West Midlands Regional Habitat Data Project

Stage 2 Report

WGB Environment for the

West Midlands Biodiversity Partnership

November 2008

WEST MIDLANDS REGIONAL HABITAT DATA PROJECT

STAGE 2 REPORT

WGB Environment for the

West Midlands Biodiversity Partnership

November 2008

WGB Environment

Bill Butcher
Minehead, Somerset
bill.butcher@wgbenvironment.net
07864 106967

West Midlands Biodiversity Partnership

Jeff Edwards
Senior Specialist, Biodiversity
Regional Biodiversity Coordinator
Natural England
Government Offices
St Phillips Place
Birmingham
B2 3PW
tel: 07766968661
jeff.edwards@naturalengland.org.uk

Contents

Background

The **Regional Habitat Data Project** aims to create a reliable spatial representation of all known priority BAP habitat across the region based on the total extent of existing reliable national, regional and local data sets. Stage 1 has been completed and this provides an audit of all the habitat data sets held by organisations within the region. Stage 2 seeks to produce an initial GIS layer of existing data sets that are easily convertible from existing surveys into priority BAP habitat. Work has already commenced with some initial data sets already translated but remain with individual Data holders. Stage 2 involves mobilising a range of organisations to undertake a structured programme of translation and digitisation work to ensure that as many data sets as possible are brought into the initial centralised composite GIS layer.

The need to establish better baseline data for habitats has been accepted by the Regional Assembly (Environmental Monitoring Group) based on advice from the Statutory Agencies. The lack of standardised data across the region was also identified as a major issue within the Landscapes for Living project and remains a problem for the current project to review the RSS habitat targets and for providing regional advice in connection with adaptation for climate change. The timing of the RSS Review is crucial and it is important that this uses the 'best available data'.

Objective

To create the initial centralised composite GIS layer of priority BAP habitats to feed into the RSS Phase 3 Review the habitat targets (Appendix B of the RSS) work as commissioned by the Regional Assembly to ensure that best available data is available for use by this project.

Methods

A candidate list of source datasets was compiled from the Stage 1 Report, the Landscapes for Living Technical Report and consultation with Jeff Edwards and Claire Edwards of Natural England. Consultation with the five Local Records Centres in the region, and Shropshire County Council, then identified further datasets held at county level that could contribute to the regional layer.

The principal criteria used for selection of source datasets for the candidate list were:

1. GIS format
2. Habitat dataset, not a sites dataset. An exception to this rule was where a sites dataset related exclusively to a single habitat.
3. Habitat data held in a classification that was capable of translation to BAP.
4. Includes data that was likely to be BAP, either directly or through translation.

5. Translation into BAP habitat was feasible without excessive referral by the contractor or sub-contractors to other data sources that would render the process impossible within the seven week timeframe.

The contractor then worked with contributing organisations to import copies of these candidate datasets into the contractor's computer, through web download, email attachment or electronic file transfer mechanisms. Datasets were acquired in MapInfo or ARC shape file format, and translated where necessary into MapInfo format.

Some datasets were supplied in multiple layer format; these were amalgamated into a single layer per dataset.

The contractor then reviewed each dataset, identifying and quantifying the BAP habitat component, and trialling translation routines from source classifications. Those datasets that were assessed as potentially providing significant BAP data were then prepared for translation.

An updated set of BAP definitions for those BAP habitats believed to be present in the West Midlands region (31) was compiled from the Integrated Habitat System (IHS), national habitat inventory document downloads and documents published in the 2007 BAP review. This latter source introduced new BAP habitats and amended the definitions of some pre-existing.

Remarkably, there is no agreed central repository of agreed BAP definitions, and there are conflicts between published versions as well as critical ambiguities in the definitions of new BAPs. The contractor attempted to resolve these during this contract, but answers were not forthcoming on some critical points. The outcome is unsatisfactory ambiguity in some data elements in the final product.

The Data Group of the West Midlands Biodiversity Partnership discussed and agreed two final products of this project:

1. A Final BAP Layer, containing all polygons of BAP habitat identified with a reasonable level of certainty, and a limited suite of attributes.

2. A Working BAP Layer, containing all of the above plus polygons of less determination certainty, and those that were close to the BAP definition. This layer would have the full set of attributes.

Data Formats for these two products, in terms of attribute scope and definitions, was proposed by the contractor, discussed by the Data Group, circulated to LRCs and finalised by the contractor in the light of comments received. (see Annex xx). It was agreed that it was desirable to include all of the

attributes contained in the National Habitat Inventory data format[1], and also to obtain consistency where possible with the emerging data format operating in the equivalent project in SE England[2].
It was also agreed that priority in this project should be given to populating and verifying a restricted range of attributes that would appear in the final BAP layer. The wider attribute set would be populated from source datasets where straightforward to do so, but not fully populated by further processing, or verified.

The Data Group also decided that the datasets should be restricted to polygons ("regions" in Arc). Linear ("polyline") and point features were therefore excluded. This decision means that Hedgerows are excluded from the products, and Rivers are only included where source datasets had mapped them as polygons rather than linear features.

Source habitat classifications

Data in source datasets was in one of the following classifications:

1. Phase 1
2. NCC/RSNC – an enhanced version of Phase 1
3. NVC
4. IHS – Integrated Habitat System
5. BAP Priority

Some datasets had local variations on the above, with additional project specific or county specific categories.
In a few cases polygons were attributed with habitat codes from more than one classification. These were generally translated using the closest code to BAP, or the dominant classification for that source dataset.
Polygons with more than one habitat code entry were generally not translated, as this format breaks the most fundamental rule in GIS habitat mapping. This was most frequent in NVC datasets.

The source habitat data was translated into BAP or non BAP habitat using the translation correspondences in IHS, Version 2, as enhanced and incorporating new BAP codes by the contractor for this project. This process was implemented in Microsoft Excel, the attributes having been exported from MapInfo.

The success of the translation, in terms of lowest loss of information in translation, was as follows:

1. BAP, IHS
2. NVC

[1] In fact technical fields used only by the inventory data capture tool were omitted from the final data format
[2] Managed on behalf of the region by Thames Valley Environmental Records Centre, latest data format kindly supplied by the Hampshire Biodiversity Information Centre.

3. NCC/RSNC
4. Phase 1

Phase 1 is the poorest because only around 20% of the source categories map directly into BAP, the remainder leaving ambiguity across several BAP or non BAP categories.

For the purposes of this translation process the following assumptions were made, in order to increase the proportion of source data successfully translated, at the expense of introducing some inaccuracies into the final dataset:

1. Phase 1 A111 translates to native woodland BAP (technically, some may be non-native, owing to non-alignment of definitions with respect to percentages in composition).
2. All polygons in the source datasets were translated as polygons, irrespective of shape and size.
3. Translations were assumed to be valid to the non-linear sections of IHS, even where technically they might in fact represent linear features and should, according to IHS mapping protocol, be mapped as linear features.

Sub-contractors were asked to focus any manual translation effort on polygons larger than the Minimum Mappable Unit for the relevant potential BAP habitat (0.25ha for most habitats, 0 in the case of ponds). However in automated translation polygons of all sizes were translated. [3]

IHS

IHS is unique amongst the classifications in use in that it incorporates both BAP habitats in their own right and all other habitats across the landscape. It can be used both as a translation product and as an original classification for survey and data capture.

Worcestershire County Council and LRC had adopted IHS for its comprehensive project of habitat mapping through aerial photo interpretation and collation of existing sources. Shropshire County Council had also adopted IHS for its long term project to bring habitat data together across the county.

The contractor offered to undertake First Pass (automated) translation of the dominant local datasets in Herefordshire, Warwickshire, Staffordshire and Birmingham and the Black Country during this project, for the potential benefit and use of the Local Records Centres.

[3] Filtering out of small polygons before copying into the product layers was a little inconsistent, in fact. Some of the earlier datasets had small polygons removed. On realising that in some datasets adjacent patches of identical habitat had been captured as separate polygons, this filtering was dropped, and all polygons were copied across. Spatial analysis could be undertaken later to join or remove the appropriate polygons.

National inventory data, in they contain only BAP data, is effectively in IHS format already. With other national or regional datasets, the product translation was interpreted in IHS format in addition to BAP.

Therefore the Working Layer is fully populated with IHS data, although of course much of it is only resolved within the IHS hierarchy to a fairly generic level, constrained by the ambiguities of the Phase 1 and NVC sources. It is a resource with potential to be developed into a consistent and useful product across the region.

This is more than a classification issue, as polygons in the Working Layer that do not represent BAP habitat, by definition, cannot be identified in terms of BAP habitat. They therefore must be identified using a consistent framework across the region, and IHS is the only system that meets this requirement.

"Second pass" translation

National inventories and other national and regional datasets supplied by Natural England were translated directly by the contractor.

Local datasets held by Local Records Centres and County Councils were translated first against automated routines by the contractor, and then checked and translated further by the Local Records Centre or County Council. In the case of Worcestershire this required little further processing in that the source dataset was in IHS format and the County Council had already filtered BAP habitats.

Attributes in the source dataset were mapped to the new data format. In almost all cases the contents were not edited[4]; cases of poor data capture and infringement of protocol rules in the source datasets are therefore all present in the products, alongside the high quality attribute data.

Previous translations by other organisations from source data to BAP were checked by the contractor, and any differences discussed with the local organisation. In some cases BAP definitions had changed since the original translation.[5]

Manual translations were effected wherever possible on a grouped basis i.e. all examples in a dataset of source habitat code X were translated into product habitat code Y with a Determination Quality of Z. This was necessary because in some cases the LRC was faced with thousands of polygons to check in the space of a few days. However where this was judged to be unsafe, either other source data (paper or other electronic) was inspected and translations undertaken polygon by polygon, or the entire habitat group was left without further translation. In the latter case the IHS

[4] Private landowner information was deleted from one dataset for data protection reasons

[5] These could well cause confusion in interpretation of the product dataset, as the original translations may still be present in the attributes, in the interests of a transparent audit trail. Users may be tempted to conclude a translation mistake in this project; while such a conclusion cannot be ruled out (!) the reason may lie here.

Main Habitat will normally be "UH0" (unidentified habitat), meaning that two or more main habitats are possible translation products of that source code, and it has not yet been possible to determine which represents reality. These polygons represent the bulk of the "uncertain" group in the Working Layer, not having made it into the final BAP layer.[6]

Following all translation, translated source datasets were linked back to the spatial data using unique identifiers for polygons, and all datasets merged into a single layer. These were not straightforward processes.

Filtering of data into the Product Layers

The Working Layer was generated first. There is a fuzzy boundary between the Working Layer and the remainder of the region not mapped in this dataset. The project was never intended to bring together all habitat data for the region. It contains all of the allowable Determination Qualities:

- Definitely is
- Probably the priority habitat but some uncertainty of interpretation
- Definitely present within polygon but not mappable
- Probably the priority habitat but some uncertainty of interpretation and not mappable
- Not present but close to definition

Together with some polygons where no Determination Quality has been assigned. These should either have a Determination Quality assigned later, or, arguably, be deleted from this dataset.

The Final BAP Layer comprises polygons with the following Determination Qualities:

- Definitely is
- Probably the priority habitat but some uncertainty of interpretation

The Final BAP Layer is thus duplicated in the Working Layer.

The rationale behind including the "Probably the priority habitat but some uncertainty of interpretation" and not the "Definitely present within polygon but not mappable" in the Final BAP Layer is based on the following observations:

1. The "Probably the priority habitat but some uncertainty of interpretation" has been used in national inventories and some other datasets, according to English Nature guidance, where

[6] Note that the IHS main habitat may sometimes be "Unidentified Habitat" even in the final BAP layer; this can occur where the BAP habitat concerned is a habitat complex i.e. crosses Biodiversity Broad Habitat boundaries, and is therefore placed in the Land Use/Management section of the IHS classification rather than Main Habitat. The main habitat may in some cases not be capable of translation from the source dataset; however the BAP habitat presence is confirmed.

the original survey was certain but the uncertainty arises only from lapse of time since that determination.[7] Experience has shown that that probability of continuity in these cases is high.

2. The "Probably the priority habitat but some uncertainty of interpretation" has been used extensively in cases of translation from source classifications in this project, even where the probability is high.

3. Validation of many "Definitely present within polygon but not mappable" polygons in national inventories in SW England has shown that this has often been used for large polygons where the probable extent of the priority habitat is a small proportion of the whole. Inclusion of these would therefore introduce large perceived errors of commission into the dataset.

4. Visualisation of a sample of overlaps in the Working Layer suggests that many derive from large polygons of "Definitely present within polygon but not mappable" from a national inventory overlapping with smaller polygons from a more precise local data source that gives a different BAP habitat.

Using the Datasets

Interpreting the spatial data

Habitat parcel boundaries have been digitised to GIS to a precision recorded in the DigitisationMapBase attribute using the source recorded in Source1Boundary or Source2Boundary or Source3Boundary as "Primary". Some have used OS Landline, others OS MasterMap as the base.

Note that there are overlaps between polygons in some places. These can arise in two ways:

1. Where a main habitat BAP and habitat complex BAP are both relevant to the polygon (e.g. Lowland Meadows, Coastal and Floodplain Grazing Marsh). These overlaps are legitimate, as the BAP definitions do not exclude them. (see recommendations for analysis of such data).
2. Where two or more source datasets ascribe different BAP habitat codes to the same polygon or part polygon. These almost certainly represent an error in one of the sources.[8] (see recommendations for treatment of such overlaps).

[7] The SW LRCs recommended to English Nature in 2005 a change in this protocol: see Habitat Data Custodianship, 2005.

[8] It's possible, but highly unlikely, that both are correct from different dates. It's also possible that both sources are incorrect.

Interpreting the attribute data

The full attribute list of both products, with definitions, is given in Annex xx. This section draws attention to some features of the most critical attributes.

BAPHabitatCode

This is the IHS code for the BAP habitat. This is the attribute that should be used as the primary attribute for querying. [9]

BAPHabitat

The name of the habitat, giving the correct spelling of the BAP habitat, generated from the code.[10]

DeterminationQuality

This is the original from the source data ("PriDet" in the national inventory format), where supplied, and has been assigned in this project otherwise, by the contractor or sub-contractors. It is a critical field that determines the positioning of the polygon in the respective datasets, and will also influence targets. For such an important attribute it is disappointingly weak and subjective. It is the source if much misunderstanding of habitat data, especially through users not taking notice of its contents, and criticising data quality unjustly. However the existence of the attribute is probably inevitable given the enormous variety of habitat survey and assessment processes in use.
There has been a tendency for different surveyors, data managers and translators to interpret the definition of the five categories in different ways, especially in respect of time elapsed since the last survey.

Interpretation Quality

This is directly from the source data, where populated.

Area

This has been generated from GIS in this project, over-writing any pre-existing data.

[9] Note that the code is given incorrectly in some national inventories. These have been corrected here.
[10] Any direct editing of this attribute risks spelling errors and invalid filtering/querying; it should be protected in the master copy.

Incid

The unique identifier for these datasets. Sequential from 1 in the Working Layer[11], and for those polygons that are in both datasets, the same Incid. [12]

LegacyID

The ID given in the source dataset (various attribute names). This should allow the audit trail back if queries arise, or there is a need to get more attribute data.

The Source attributes

These are key audit trail attributes, showing up to three sources for the boundary definition and habitat identification. These are normally quoted directly from the source datasets collated for this project. However, as many of these are themselves collations, some ambiguity of audit trail arises. It would be impractical to develop several tiers of sources for multiply–collated datasets. Source 1 is given as the immediate predecessor source where no previous sources are given.[13]

The IHS attributes

IHS codes are given in 7 attributes from IHSMainHab to IHSMan2 and these are translated into text in the following 7 attributes from IHSMHabTxt to IHSManTxt2. The codes are collated into a "multiplex" in IHSSummary. [14]

The IHS data is as recorded in the source dataset where it is already being used in original data collection, and translated from the source data elsewhere. Some of the originally recorded IHS data needs translating to the latest IHS version.

Key assumptions made in translation

The objective was to develop, by the deadline, at datasets that would be the closest approximation to current reality of the habitat resource, allowed by the dual constraints of the quality of the source

[11] With some gaps, where polygons have been deleted for various reasons during this project

[12] National recommendations call for polygon ids to be started with the relevant county code; we started with this intention here but it proved too complex to manage in a regional dataset. This could be amended later.

[13] Note that the immediate predecessor dataset is always noted in the attribute "previous dataset"

[14] Although an attribute IHSText was originally included in the format, as used in SE England, it seemed rather redundant as it can be easily compiled from the 7 existing IHS text fields if required. It has therefore been replaced by a "previous dataset" attribute (see above).

data and time available to translate it in this project. The product is a far from perfect reflection of that reality, but is usable within its known limitations and represents an information resource that should promote further work and quality improvement in the future.

To meet this objective, source datasets were prioritised on the probable yield of usable BAP priority habitat data. Within these selected datasets data was again prioritised for probable yield.

The higher quality "Determination Quality" assessments were also prioritised according to the following order:

1 Definitely is

2 Probably the priority habitat but some uncertainty of interpretation

3 Definitely present within polygon but not mappable

4 Probably the priority habitat but some uncertainty of interpretation and not mappable

5 Not present but close to definition

The consequence of this, when combined with the fact that the dominant source habitat classification was Phase 1, generated a higher representation in the product datasets of those BAP habitats that have straightforward mappable definitions and good correspondence with Phase 1. Data for BAP habitats with complex definitions, ambiguities in mappable definitions or those that had only recently been defined, was rarely generated in this project. Analysis of these biases will be essential in the Habitats Targets Review project.

Spatial Overlaps

Spatial overlaps have not been resolved.[15] This process will require resourcing on a scale much larger than this project. Local knowledge and access to detailed site data, where available, is essential for this work. Discussion of overlap issues and a proposed procedure to resolve overlaps are presented in Annex 3.

Known weaknesses in the regional datasets

1. Spatial overlaps – see above. One implication is that the total BAP area is not the sum of the total of individual BAP areas – double counting should be avoided in analyses.

[15] Note that the selection of certain Determination Qualities for the Final BAP Layer is, in one sense, a resolution of some of the overlaps. However many remain, even in the Final BAP Layer. The resolution of these should take into account all information in the Working Layer, and other data sources where available.

2. Low or absent coverage of certain BAP habitats known to be present in the West Midlands Region:

BAP Habitat	Reason
Arable Field Margins	Relatively ephemeral habitat; poorly defined; no surveys undertaken for this habitat
Calaminarian Grasslands	Presence in region needs confirmation; absent from the dataset
Coastal and Floodplain Grazing Marsh	A landscape scale BAP habitat that requires specific inventory compilation using vegetation and physical data e.g. floodplain limits. No systematic inevntory compilation has taken place in region.
Eutrophic Standing Waters	Few of the standing waters have been assessed for trophic status. UK lakes GIS dataset (which also lacks trophic status data) received too late to be used in the project.
Hedgerows	Linear habitats not included in the dataset. Hedgerow surveys extremely resource intensive and usually only sampled.
Inland Rock Outcrop and Scree Habitats	Linear habitats not included in the dataset.
Lowland Mixed Deciduous Woodland	Much of the woodland data is recorded as UK BAP Priority Woodland [generic], which has not been assessed for specific BAP type. Most of this is likely to be Lowland Mixed Deciduous Woodland.
Wood-Pasture & Parkland	Extremely complex definition that is resource intensive to apply. Parkland generally recorded at low precision; wood-pasture has very low data coverage.
Mesotrophic Lakes	Few of the standing waters have been assessed for trophic status. UK lakes GIS dataset (which also lacks trophic status data) received too late to be used in the project.
Oligotrophic and Dystrophic Lakes	Few of the standing waters have been assessed for trophic status. UK lakes GIS dataset (which also lacks trophic status data) received too late to be used in the project.
Open Mosaic Habitats on Previously Developed Land	A new BAP habitat from the 2007 review. Practically no relevant survey or evaluation to date.
Ponds	BAP definition has detailed criteria – survey data to evaluate against criteria is sparse.
Rivers	Ambiguity in new BAP definition. Widely different approaches to recording and GIS protocol across region. [16]
Traditional Orchards	New BAP definition has not yet been fully applied in survey and evaluation.
Upland Birchwoods	Presence in region needs confirmation; absent from the dataset
Upland Flushes, Fens and Swamps	Presence in region needs confirmation; absent from the dataset

[16] In some counties all rivers and streams have been evaluated as BAP, in others not. Once clarification of the BAP definition is available, the dataset attributes should be corrected across the region accordingly.

3. Relatively low coverage of the known habitat resource in Shropshire, Telford and Birmingham and the Black Country.
4. Low spatial precision of some data – see DigitisationMapBase attribute. Data digitised at 1:25000 or smaller scales should ideally be re-digitised zoomed in over OS MasterMap or orthorectified aerial photoographs. Some national inventory data from modelled sources appears as a raster-like grid (see heathland, for example). Furthermore there are known displacements of up to 10 meters between OS Landline and OS MasterMap. All data should in due course be brought up to OS MasterMap standard.
5. Low currency of some data – see Source1Date etc.
6. Low accuracy of some habitat identifications – especially in national inventory sources. This is clearly the case in viewing some of the overlaps.
7. There appears to be an untraceable technical data issue with the Source1Txt attribute, that prevents its translation from MapInfo to Shape format[17]. For this reason Source attributes are omitted from the Final BAP Layer. Unless or until this is solved, Shape File versions of the Working Layer will also lack the important Source1 attribute.
8. The Source1, Source1 and Source3 attributes are not linked to a metadatabase. Such linkage would have been way beyond the resources of this project. Source1 is a copy of the Source1 entries in the contributing datasets, if populated, or the name of the contributing dataset, if not previously populated.
9. MapInfo thematic map views of the Working Layer insist on showing an outline around those polygons that lack a BAPHabitat (mostly rivers). These lines distort the view, especially when zoomed out.
10. For a very small number of polygons, the attribute data may have become linked to the incorrect polygon. MapInfo appears to have the occasional glitch in automatic numbering of rows; while these have been mostly trapped in validation, a few may have slipped through. These will only come to light through use of the dataset.
11. The Warwickshire Habitat Biodiversity Audit dataset contains some polygons that are digitised to grid square boundaries(i.e. artificially split). As the MMU was applied on the basis of patch size, which in these cases is false, some peculiar shape polygons with a straight edge along the grid line occur in the dataset.

Recommendations

1. Ongoing management of datasets

The datasets are managed under a regime of continuous and targeted quality improvement, fully integrated with source datasets, especially those managed by Local Records Centres. The South East

[17] There is no apparent problem with this attribute in MapInfo copies.

England regional arrangements and the SW LRCs' report for English Nature in 2005 on Habitat Data Custodianship could provide models.

2. Working Layer scope

The Working layer should be re-defined in scope to aim to include all land that is

- possibly a BAP Habitat but insufficient data to identify or include in the regional BAP layer
- probably the BAP habitat but not mappable
- not a BAP habitat but close to the definition
- land that is capable of restoration to a BAP habitat (rather than expansion) – in BAP target terminology (see Annex 4)

3. Translation of more datasets

More datasets are targeted for translation into both layers – those that are likely to yield BAP habitat or land in the recommended scope of the working layer (see above), perhaps with an emphasis on those habitats that are relatively poorly represented in the current layers.

4. Population of more attributes within the working layer

There should be a systematic review and population of all attributes within the working layer.

5. Audit trails

There needs to be a re-examination of the audit trail logic between this dataset, the datasets that contributed to it and the original sources that contributed to those datasets. This will include the construction of a metadatabase to accommodate metadata on the sources identified in the attributes

6. IHS

The potential use of IHS as both the standard habitat classification for the working layer and the key local contributing datasets should be reviewed.

7. BAP definitions

Natural England/ JNCC/ the UK BAP should be prevailed upon to develop and manage a central web-based resource of agreed mappable definitions of BAP habitats, to avoid further confusion and resource wastage.

8. Overlaps

A process should be put in place for work to resolve all overlaps in the final BAP layer. A suggested procedure is presented in Annex 3.

9. Inclusion of linear habitats

Linear habitats should be included in future versions of the datasets. This will require a standard protocol for management of linear and polygon data across all main contributing datasets.

10. Minimum mappable units (MMUs) and point features

A rigorous MMU standard for BAP habitat data in the region should be developed. It should then be applied retrospectively to these datasets and in future data management, including in the key contributing datasets. Point features, for recording of habitats below the MMU in patch size, could be part of this protocol if required.

11. Unique Ids

The custodians of these datasets and all contributing datasets should be asked to pay meticulous attention to the management of unique Ids for polygons, within and between datasets. This is relatively straightforward to do, but if not followed, can double the resource required for collation projects of this sort.

12. Change over time

The datasets should evolve in due course into a dynamic system capable of analysing trends and other changes over time.

13. Updates by area rather than habitat

Commissioning organisations should consider the merits of arranging for habitat updates over defined areas of land – districts, counties etc. – rather than individual habitats over wide areas. This is in the interests of efficiency, through avoidance of the overlap issue at source.

14. Recording of BAP habitat in original data capture

Ideally BAP habitat and non BAP habitat should be included in such projects. BAP habitats should be identified and mapped as such in the original data capture process, rather than through post hoc translation. Again this would have strong efficiency advantages.

15. Including Habitat Condition

Data on habitat condition should be included in the datasets in the future. Habitat condition should be recorded on a habitat parcel basis. The addition of a small number of attributes to the data format would accommodate this.

Annex 1 Data Format

These are the attributes included in the Working Layer. A subset (in bold) appears in the Final BAP Layer.

Attribute	Label	Definition	Attribute source
BAPHabitat	**BAP Priority Habitat Name**	**Priority habitat name e.g. Lowland heathland**	**HIDCT**
BAPHabitatCode	**BAP Priority Habitat Code**	**IHS code for the priority habitat**	**HIDCT**
DefinitionVersion	Habitat Definition Version	Priority Habitat definition version used for determination of habitat (e.g. 1.3)	HIDCT
DeterminationQuality	**Priority Determination Quality**	**Categorises the accuracy with which the priority habitat has been determined e.g. Definitely is**	**HIDCT**
DeterminationComments	**Determination Comments**	**Comments relating to the determination quality**	**SERHL**
InterpretationQuality	Interpretation Quality	Selected from list. This is a combination of the assessment of the quality of the original habitat identification in the data source and the relationship between the original habitat type and the priority habitat type.	HIDCT
Area	**Area**	**Area of polygon in hectares as calculated by GIS**	**SERHL**
Incid	**Polygon reference number**	**This is a unique reference number that is generated when data for a habitat polygon is first entered. It has to be an unique value across the whole country, thus it has a reference to the county and then the unique value given in the county. See incid worksheet for details.**	**HIDCT**
LegacyID	**Legacy Polygon ID**	**Previously used polygon id in source dataset - kept for backward-referencing**	**SERHL**
LegacyHabitatCode	**Legacy Habitat Code**	**Code for the habitat used in its original dataset. Do not enter the text description here but in the Habitattxt. The habitat code may be the same as in Source, but when several datasets have been used to build your dataset, this may not be the case.**	**SERHL**
LegacyHabitatText	**Legacy Habitat Text**	**Previously used habitat text - kept for backward-referencing**	**SERHL**
LegacyClassification	**Legacy Classification**	**Previously used habitat classification system - kept for backward-referencing**	**SERHL**

Source1	Source 1 reference number	Index to Metadata entry in Access database. There is no protocol for which data sources are allocated to Source1 as opposed to Source2 or Source3. If more than three data sources are used choose the most influential three sources, including the primary sources for both boundary and habitat identification, and the (C) confirmation source if used. This is a lookup reference to the metadatabase that you use and lists the titles of all named datasets. It also provides an automatic link to a metadata reference code	HIDCT
Source1txt	Source 1 title	Title of source data set	HIDCT
S1captdate	Source 1 date	The date of the data source	HIDCT
S1habclass	Source 1 habitat classification	The habitat classification in which the source dataset was compiled	HIDCT
S1habtype	Source 1 habitat type	The habitat code for the polygon in the source classification (note this will be used	HIDCT
S1boundary	Source 1 boundary	Indicates if this source was used as the Primary or Secondary source for the boundary.	HIDCT
S1habid	Source 1 habitat identification	Indicates if source provides a Primary or Secondarysource of the resolved PHT or is it just a Contributor?	HIDCT
Source2	Source 2 reference number	see source 1	HIDCT
Source2txt	Source 2 title	see source 1	HIDCT
S2captdate	Source 2 date	see source 1	HIDCT
S2habclass	Source 2 habitat classification	see source 1	HIDCT
S2habtype	Source 2 habitat type	see source 1	HIDCT
S2boundary	Source 2 boundary	see source 1	HIDCT
S2habid	Source 2 habitat identification	see source 1	HIDCT
Source3	Source 3 reference number	see source 1	HIDCT
Source3txt	Source 3 title	see source 1	HIDCT
S3captdate	Source 3 date	see source 1	HIDCT
S3habclass	Source 3 habitat classification	see source 1	HIDCT
S3habtype	Source 3 habitat type	see source 1	HIDCT
S3boundary	Source 3 boundary	see source 1	HIDCT
S3habid	Source 3 habitat identification	see source 1	HIDCT
DigitisationMapBase	DigitisationMapBase	Selected from list - OS MasterMap, Landline	SERHL

		etc.	
LinearFeature	Boundary or Linear Feature	Use this if the polygon should technically be allocated to the Boundary and Linear Feature Broad Habitat, and therefore should not technically be allocated to this BAP habitat	WMRHL
SiteRef	Site/File reference	Freehand data entry upto 100 characters. Any file reference(s) that may be available this feature. This is used primarily for administration of LRC data	HIDCT
SiteName	**Site name**	**Freehand data entry upto 100 characters. Any site reference(s) that may be available this feature. This is used primarily for administration of LRC data, and may be used to provide a site name (SSSI, reserve etc) if one already exists and relates wholly or in part to the site.**	**SERHL**
Createdate	Date polygon identified/captured	Date when the polygon was identified/…	HIDCT
Createdby	Created by	Person who made the classification	HIDCT
Determiner	Determiner	Person who decided the habitat id	HIDCT
County/Unitary	**County/ unitary**	**In which the polygon falls**	**SERHL**
Sourceorganisation	**Source organisation**	**e.g. LRC name, supplying the source collated dataset to the regional habitat layer**	**WMRHL**
PreviousDataset	Previous Dataset	The name of the immediate predecessor dataset used in the Regional Habitat Data Stage 2 Project, October 2008	WMRHL
Aditionalusefulinformation	Aditional useful information	Any additional comments about the polygon which may help in the translation process.	SERHL
Generalcom	General comments	Any additional comments about the polygon, habitat, data capture process, etc NOT included elsewhere, and which are necessary to give a proper understanding of the site.	SERHL
IHSAutoTranslation	Automatic IHS translation	IHS habitat code in first pass of the translation (automatically given when using translation tool). Leave blank if translator tool was not used	SERHL
IHSTranslationAudit	IHS translation audit trail	IHS translation audit trail: automatically given with translation when using translation tool. IHSversion:Date:operator initials:procedure(if done without translation tool-Manual)	SERHL
IHSSummary	IHS Multiplex code	IHS multiplex code: This is the final habitat classification in IHS, automatically given with translation when using translation tool.	SERHL
Ihsmainhab	IHS Main Habitat Code	IHS Main Habitat Code	HIDCT

Ihsmatrix1	IHS Matrix 1 Code	IHS Matrix 1 Code	HIDCT
Ihsmatrix2	IHS Matrix 2 Code	IHS Matrix 2 Code	HIDCT
Ihsorigin1	IHS Origin 1 Code	IHS Origin 1 Code	HIDCT
Ihsorigin2	IHS Origin 2 Code	IHS Origin 2 Code	HIDCT
Ihsman1	IHS Management 1 Code	IHS Management 1 Code	HIDCT
Ihsman2	IHS Management 2 Code	IHS Management 2 Code	HIDCT
Ihsmhabtxt	IHS Main Habitat Text	IHS Main Habitat Text	HIDCT
Ihsmat1txt	IHS Matrix 1 Text	IHS Matrix 1 Text	HIDCT
Ihsmat2txt	IHS Matrix 2 Text	IHS Matrix 2 Text	HIDCT
Ihsorig1txt	IHS Origin 1 Text	IHS Origin 1 Text	HIDCT
Ihsorig2txt	IHS Origin 2 Text	IHS Origin 2 Text	HIDCT
Ihsmantxt1	IHS Management 1 Text	IHS Management 1 Text	HIDCT
Ihsmantxt2	IHS Management 2 Text	IHS Management 2 Text	HIDCT
IHSversion	IHS version	IHS version number	SERHL
Broadhab	Biodiversity Broad Habitat		HIDCT
Nbnprihab	NBN Priority Habitat Type	NBN Habitat Dictionary biotope key for priority habitat type e.g. *NBNSYS0000004618*	HIDCT
Phabfeanot	Priority Habitat Feature Notes	A list of other key habitat features that are of relevance to the habitat	HIDCT
TargetNote	Target Note	Descriptive text of habitat parcel	HIDCT
Versionno	Version number	Habitat parcel version number (1 for new polygon)	HIDCT
Moddate	Modification date	Date when the polygon classification has been modified (if applicable)	HIDCT
Modby	Modified by	Name of the person who modified it (if applicable)	HIDCT
Modsmade	Modification made	Modification made e.g. Boundary	HIDCT
Modsreason	Modification reason	Reason for modification e.g. Change in habitat distribution	HIDCT
Modscommen	Modification comment	Update comment	HIDCT
ModIHSSummary	Modified IHS Summary	Concatenation of IHS habitat and multiplex codes following this update	SERHL

Source codes:

HIDCT – Habitat Inventory Data Capture Tool

SERHL – South East England Regional Habitat Layer

WMRHL – West Midlands Regional Habitat Layer

Annex 2 Datasets

Datasets Reviewed and Used to contribute to Working Layer

County	Dataset number	Dataset Name	Source
Shropshire	1	**Shropshire Habitat Mapping Project**	Shropshire County Council
	2	Shropshire grasslands nvc 1995-5	Shropshire County Council
	3	Shropshire grasslands nvc 1996	Shropshire County Council
	20	Shropshire quarry survey	Shropshire County Council
Worcestershire	4	**Worcestershire Habitat Inventory**	Worcestershire County Council
	5	Special Wildlife Sites	Worcestershire BRC
	6	Worecstershire WT Reserves	Worcestershire BRC
Warwickshire	7	**Warwickshire Habitat Biodiversity Audit**	Warwickshire BRC
	8	Warwickshire Wildlife Sites Project	Warwickshire BRC
	9	Warwickshire BAP Habitats	Warwickshire BRC
Herefordshire	10	**Herefordshire BAP Priority Habitats**	Herefordshire BRC
	11	Herefordshire Native Woodland Restoration Project	Herefordshire BRC
	19	Herefordshire Grassland Inventory Updates outstanding	Herefordshire BRC
	12	**Herefordshire Phase 1**	Herefordshire BRC
Staffordshire	13	**Staffordshire Habitats Surveys 2003+**	Staffordshire Ecological Record
	14	Grasslands inventory update work	Staffordshire Ecological Record
	15	Heathlands inventory update work	Staffordshire Ecological Record
Birmingham/ Black Country	16	**Birmingham SINCs**	EcoRecord
	17	Various Phase 1 Surveys outside of SINCs	EcoRecord
Staffordshire	21	**WoodPasture1999_2000_region [Staffs]**	Natural England
Shropshire & Staffordshire	22	**Meres_2005_region**	Natural England

Shropshire & Cheshire	23	Meresmoss	Natural England
Region	24	**Lowland_calcareous_grassland_WM_region**	Natural England
Region	25	**Lowland_dry_acid_grassland_WM_region**	Natural England
Region	26	**Lowland_meadows_WM_region**	Natural England
Region	27	**Purple_moor_grass_and_rush_pasture_WM_region**	Natural England
Region	28	**Undetermined_grassland_WM_region**	Natural England
Region except Warwickshire	29	**West_Midlands_parklands_minus_4c**	Natural England
Warwickshire	30	**Warwickshire_parklands**	Natural England
Staffordshire	31	**Staffordshire_BAP**	Natural England
Staffordshire	32	**BAP_data_from_advisers_submissions**	Natural England
Region	33	**BAP_habitats_from ecological_reports**	Natural England
England	36	**Blanket bog national inventory**	Natural England
England	37	**Coastal and floodplain grazing marsh national inventory**	Natural England
England	38	**Fens national inventory**	Natural England
England	39	**Lowland beech and yew woodland national inventory**	Natural England
England	40	**Lowland heathland national inventory**	Natural England
England	41	**Lowland mixed and deciduous woodland national inventory**	Natural England
England	42	**Lowland raised bogs national inventory**	Natural England
England	43	**Reedbeds national inventory**	Natural England
England	44	**Upland heathland national inventory**	Natural England
England	45	**Upland mixed ashwoods national inventory**	Natural England
England	46	**Upland oakwoods national inventory**	Natural England
England	47	**Wet woodland national inventory**	Natural England
Region	48	**Forestry Commission Woodlands NVC & BAP**	Forestry Commission
Region	49	**Land Cover Map 2000**	CEH
Region	50	Forestry Commission Semi-natural scores	Forestry Commission
Region	51	**ENSIS SSSI Data**	Natural England
UK	52	UK Lakes	Environment Agency
Staffordshire	53	**Lowland heathland regional update - Staffordshire**	Natural England

Annex 3 Overlap resolution

The resolution of spatial overlaps will be a complex and time consuming process. It should be noted that national inventories have normally been generated independently of each other. Overlap resolution has only previously been attempted on a restricted range of source datasets within some Local Records Centres.

One source of complexity is that conceptually some types of overlap are allowable. For example:

1. Some BAP Habitats can overlap by definition. There is currently no published document defining such overlaps. A matrix generated jointly by JNCC and SERC in 2003 is available as a source file as part of IHS.[18]
2. Some habitat classifications have protocols for spatial overlap; others are silent on the issue. IHS is designed to avoid overlaps in its coding structure.[19]
3. The Determination Quality definitions conceptually allow spatial overlap between, for example, two polygons each with Determination Quality "Definitely present within polygon but not mappable". There are many such examples in these datasets.

Other complexities can arise from the following:

4. The time dimension[20]. Two or more polygons at the same location may represent habitat change rather than errors. Distinguishing between these possibilities is sometimes a matter of great expertise and may not always be possible.
5. Most overlaps are partial, spatially. Therefore, even when a robust hierarchical decision can be made between overlapping polygons, there often remains another decision on whether to also delete the area of the non-preferred polygon that does not overlap with the preferred polygon.
6. It's possible that in some locations the best final product would be obtained by combining the boundary of one polygon with the attributes of another overlapping polygon.
7. Arc GIS and MapInfo handle spatial overlaps differently. Arc is much more rigorous in its typology; therefore MapInfo can be easier to use in manipulating datasets with overlaps.

Assessment of the general quality (high precision, currency, accuracy) of contributing datasets can be used to generate a hierarchy for use in resolving overlaps. All other things being equal (which

[18] Note that this matrix currently excludes new BAP habitats defined in 2007

[19] There are a few exceptions to this rule, made necessary by poorly structured BAP definitions.

[20] The datasets imply a best available representation of the resource at the time of compilation, although the sources can range from 1 to 20+ years before present. There is no attempt to record change through time. Some LRCs in the region have systems for managing habitat datasets that include the time dimension, but there is no standard methodology.

they never are!) the following general hierarchy of source datasets might be used to generate a first draft of overlap resolution:

Group	Datasets
1	Regional Habitat Inventory Updates Lowland heathland regional inventory update Meres 2005 Region
2	Warwickshire Habitat Biodiversity Audit Staffordshire Ecological Record Habitats Herefordshire BRC BAP Data EcoRecord Worcestershire Habitat Survey Shropshire Habitat Survey
3	Herefordshire Phase 1 Non Water Herefordshire Phase 1 Water Warwickshire Parklands Forestry Commission Woodlands NVC BAP
4	ENSIS SSSI Data Natural England National Habitat Inventory Agri-Environment Ecological Reports 2000-2008 BAP Data from Adviser Submissions, Staffordshire
5	West Midlands Parklands minus 4C

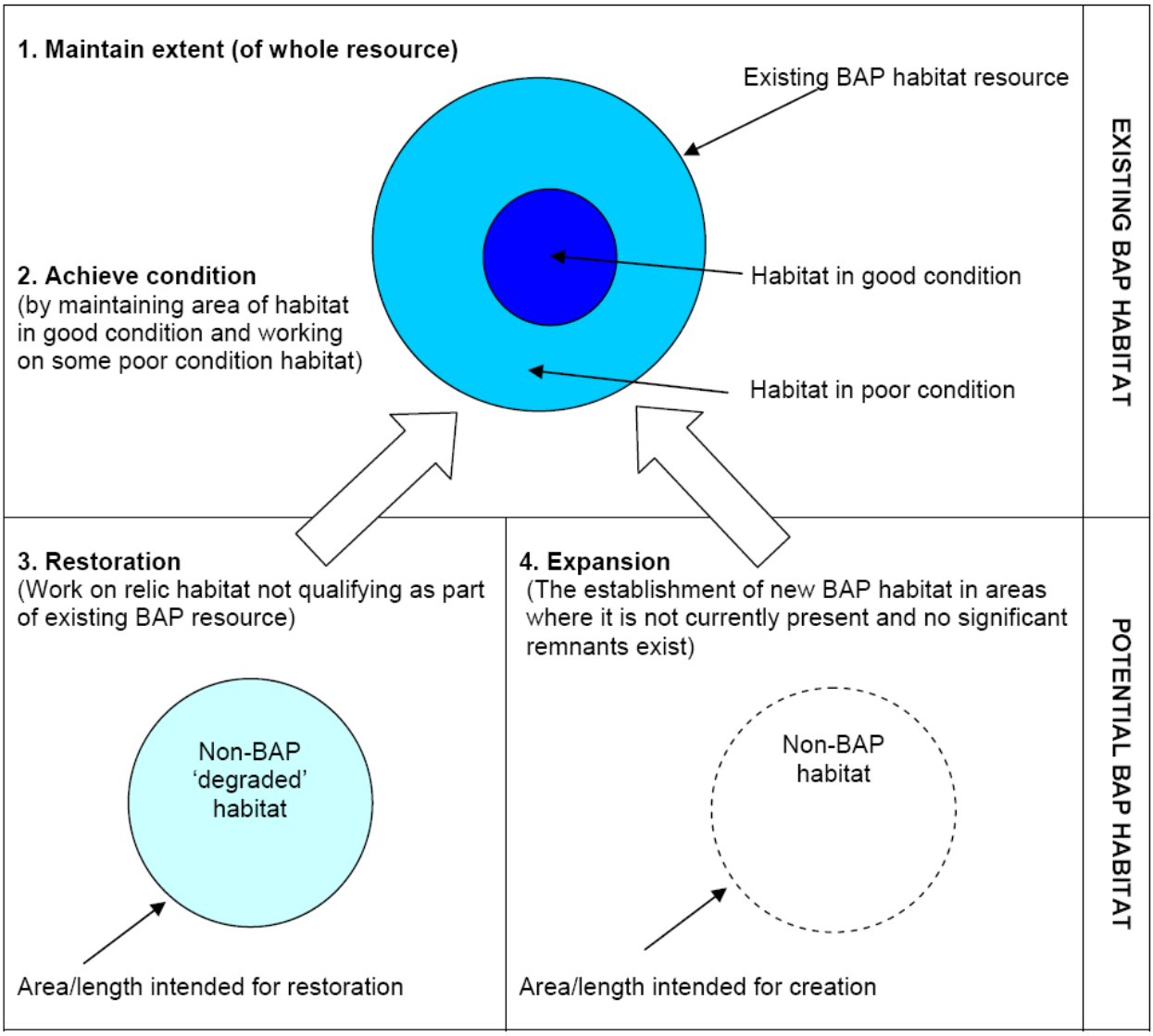
The procedure would be to examine each overlap; where polygons arising from source datasets in different groups, the polygon from the higher numbered group is selected for possible deletion. Attributes would then be used to make a final decision, the key attributes in this respect being Determination Quality, LegacyHabCode, Source1Date, Interpretation Quality, DigitisationMapBase. Supporting detailed data and aerial photographs should also be used where possible. A clear audit trail of the decision making process should be recorded.

There also needs to be a validation of the logic between the BAPhabitat attribute on the one hand and the IHS attributes that contain BAP habitats on the other. This is especially in respect of habitat complexes. For example a Lowland Meadow occurring within a Coastal and Floodplain Grazing Marsh, coded as GN1.CF1 in IHS must either be coded as GN1 or CF1 in the BAP Habitat attribute, not both. There are a few instances of this situation in this dataset. The main habitat (sensuo IHS) should normally be selected (GN1 in this case) but the under-representation of habitat complexes (CF1 etc.) resulting needs to be remembered. Alternatively, and more robustly, for statistical purposes, IHS codes should be fully analysed for main habitat BAPs and habitat complex BAPs – appearing in either the IHSMan1 or IHSMan2 attribute – so that each BAP is accurately reported; the

total area of all BAPs with this approach can only be analysed by GIS, not by summing the areas of individual BAPs.

Annex 4 BAP Target Terminology Definitions

(see recommendations on scope of Working Layer, and inclusion of habitat condition)



Note: In terms of this diagram the final BAP layer contains the known extent of the two dark blue circles, , existing BAP habitat, but is silent on condition. It will also have a little of the light blue (some polygons recorded as the "Probably BAP" Determination Quality, but, in reality, not).

The Working Layer contains some of the unknown extent of the dark blue circles (but its definition within the dataset is unknown) plus some, but not all, of the light blue circle. It will also contain a little of the dotted line circle – note the distinction in definition between the light blue and dotted circles is fuzzy and probably not capable of full definition in a GIS dataset.

Comprehensive datasets, such as the Warwickshire Habitat Biodiversity Audit, will encompass all three circles, although the distinctions between the three will not necessarily be clear. The Worcestershire Habitat Survey, in IHS format, also contains all three circles, and is clearer on the definition of the existing BAP habitat, by virtue of the inclusion of BAP habitat in the IHS classification.