Data Standards for collation and digital storage of biodiversity habitat survey

This document, developed by the Cumbria Local Nature Partnership and Cumbria Biodiversity Data Centre, sets out agreed standards for the collection and storage of biodiversity habitat data in Cumbria. It has been produced to aid the development of a shared Cumbria-wide habitat evidence base, which can be used to help guide/evaluate both the nature recovery and wider natural environment agendas.

New and up-to-date habitat data is critical to a range of current initiatives/programmes, including planning and implementation of the Local Nature Recovery Strategy (and associated partner plans and strategies), and the delivery/evaluation of biodiversity net gain and natural capital investment programmes. To provide data which can work across all of these linked agendas, and can be shared and used in an efficient manner, requires data to be collected and stored in a consistent manner and/or according to agreed standards. Similarly aligning (where possible) local data collation standards with those of key national data sets (for example the updating/maintaining the Priority Habitat Inventory) will ensure more fluid flow of data from national to local and vice-versa.

The standards below are based on current work within the Cumbria LNRS evidence-base work and on developing standards within ALERC (Association of Local Environmental Records Centres) and Natural England (thanks to both organisations for access to their current thinking)

1. Overarching principles and standards for habitat data collation and storage

- **Data Sharing** - All commissioned habitat surveys should, as a starting point, be designed so that data can be readily shared between organisations, and in particular that data can be provided to CBDC to be utilised across Cumbria and to support the Local Nature Recovery Strategy evidence-base. Data should wherever possible be provided in formats which can be utilised by those without appropriate licences (i.e. conform to open licence standards) and be useable within freely available programmes (e.g. QGIS). Note this does not preclude use of other systems, however where possible data should also be collated and provided in more easily accessible formats.

- **Data permissions** – Data should be collected with appropriate permissions to share across the partners. Permissions should always be obtained from land managers/owners, contractors and developers for collation of survey information, and these should include clear permission to both store and share the data (excluding any personal data, or data explicitly covered under GDPR or other data legislation).

- **Structure of data** (i.e. data attributes) – Data should be recorded and stored according to standards as outlined in this document, assisting with both sharing, repeatability and usefulness of information.

- **Transparency of data** – Data should be provided/shared as complete and full datasets, ensuring that data can be interpreted more accurately and effectively.

2 Habitat Survey and Classification systems

2.1 Important considerations

Improving accessibility of data, and the ability to share this across different programmes and between partners is important if we are to develop more cost effective approaches to habitat/ environmental survey. Key to this is ensuring that appropriate land owner/manager permissions are obtained at the outset, permitting the 'commissioning' organisation to share data more widely. These permissions should be obtained in writing as part of an audit trail for future users. It is up to the 'commissioning' organisation to ensure that data collation, storage and sharing abides by GDPR or other data legislation/guidelines.

Ideally 'commissioning' 'organisations will consult with other partners regarding potential to gather other information or to work collaboratively on wider habitat survey programmes.

2.2 Habitat Classification

Habitat surveys are generally carried under either Phase 1 classification or National Vegetation Classification methodologies. More recently the UKHab classification has come more to the fore as it is linked directly to Defra's Biodiversity Metric. Systems to convert habitat codes within different vegetation classifications are available. The Cumbria Habitat basemap produced as part of the Cumbria Local Nature Recovery Strategy is based to a large extent on **UK priority habitat categories** and where possible individual data sets arising from different habitat classifications/methodologies have been converted to these 'broader' 'habitat types (whilst retaining more detailed habitat data within digitised attributes).

UKHab is likely to increase in importance over time (due to its inter-relationship with the Biodiversity Metric), so it is recommended that as habitat surveyors become more highly skilled with this methodology it should become the 'go to' survey classification for Cumbria. **However in the meantime both Phase 1 and NVC mapping should be utilised as a basic standard for habitat survey and when relevant the habitat should be identified using the UK Priority Habitat Classification.**

2.3 Habitat condition

Assessments of the condition of habitats are important for a number of programmes/agendas including:

- to guide delivery of and monitor success of nature recovery
- to guide where best to target investment
- as a vital part of calculating biodiversity units within biodiversity impact assessment of development and for natural capital investment

Current habitat condition assessment techniques include JNCC/Natural England's Common Standards approach (utilised on SSSI), the Defra biodiversity metric 3.1 (linked to the UKHab classification) being used as part of mandatory net gain, and developing condition assessments such as that under the Woodland standard of the Sustainable Farming Incentive pilot.

As part of biodiversity net gain it will be critical that habitat condition data is collected and submitted as part of baseline ecological reports to enable independent assessment of the metric calculation. Collating data to defined standards is therefore vital to enable repeats of the metric calculation over time. Similar data standards are needed to monitor the success of wider biodiversity investment.

At present it is suggested that JNCC/Natural England's common standards approach to recording habitat condition is utilised as part of wider surveys/monitoring on land outside of SSSI as well as that within it. All nature recovery programmes/project implemented should carry out a baseline habitat condition assessment as part of the work, enabling evaluation of success in future years. Development planning and natural capital investment programmes should utilise the Biodiversity Metric and provide this data, along with habitat survey data as outlined in Section 3 below.

3 Habitat Data Storage

3.1 Attributes

The following table outlines the attributes that should form the basis of GIS vector data and should accompany any data submitted to CBDC.

Format: Geopackage, ESRI shapefile

Attribute	Data type	Description
Habitat	Character (250)	Habitat names as received from data source/Survey.
LNRNHab	Character (200)	Collated habitats classified where appropriate according to the UK Priority Habitat classification.
HabCode	Character (20)	Habitat codes to be used in the NE habitat Modelling. To be completed by CBDC.
Date	Date	Dataset reference date
Confidence	Character (20)	Copied from data source or as assessed during the mapping process. For all recent survey is 'High'. If in doubt leave it blank.
Source	Character (200)	Source of datasets (e.g. survey or project name).
Organisati	Character (150)	Name of datasets supplier
Comments	Character (250)	Additional information from the original datasets (e.g. if the habitat is PHI habitat or include information on habitat restoration. If these information available for most of the datasets I would create a separate column but at the moment the best way is to store them in the comment column).

Additional Data if available		
Site_Name	Character (250)	Name of site surveyed
Hab_Cond	Character (200)	Habitat condition
Map_by	Character (50)	Name of person carrying out habitat mapping
Contractor	Character (50)	Name of contractor carrying out survey
BaseMap	Character (30)	Basemap used to digitise habitat boundaries

3.2 Digitisation of survey data

Habitat data should be digitised according to specific procedures and guidelines (see Appendix 1). Ideally this should be specified as an output part of 'commissioned' survey work. This is important so that data can be used across key programmes and initiatives (e.g. biodiversity net gain, updating of national inventories, evaluation of progress across the nature recovery agenda etc.). Well mapped/digitised data also enables better and more precise analysis and modelling.

To achieve this, the following principles should be observed:

- Digitisation is performed in a GIS and saved in a file type that is widely readable by GIS programs such as QGIS or a geopackage.
- The file contains a number of mapped polygons, the information/attributes for which is stored in a table conforming to the standards outlined above.
- Digitisation of mapped polygons is wherever possible based on existing mapped polygons of land use and where possible this mapping should be Ordnance Survey MasterMap (noting not everyone has access to MasterMap).
- Digitisation should ensure that there are no gaps or overlaps between new digitised habitat polygons or existing MasterMap polygons. This is called “snapping to MasterMap” and ensures digitised habitat polygons can be used seamlessly in conjunction with MasterMap itself or any other polygons based on MasterMap.

In addition to these principles, the appendix contains more details on how to produce habitat polygons that “snap to MasterMap” as produced by Natural England.

Appendix 1:

Digitising NEW Polygons – Natural England Principles (‘borrowed’ from draft ALERC standards)

1.1 All data should be mapped as polygons. There is no maximum polygon size. The size of the polygon is determined by the extent of the contiguous habitat to which associated attribution can be applied. Habitats within the same 1km square and overlapping into other 1 km squares can be mapped as multi-part polygons.

1.2 There is no defined minimum mappable size.

1.3 Polygons may not extend across roads (as defined with metaling on the OS MasterMap data) or used railways. Polygons may not extend across any rivers that are mapped as polygons.

1.4 Where a boundary follows an OS MasterMap feature the OS MasterMap feature should be copied so that the habitat feature uses its geometry.

1.5 Where a boundary follows part of an OS MasterMap feature the digital boundary should be snapped along the OS MasterMap feature so that the digitised boundary and OS MasterMap feature both share the same geometry where appropriate.

1.6 Where a boundary does not follow an OS MasterMap feature, such as where the boundary follows a feature on a aerial photograph, scanned and geo-rectified map (maybe field or historical), the digitised boundary should be captured with sufficient nodes that the digitised feature takes on the shape of the feature on the source material at a scale of 1:2500.

1.7 Where a boundary is shared between two (or more) polygons the boundaries should all share the same geometry. Thus there should be no slivers or gaps between polygons with shared boundaries.

1.8 The overall dataset should be ‘flat’, that is to say, there should not be any overlaps between polygons within the dataset.