the authority to manage the dataset. (Please note: The National Biodiversity Data Centre claims no rights to data provided to the Centre, but does provide a service to display data as part of its data management and mapping system).
Organisation	A brief overview of the role or objectives of the organisation or group.
Primary contact	The person who is the first contact for matters pertaining to the dataset.
Postal address	A postal address for the primary contact. For private individuals this information will not be made available to third parties without the prior consent of the provider. The Data Centre will be the default contact point for all private individuals unless otherwise specified.
Telephone number	A telephone number for the primary contact. For private individuals this information will not be made available to third parties without the prior consent of the provider. The Data Centre's telephone number will be the default contact number unless otherwise specified.
E-mail address	E-mail address for the primary contact
Website	A website address for an organisation or individual where one exists

Section 2	Dataset Information
Dataset title	The title is the name by which the dataset should be known. If the dataset does not have a title it should be given one that describes its content. Best practice is to include the data subject, date range and geographic extent, for example <i>Butterfly Distribution in Ireland, 1994 to 2004</i> . The data provider's name does not need to be listed here as the providers name is always listed with the dataset in the mapping system.
Data description	A general description of the dataset. This should summarise the subjects included, for example the species or habitats. Where there is an existing description of the dataset this should be included here.
Method of data capture	A concise description of the methods and techniques used in data collection. This would include the criteria for record inclusion, the methods used to verify records and validate the dataset. Also, where possible, details of the methods used in data capture should be provided (for example, direct observation, pitfall traps, kick sampling). Any areas of uncertainty with regard to any aspect of data capture and collation should be highlighted.
Purpose of data capture	The purpose for which the records were originally collected. If the creation of the database was for a different purpose other than from the original data collection, this should be stated. For example, the database may have been created to identify national distribution patterns, but the individual records may have been made as part of a site-based Environmental Impact Assessment.

Geographic coverage	A description of the geographic extent and coverage of the dataset. For survey-based data this may be a larger geographic area than the occurrence of data within the dataset. For example, a butterfly survey for all of Ireland may only have generated records in some counties for some species.
Temporal coverage	A description of the recording level over time. For some survey-based data this time period may be longer than that described by the records the survey generated. For example, a bird survey in 1994 may only generate records from April to September for summer migrants, but it is important to highlight that survey activity was maintained throughout the year.
Data quality	Provide a brief description of the quality control processes that have been applied to the dataset. State the general level of confidence in the data source, processing, validation and record collection.
Dataset original format	The data format originally used to store the dataset. If the data were stored in a biological recording software application, such as Recorder or MapMate, then the version of the software should be included.
Dataset status	Describes whether the dataset is complete. It should also indicate if the dataset has been published.
Additional information source	Details of where additional information about the data can be found. For example, this could refer to an organisation's website, a publication or an individual to contact.

Section 3	Level of Access
Display resolution	Sets the resolution at which data can be displayed on the mapping system. To make optimal use of the mapping system as a conservation tool, display of data at full or 1km resolution is encouraged.
View recorder	Indicate if recorder's name will be shown with each record.
Allow downloads	Indicate if downloads of the raw data are permitted.
View attributes	Indicate if additional attributes can be viewed.
View sensitive	Indicate if sensitive data can be viewed.

Section 4 (Optional)	Survey Information
	When a dataset comprised a number of individual surveys, details of the separate surveys should be provided.
Survey key	Unique survey key as specified in the dataset.
Survey title	The title to identify the specific survey.
Data subject	A general description of the dataset.
Method of data capture	A concise description of the methods and techniques used in data collection.
Purpose of data capture	The purpose for which the survey was originally undertaken.

Geographic coverage	A description of the geographic extent and coverage of the survey.
Temporal coverage	Indicate the start and end date of the survey.
Data quality	Provide a brief description of the quality control processes that have been applied to the dataset.
Additional source of information	Details of where additional information about the data can be found.

Section 5 (Optional)	Attribute Information Provides details of additional record attributes contained in datasets. This section should be completed for each record attribute defined in the database.
Attribute name	Name of the record attribute as specified in the datafile. The maximum length of an attribute name is 40 characters.
Attribute description	Description of a record attribute. The maximum length of an attribute description is 300 characters.

Section 6	Dataset Constraints (for internal Data Centre use only)
Access constraints	
Use constraints	