OxCam LNCP Natural Capital Baseline Assessment:

Data Report

Author:

Dr Jim Rouquette

Reviewed by:

Dr Alison Holt, Natural Capital Solutions
Ceri Lewis, Environment Agency

Contact details:

Dr J.R. Rouquette

Natural Capital Solutions Ltd

www.naturalcapitalsolutions.co.uk
jim.rouquette@naturalcapitalsolutions.co.uk

Report prepared for:

OxCam Local Natural Capital Plan Project, Environment Agency

Version: Final

May 2020

Contents

1.   Introduction

Accurate knowledge of which habitats are present, their extent and spatial location is an important pre-requisite for properly understanding natural capital, the benefits that it provides and the opportunities to enhance it. This is an area in which Natural Capital Solutions have been involved for some time, having already produced detailed habitat basemaps for Northamptonshire, Cambridgeshire and Buckinghamshire (including Milton Keynes). The Oxford-Cambridge Arc Local Natural Capital Plan team have therefore commissioned Natural Capital Solutions to undertake a project with the following five aims:

1. Create a detailed habitat basemap based on the best available existing data for Bedfordshire and (separately) for Oxfordshire.
2. Merge these maps with already existing habitat basemaps for Northamptonshire, Cambridgeshire and Buckinghamshire to produce a map for the whole OxCam Arc.
3. Produce a lower resolution raster (gridded) version of the habitat basemap for pubic release, along with a series of raster maps showing the location of key individual habitat types.
4. Carry out a detailed assessment to compare the basemap for Oxfordshire with other basemaps produced using a similar but entirely independent approach and with different data inputs.
5. Write reports on the processes involved in producing the basemap and raster maps (a data report), the basemap comparison, and compile a report on the lessons learnt.

This report is the first of those reports: the *Data report*. Section 2 provides an outline of the mapping approach undertaken, and a detailed description of all the data sets used as input. Section 3 briefly describes technical aspects of the vector output (the main habitat basemap), before Section 4 describes the methods and outputs of the raster conversion. Recommendations are reserved for the *Lessons Learnt* report. Technical appendices at the end provide detail on each processing step involved in creating the basemaps (Appendix 1) and give a full list of habitats mapped and the different ways in which the habitats were grouped (Appendix 2).

Two versions of the basemaps have therefore been produced as outputs of this project. The detailed vector version, based on OS MasterMap, which is only available to project partners that meet license conditions and the low resolution raster version which provides an indicative map of broader habitats across all five counties, which will hopefully be freely available for all to view online at a scale agreed with third part data owners.

2.   Producing the natural capital baseline maps

2.1   Approach to mapping habitats

The same general approach was used for producing the habitat basemaps across all five counties in the OxCam Arc. This used Ordnance Survey MasterMap Topography Layer polygons as the underlying mapping unit and then used a series of different data sets to classify each polygon to a detailed habitat type and to associate a range of additional data with each. Polygons within MasterMap separately map each building, field, road section, verge, pavement and so on, but only provides limited information on the land use or habitat of these polygons. The aim was to use the best available data sets to classify the polygons more accurately, although this varied for each county mapped. Locally derived habitat data provided the most

accurate and detailed information, but the availability and quality of this local data varies considerably from county to county and led to the greatest differences between maps. Local data was supplemented by a number of national data sets which provided a more consistent but less accurate underpinning for each map. The data that was used to classify habitats for Oxfordshire and Bedfordshire is shown in Box 1 and are described fully in Section 2.2.

Box 1: Data used to classify habitats in the basemaps for Oxfordshire and Bedfordshire:
- OS MasterMap Topography layer
- Phase 1 habitat data for Oxfordshire – supplied by Thames Valley Environmental Records Centre (TVERC)
- Habitat data for Bedfordshire – supplied by Bedfordshire and Luton Biodiversity Recording and Monitoring Centre (BLBRMC)
- Natural England Priority Habitats Inventory
- Centre for Ecology and Hydrology (CEH) Landcover Map 2015
- OS Mastermap Greenspace data
- OS Open Greenspace data
- Built-up Area Boundaries data
- Ancient Woodland Inventory data

Our approach to producing the habitat basemaps was based on the **EcoServ-GIS toolkit[1]**. However, this toolkit was developed several years ago and since that time a number of bugs have developed which means that the original toolkit no longer works properly, therefore cannot be used off the shelf. It is also not able to fully integrate Phase 1 habitat data, National Forest Inventory data and other key data sets. Hence we used the core processes and rules from EcoServ-GIS, but with a large number of bespoke modifications and additional steps. Further information on how polygons were assigned to habitats is provided in Box 2 (overleaf) and full details of each processing step are provided in the technical appendix (Appendix 1). Polygons were classified into Phase 1 habitat types and were also classified into broader habitat groups, as outlined in Section 3. All mapping was undertaken in ArcGIS.

Please note that the basemaps provide the best approximation of habitat types that can be achieved based on available data, but have not been ground-truthed and will inevitably contain errors. Although the habitat basemaps for the five counties have been produced using the same method, different data sets were available for each county, so there will be some difference in the quality of the outputs. For example, the habitat data supplied for Oxfordshire covered almost the whole county and was highly detailed, whereas the local habitat data supplied for Bedfordshire was much less extensive. A particular challenge was classifying polygons where more than one habitat was present. Mixed habitats containing woodland and scrub, or grassland with woodland were classified in some detail, but not all combinations of habitats could be accommodated. In other areas, where there was a mismatch between data sources, or where land use is changing rapidly classifying the polygons remained a challenge. These issues are explored in much more depth in the *basemap comparison report for Oxfordshire*, which also explores the effect of using different data and a slightly different approach.

---

[1] Winn, J.P., Bellamy, C.C. & Fisher, T. 2018. EcoServ-GIS: a toolkit for mapping ecosystem services. *Scottish Natural Heritage Research Report No. 954.*

Box 2: Assigning habitats

Our approach to assigning habitats is illustrated pictorially below. OS MasterMap Topography Layer is the most detailed and accurate mapping available across Great Britain and identifies all roads, buildings, fields and other features as individual polygons (shown pictorially as the black layer below). However, information on the habitat of these features is limited. We used a series of rules and other layers to classify each polygon. For example, we used rules to assign features as houses, gardens, industrial / commercial buildings and so on.

The data sets listed in Box 1 and described below were then overlain in turn (example shown as red layer in Fig.1) and the degree of overlap calculated using zonal statistics. These often do not match precisely so, for example, if a polygon marked as semi-natural broadleaved woodland (A1.1.1) (red polygon on left, below) overlaid houses, gardens and a polygon identified as non-coniferous trees in OS MasterMap, we could now assign the non-coniferous tree polygon more accurately as semi-natural broadleaved woodland, but the houses and gardens would be left unchanged.

We were also able to check the accuracy of the data using additional sources, where possible. So, for example, if a locally derived habitat polygon identified arable land (J1.1, red polygon on right, below), and this matched a OS MasterMap polygon identified as agricultural land, we would also check that the CEH Land Cover Map 2015, also classified the land as arable. This is one of the key benefits of having multiple data sources to help provide confidence in the basemap.

A number of additional rules and layers were used to gradually build up as complete a picture as possible. For example, areas identified as improved grassland, but within urban areas, were classified as amenity grassland. All polygons were assigned to a Phase 1 habitat type, although areas currently undergoing development were marked as unclassified. Upon initial completion, the basemap was checked against Google maps, focussing on areas where classification was uncertain, and manual alterations were made in a number of places.

More information on the basemapping methodology, including a detailed description of all the steps involved, is provided in Appendix 1.
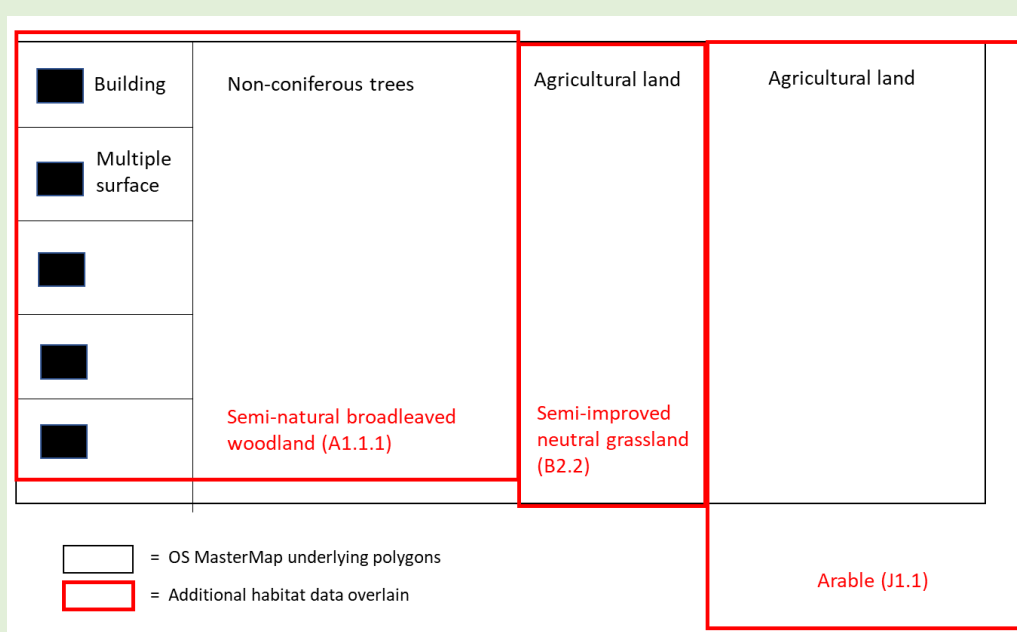
2.2   Data sets used

OS MasterMap Topography Layer

Source: Ordnance Survey

Cost: High (£££ tens of thousands for a county) for commercial use, but available to public sector under Public Sector Geospatial Agreement (PSGA), which has recently replaced the Public Sector Mapping Agreement (PSMA).

Restrictions: Currently only available under licence (although will be partially freely available from summer 2020[2]), outputs retaining original polygons can't be passed on to others unless they hold a license.

Description: Provides the most detailed vector mapping available across GB, mapping all buildings, roads and land polygons. We used the Area (polygon) layer. Does not include individual trees or field boundaries.

Use in basemap: Provides the underlying mapping polygons on which other data is overlain to classify habitats. Attributes "Descriptive Term", "Descriptive Group" and "Make" also used to help classify habitats.

OS MasterMap Greenspace

Source: Ordnance Survey   *Cost:* Free to license holders only

Restrictions: Only available to holders of the Public Sector Geospatial Agreement (PSGA).

Description: Based on MasterMap polygons, this highlights all green spaces in larger urban areas and gives their function (e.g. Amenity – transport, cemetery, school grounds, playing field etc.) and in some cases their form (e.g. manmade, open water, woodland etc). Note that OS MasterMap Greenspace only covers larger urban areas, not smaller towns and villages.

Use in basemap: Provides information on function, which is a useful addition to the basemap, especially for use in mapping public accessibility and ecosystem services. Also helps with assigning habitats (e.g. Amenity – transport is assigned the habitat of amenity grassland (J1.2)).

OS Open Greenspace data

Source: Ordnance Survey   *Cost:* Free

Restrictions: Subject to the terms of the Open Government Licence

Description: Provides a simplified version of the MasterMap Greenspace data, focussing on public open spaces such as public parks, playing fields, sports facilities, play areas and allotments. Does not include amenity greenspaces or private gardens.

Use in basemap: Used to supplement the OS MasterMap Greenspace data, providing information on greenspaces in smaller urban areas and villages. Can be used in place of the above where OS MasterMap Greenspace is not available.

Natural England Priority Habitats Inventory (PHI)

Source: Natural England	(data.gov.uk) *Cost:* Free

Restrictions: Subject to the terms of the Open Government Licence

Description: Identifies the geographic extent and location of Natural Environment and Rural Communities Act (2006) Section 41 habitats of principal importance (previously known as UK BAP habitats), hence the

---

[2] https://www.ordnancesurvey.co.uk/business-government/tools-support/open-mastermap-programme

most important semi-natural habitats in England. Note that for this project woodland habitats were excluded as more reliable sources of information were available for these habitats.

Use in basemap: Used to classify semi-natural habitats, unless more detailed locally collected information was available (see Phase 1 habitat data below).

CEH Land Cover Map 2015 vector (LCM 2015)

Source: CEH   *Cost:* Medium (££ hundreds for a county)

Restrictions: Subject to licence conditions. Output maps that have been created using this data cannot be shared with un-licensed users except as low resolution rasters.

Description: Available as either a vector or raster product, this project used the vector version. Provides a habitat map derived from Landsat satellite data.

Use in basemap: Primarily used to distinguish arable from pasture. Other LCM 2015 habitat types were only used to confirm habitat classifications, as alternative data sources were available that provided more reliable (ground-truthed) and higher resolution data for most semi-natural habitats.

Built-up Area Boundaries data

Source: ONS (data.gov.uk)   *Cost:* Free

Restrictions: Subject to the terms of the Open Government Licence

Description: Provides vector boundaries for built-up areas in England and Wales based on the 2011 census. Provides an approximate outline of all urban areas.

Use in basemap: This data set was used to assist in classifying habitats. For example, areas identified as grassland in urban areas (outside of gardens) were classified as amenity grassland (code J1.2), rather than agricultural improved grassland (B4).

Ancient Woodlands Inventory (AWI)

Source: Natural England (data.gov.uk)   *Cost:* Free

Restrictions: Subject to the terms of the Open Government Licence

Description: Identifies ancient woodland sites in England (parallel data sets are also available for Scotland and Wales) based on old maps and a variety of other sources.

Use in basemap: Not used to classify habitats, but can be used to highlight ancient woodland sites on the basemap.

Phase 1 habitat data for Oxfordshire

Source: TVERC (Thames Valley Environmental Records Centre)   *Cost:* Medium-high

Restrictions: Subject to licence conditions. Output maps that have been created using this data cannot be shared with un-licensed users except as low resolution rasters.

Description: Detailed habitat survey covering most of Oxfordshire, based on a mixture of field survey data and aerial photograph interpretation. Habitats have been mapped to identify S41 Habitats of Principal Importance and JNCC Phase 1 habitats.

Use in basemap: This data set was a key source for classifying habitats across rural Oxfordshire. Both Habitats of Principle Importance and Phase 1 habitat classifications were used.

Habitat data for Bedfordshire

Source: Bedfordshire and Luton Biodiversity Recording and Monitoring Centre   *Cost:* Medium-high

Restrictions: Subject to licence conditions. Output maps that have been creating using this data cannot be shared with un-licensed users except as low resolution rasters.

Description: Series of individual layers showing the location of certain key habitats (S41 Habitats of Principal Importance) in Bedfordshire and Luton. In addition, layers showing the location of neutral grassland based on old (1980s) Phase 1 habitat survey data but updated using aerial photo interpretation.

Use in basemap: The different S41 data layers were amalgamated into a single layer and then cross-referenced and amalgamated with the PHI data set. Used to classify semi-natural habitats in Bedfordshire. The neutral grassland layers were used as a cross check at the end of the mapping process.

National Forest Inventory (NFI)

Source: Forestry Commission   *Cost:* Free

Restrictions: Subject to the terms of the Open Government Licence

Description: A vector map covering all forest and woodland area over 0.5 hectare with a minimum of 20% canopy cover (or the potential to achieve it) and a minimum width of 20 metres. It classifies woodlands into broadleaved, conifer, mixed predominantly broadleaved, mixed predominately conifer, and a number of other categories.

Use in basemap: Was used to clarify the type of woodland previously identified on the basemap.

DTM

Source: Various   *Cost:* Medium

Restrictions: Depends on the source, but subject to license restrictions

Description: Digital Terrain Model, which can be created from a variety of sources. For this project we used a 5m DTM provided by the Environment Agency under license. A free 50m resolution version is available from OS (OS Terrain 50), but that was not considered to be detailed enough for this project. It is anticipated that a free 1m or 2m DTM will become available from the Environment Agency based on LIDAR measurements in the near future.

Use in basemap: For lowland areas (such as the Ox-Cam Arc) the use of a DTM is not critical, although it can add useful data on heights and slopes for each polygon. For study areas including upland areas, slope and height are used to help classify habitats into unimproved, semi-improved or improved grassland types and to identify certain upland habitats.

Boundary-Line

Source: Ordnance Survey   *Cost:* Free

Restrictions: None

Description: Maps administrative boundaries.

Use in basemap: Used to map the study area boundaries

3. The natural capital basemaps

3.1 Basemap attributes

The main output was a vector map (shapefile), based on OS MasterMap polygons, but with a detailed habitat type and other attributes assigned to each polygon. The key attributes assigned to each polygon are shown in Table 1 below.

Table 1. Description of the attributes of the habitat basemaps.

Attribute	Description
OBJECTID	Parcel identifier for each polygon in the basemap.
Toid	Topographic Identifier assigned by Ordnance Survey in MasterMap. Each TOID is unique and enables cross reference back to MasterMap source data.
Ph1code	Detailed habitat code applied to each polygon. Based on JNCC Phase 1 habitat codes, but includes some additional categories. A full listing of all the habitat codes is given in Appendix 2.
HabNmPLUS	Habitat description that matches the Ph1code (above). This is the most detailed level of classification of the habitats. E.g. Woodland, Broadleaved, Semi-natural.
HabBroad	Broader grouping of habitats that classifies and groups them at a less detailed level e.g. Woodland, Broadleaved. Further details in Appendix 2.
Ph1ColBestFit	Used to display the most detailed habitat classification. The Ph1code is adjusted to fit standard Phase 1 codes and symbology.
HabType2	A further broader grouping of the habitats, created in discussion with the OxCam LNCP team to reflect local interest. Considered best for displaying the data over larger areas. The habitat types are shown in Table 2, with the relationship with the other habitat groupings shown in Appendix 2.
Shape_Length	Standard measure of polygon length.
Shape_Area	Standard measure of polygon area.

The final Oxfordshire basemap contained 1.38M polygons and the Bedfordshire basemap contained 1.08M polygons, each of which was classified to an appropriate habitat type. When the basemaps for all five counties were merged, the combined basemap contained 6.81M polygons and covered an area of 1,148,900 ha or 11,479 km^2.

3.2 Projection and display

The basemap and the rasters described in Section 4 were projected using the British National Grid.

Two layer files were created to enable the map to be displayed using a consistent symbology across different maps and by different project partners. One was created to display the most detailed Phase 1 habitat types (use Ph1ColBestFit) based on standard Phase 1 colours and symbology and one was created for the broader HabType2 grouping.

3.3   Habitat classifications

Habitats were classified into detailed Phase 1 habitat types, but with a number of additions to cater for mixed habitats and further urban habitats and land uses. In total, 99 habitat types were present across the OxCam Arc and a full list is provided in Appendix 2. These were given a code (Ph1code) and a name (HabNmPLUS).

In addition to the detailed habitat types, we also grouped the habitats into two broader groupings, known as HabBroad and HabType2. There were 31 HabBroad categories across the OxCam Arc and 18 HabType2 categories. The HabType2 groupings were used to display habitats on the low resolution rasters (Section 4) and are best for displaying the maps over a large area. They were developed in consultation with the OxCam LNCP team and their technical group. The HabType2 categories and their constituent habitats are shown in Table 2, and the number of polygons, area and % cover of each of the HabType2 categories across the whole OxCam Arc is shown in Table 3. A full list of how each detailed habitat type fits into HabBroad and HabType2 is shown in Appendix 2.

Table 2. HabType 2 categories and brief description of the habitats that fit within each category.

HabType2	Sub-categories
Cultivated / disturbed land	Predominantly arable, also includes allotments
Uncertain agriculture	Agricultural land of unknown habitat
Improved grassland	Agricultural improved grasslands, amenity grassland
Semi-natural grassland	Acid, neutral, calcareous, rough and semi-improved grasslands, also those with scattered trees / scrub
Marshy grassland	Floodplain grazing marsh, purple moor grass and rush pasture, other wet grassland
Heathland	Wet and dry heath, acid grassland heath mosaic
Fen, marsh and swamp	Fen, flush, swamp and reedbed
Broadleaved woodland	Semi-natural, plantation and unknown broadleaved woodland, plus orchards
Coniferous woodland	Semi-natural, plantation and unknown coniferous woodlands
Mixed woodland	Semi-natural, plantation and unknown mixed woodlands
Scrub	Dense or scattered scrub
Trees / Parkland	Broadleaved, coniferous and mixed parkland and scattered trees on improved grassland
Water	Standing and running water of all types
Rock, exposure and waste	Quarry, landfill, spoil, and natural rock outcrops
Built-up areas and infrastructure	Buildings, roads, pavements, railways and paths
Garden	Domestic gardens, also includes larger garden areas around non-domestic buildings
Mixed / other / uncertain	Mixed habitats, unknown grassland, tall herb and fern, bare ground and other habitats
Unclassified	Land currently being developed

Table 3. The number of polygons, area and % cover of each HabType2 category across the whole of the OxCam Arc.

HabType2	Frequency	Area (Ha)	% Cover
Cultivated / disturbed land	76,020	620,391	54.0
Uncertain agriculture	7,494	9,163	0.8
Improved grassland	788,929	224,462	19.6
Semi-natural grassland	39,273	34,791	3.0
Marshy grassland	4,142	8,713	0.8
Heathland	268	310	0.0
Fen, marsh and swamp	1,283	923	0.1
Broadleaved woodland	47,766	54,738	4.8
Coniferous woodland	5,787	8,975	0.8
Mixed woodland	8,400	14,064	1.2
Scrub	8,797	3,196	0.3
Trees / Parkland	70,212	19,064	1.7
Water	94,637	16,675	1.5
Rock, exposure and waste	567	2,784	0.2
Built-up areas and infrastructure	3,630,312	69,569	6.1
Garden	1,983,733	49,545	4.3
Mixed / other / uncertain	30,265	7,025	0.6
Unclassified	9,258	3,477	0.3
TOTAL	**6,807,143**	**1,147,867**	**100**

4.   Rasters

4.1   Raster showing all habitats

License restrictions on the data inputs used to create the basemaps meant that it is not possible to share these outputs widely, but creating a publicly accessible output was considered to be a key goal of the OxCam LNCP project. Therefore, a new raster version was created for public access, which removes the underlying polygons and some of the attributes. The resolution of the rasters was determined following discussion with the data providers to fit with their requirements, hence the resolution is lower than we would have preferred. The map, therefore, cannot be used to identify field-level detail, and small areas of habitat are lost, but it is still useful for examining the landscape-scale picture over wide areas.

The key steps involved in the production of the raster is outlined below:

1. The detailed vector basemaps for all 5 counties (Bedfordshire, Buckinghamshire (including Milton Keynes), Cambridgeshire, Northamptonshire and Oxfordshire) were merged.

2. The merged basemap was converted to a raster (polygon to raster) using HabType2 as the habitat classification. Each pixel on the raster was assigned a habitat based on the maximum combined area of the habitats, which means that the area of each different habitat occurring within the pixel was summed and the habitat type with the largest summed area within the pixel determined the habitat applied to that pixel. Raster maps were produced at two different resolutions: 100m and 500m.

3. A symbology layer was created that matches the vector basemap symbology layer for HabType2. This allows for easier interpretation and enables consistent display by different partners.

4.2   Rasters showing individual habitats

In addition to creating an overall habitat map to mirror the main vector basemap, we also created low resolution raster maps of a number of key habitats. This highlights the location and relative abundance of key habitats and also reveals where they occur where they would not have shown on the overall raster as they were not the main habitat in a pixel. Following discussion with the data providers, these rasters have been created at 500m resolution.

The key steps involved in the production of the individual habitat rasters is outlined here:

1. In the combined vector basemap for the whole OxCam Arc, an additional column was created for each habitat of interest and each polygon was assigned a 1 or 0 to identify the presence or absence of the habitat.

2. A 10m resolution raster was created for each habitat in turn, snapped (spatially aligned) to the overall raster. Each 10m pixel therefore displayed 1 or 0 depending upon if the habitat was the majority habitat in that pixel or not.

3. Each high resolution raster was then aggregated to the desired cell size (resolution). This was based on the mean value of the input cells (the mean of each 10m pixel). Hence for each pixel in the output raster a score was calculated between 0 and 1, showing how common the habitat was. These rasters were also snapped to the overall raster, so that all of these individual habitat rasters were spatially aligned, with pixels exactly matching.

The habitats for which rasters were created tended to focus on biodiversity rich habitats that were rarer across the study area, so may be overlooked in the main rasters. They were selected with reference to habitat statistics calculated from the basemap and in conjunction with the OxCam LNCP team and their technical group. Rasters were created for the habitats listed in Table 4.

Table 4. Habitats for which raster maps were created and their constituent detailed habitats.

Mapped habitat	Detailed habitats included and Ph1code
Acid grassland	Unimproved acid grassland (Ph1code B11) and Semi-improved acid grassland (B12)
Neutral grassland	Unimproved neutral grassland (B21) and Semi-improved neutral grassland (B22)
Calcareous grassland	Unimproved calcareous grassland (B31) and Semi-improved calcareous grassland (B32)
Floodplain grazing marsh	Floodplain grazing marsh (B4f)
Fens	Flush and spring, basic (E22), Valley mire (E31), Flood-plain mire (E33) and Lowland fen (E3/F1)
Reedbeds	Swamp (F1)
Heathland	Heath (unknown type) (Du), Wet dwarf shrub heath (D2), acid grassland, dry heath mosaic (D5) and acid grassland, dry heath mosaics with trees or scrub (D5_Au and D5_Bu_Au)
Scrub	Scrub (unknown) (A2), Dense/continuous Scrub (A21) and Scattered scrub (A22)
Orchards	All orchards (traditional and non-traditional) (A112o)

Appendix 1: Processing steps to produce the basemap

Step 1: Prepare input data sets
All data sets were prepared for use. The MasterMap Topography Layer was adjusted so that field names were consistent, and a number of other small adjustments were made to enable automatic processing in EcoServ-GIS. A DTM was created for each county by loading each individual tile into a raster catalogue and then creating a mosaic for the whole area. Preparation for the other input layers is described in the steps below. All input layers were checked for duplicates and geometry was checked and repaired. All were projected using British National Grid.

Step 2: Create initial basemap framework
A study area boundary was created based on Boundary-Line data and a 1km buffer was created around this. The buffer is not really necessary when mapping habitats for the natural capital baseline, but is useful when modelling ecosystem services in follow-on work as models routinely consider the area around habitats as well as the habitat itself. Creating a basemap with a buffer ensures that all ecosystem service models will be accurate to the edge of the study area (note that simple metric approaches do not require a buffer as they do not consider the effect of distance). The MasterMap Area dataset was used as the initial input and was clipped to the study area plus 1km buffer boundary. A large number of blank fields were then added, which are filled in subsequent steps. Pylons and other structures above ground level were removed, as were overlapping polygons and very small polygons (<5 m^2).

Step 3: Split basemap
The basemap was split into chunks (in this case 2 chunks) to speed up processing in the subsequent steps and avoid issues of the programme freezing or crashing. Chunking is a well-used approach in analysing very large GIS data sets. The Oxford initial basemap had more than 1.5M polygons and we would recommend that chunking is done on any datasets with more than 1M polygons, or considerably less if a computer with low specifications is being used.

Step 4: Add greenspace typology to the basemap (OS MasterMap Greenspace and OS Open Greenspace data)
A new column was added to the OS MasterMap Greenspace data and each polygon was classified to a final function. In addition, a column was added indicating if the polygon was publicly accessible and this was also assigned by a set of simple rules. For example, "Amenity – Residential or Business" and "Amenity – Transport" were both classified as amenity greenspace with public access, "Institutional Grounds" and School Grounds" were classified as amenity greenspace but with no public access. In all cases, those with a Primary Form of "Manmade Surface" were removed from selections.

A similar process was undertaken with the OS Open Greenspace data. Data was only selected from this data set for locations where there was not OS MasterMap Greenspace data. The OS MasterMap Greenspace data only covers larger urban areas, so OS Open Greenspace was used to identify greenspaces in village and rural locations such as country parks. The two non-overlapping data sets were then merged and used as an input into the basemap. This greenspace layer was overlain onto the basemap and where a greenspace covered the majority (>50%) of a basemap polygon, each respective basemap polygon was classified with the relevant greenspace typology and access information.

Step 5: Add Natural England PHI (Priority Habitats Inventory) data to the basemap
A new column was added to the PHI data and each habitat type was classified into a habitat type that was consistent with the habitat classification being used in the basemap. The PHI vector data were then converted into a high resolution (2m) raster with each habitat type being given a different value. The percentage area of each MasterMap polygon that is covered by each habitat type raster was then calculated using zonal statistics and these data are added to the Basemap habitat fields.

For Bedfordshire the habitat data supplied by the Records Centre was compared to the PHI data. There was considerable overlap but also a number of polygons that appeared in one data set but not the other. The areas identified in the PHI data set that were not in the Bedfordshire data set were extracted and sent to the Bedfordshire and Luton Biodiversity Recording and Monitoring Centre for checking. Sites that were considered sensible to include from the PHI dataset were then combined with the Bedfordshire data and the process described in the previous paragraph was followed.

Step 6: Add CEH LCM 2015 data to the basemap
The process followed was very similar to PHI (above). A new column was added to the LCM layer and each habitat type in the land cover map was classified into a habitat code. The layer was then converted into a 5m resolution raster and the percentage area of each habitat type in the raster was calculated for each polygon in the basemap.

Step 7: Add Built-up Areas boundary data to the basemap
This layer was added to the basemap by simply calculating and recording if each polygon in the basemap fell within the built-up area or not.

Step 8: Add AWI (Ancient Woodlands Inventory) data to the basemap
This step determines the percentage overlap between each polygon in the Basemap with the AWI layer following the same method as for the PHI and LCM data.

Step 9: Add Phase 1 habitat data to the basemap (Oxfordshire only)
This followed a similar procedure to the PHI and LCM data. The TVERC habitat layer was converted into two fine scale 2m rasters, one based on Phase 1 habitat types, and one based on Habitats of Principle Importance, and zonal statistics were calculated on the overlay between the Basemap and the habitat rasters. Due to the large number of habitat types present (69 Phase 1 habitat types and an additional 23 Habitats of Principal Importance), rather than recording the percentage cover of each habitat type in each basemap polygon, habitats were assigned to a basemap polygon based on the majority habitat in each polygon.

Step 10: Classify all basemap polygons
This is the most important step in the basemapping process and involved applying a series of rules to determine the final habitat classification of each polygon in the basemap. Initially habits are assigned based purely on attributes within MasterMap. For example, domestic buildings (code J3.6.0) are identified based on size criteria, and private gardens (J5.6) are identified based on size and proximity to houses. Woodland codes are assigned based on the presence of broadleaved or coniferous trees recorded in MasterMap Descriptive Term. Mastermap is good at recording trees, water and urban features, but is much more vague about non-urban non-wooded habitats, which are often called "general surface" or "agricultural land".

This first assignment of habitats is then adjusted using information obtained from the other data sets using a further series of rules and priorities. The final classification of habitat will depend on the data available for the particular basemap being produced. Many of these rules were applied using EcoServ-GIS and these are described in detail in the EcoServ-GIS technical report[3].

As EcoServ-GIS does not enable Phase 1 habitat data to be properly incorporated, a further series of rules were developed to include this data. We ran these after running the EcoServ-GIS rules outlined above, to alter the initial habitats applied, in light of the additional data. For Oxfordshire, the habitat data provided by TVERC was up-to-date and considered reliable so was given precedence over the other data sources in non-urban areas. However, it is less detailed than MasterMap, often recording a block containing other features (such as roads) the same as the land. Hence a series of selections were used so the TVERC habitat was applied only where a number of conditions were met (see Box 3 for an example).

Box 3: Example of a selection rule applied to assign habitats:

TVERC habitat = 27 AND Make = 'Natural' AND NOT Theme = 'Water' AND NOT Theme = 'Roads Tracks And Paths' AND HabCode_B NOT LIKE '%A%'

This would select all polygons where the overlay with the TVERC data set had suggested were habitat 27, but only if the surface was definitely natural, was not water, was not a road, path or track, and MasterMap had not recorded trees being present. These polygons were then recorded as habitat 27, which was unimproved neutral grassland and given the final habitat code of B2.1.

A second selection was then run to obtain the polygons identified as 27 but that had been recorded as containing trees (and not water, manmade or tracks). These were then assigned a final habitat of unimproved grassland with trees or scrub (code Bu1/A2,A3).

Polygons that were not natural, or were water or tracks were not altered and their original code was retained. These rules were applied for each TVERC habitat type in turn.

The locally derived habitat data for Bedfordshire was much less extensive or up-to-date than for Oxfordshire (4.5% coverage for Bedfordshire compared to 87.7% for Oxfordshire), but was amalgamated into the PHI data. The neutral grassland data sets were assessed after a habitat had been assigned to each polygon and were used more for manual checking and resolving habitats where other data was missing or contradictory.

Step 11: Adjust woodland habitats in basemap using NFI (National Forest Inventory) data
The NFI dataset is considered to provide reliable and up-to-date information on woodland type. Hence it was used to adjust the woodland habitat type for polygons where information was based purely on MasterMap data. Only the broadleaved, conifer and mixed categories from the NFI layer were used. It was not required for sites that had previously been classified using additional data into categories such as semi-natural broadleaved woodland (A1.1.1). This step was also not possible within EcoServ-GIS, so was run manually to adjust polygons after habitats had been initially assigned as described above.

---

[3] Winn, J.P., Bellamy, C.C. & Fisher, T. 2018. EcoServ-GIS: a toolkit for mapping ecosystem services. Technical Report: ES1 BaseMap. *Scottish Natural Heritage Research Report No. 954.*

Step 12: Manual checking of basemap

Basemaps produced using the approach described here use the best available data, but will always be subject to some error. It is not feasible to check the whole of a basemap, but if time and resources allow, it is worthwhile carrying out some checks. It can be particularly fruitful to engage local stakeholders in this process who know the area really well. This was not possible for Oxfordshire or Bedfordshire, but a short series of checks were still carried out. This was focussed around identifying any polygons that had received a null classification and in sites with uncertain classifications (for example, B4/J11 is a classification applied to some sites where the habitat is not properly known). These sites were manually checked both by examining all the input data layers to determine the most likely habitats and by examining aerial photos on Google Maps.

Step 13: Add attributes table to the basemap

Once the basemap habitats have been fully classified a series of attributes can be added to the basemap table for each polygon. This can include broader habitat classifications or habitat specific scores for use in ecosystem services mapping (e.g. the mean carbon storage value for each habitat type) or any other attributes required. A separate table showing these attributes was created and was then joined to the basemap.

Step 14: Presenting the final habitat basemap

The final step was to create the final basemap for sharing. The different parts (chunks) of the basemap were merged together, the map was clipped to the county boundary and the extraneous fields used to classify the habitats were deleted to reduce the size of the final product. The geodatabase was converted to a shapefile to enable sharing across GIS platforms. Two layer files were created to enable the map to be displayed using a consistent symbology across different maps, and by different project partners. One was created to display the most detailed Phase 1 habitat types and one was created for the broader HabType2 grouping.

Appendix 2: All habitats occurring across the OxCam Arc and the classification hierarchy

Ph1code	HabNmPLUS	HabBroad	HabType2
A11	Woodland,Broadleaved,(unknown type)	Woodland, broadleaved	Broadleaved woodland
A11/A2	Woodland,Broadleaved,(unknown type)/ with scrub	Woodland, broadleaved	Broadleaved woodland
A111	Woodland,Broadleaved,Semi-natural	Woodland, broadleaved	Broadleaved woodland
A112	Woodland,Broadleaved,Plantation	Woodland, broadleaved	Broadleaved woodland
A112o	Woodland,Broadleaved,Plantation,(orchard)	Woodland, broadleaved	Broadleaved woodland
A12	Woodland,Coniferous,(unknown type)	Woodland, coniferous	Coniferous woodland
A12/A2	Woodland,Coniferous,(unknown type)/ with scrub	Woodland, coniferous	Coniferous woodland
A121	Woodland,Coniferous,Semi-natural	Woodland, coniferous	Coniferous woodland
A122	Woodland,Coniferous,Plantation	Woodland, coniferous	Coniferous woodland
A13	Woodland,Mixed,(unknown type)	Woodland, mixed	Mixed woodland
A13/A2	Woodland,Mixed,(unknown type)/ with scrub	Woodland, mixed	Mixed woodland
A131	Woodland,Mixed,Semi-natural	Woodland, mixed	Mixed woodland
A132	Woodland,Mixed,Plantation	Woodland, mixed	Mixed woodland
A2	Scrub,(unknown),	Scrub	Scrub
A21	Scrub,Dense/continuous,	Scrub	Scrub
A22	Scrub,Scattered,	Scrub	Scrub
A31	Parkland/scattered trees,Broadleaved,	Trees / Parkland	Trees / Parkland
A31/A2	Parkland/scattered trees,Broadleaved,/ with scrub	Trees / Parkland	Trees / Parkland
A32	Parkland/scattered trees,Coniferous,	Trees / Parkland	Trees / Parkland
A32/A2	Parkland/scattered trees,Coniferous,/ with scrub	Trees / Parkland	Trees / Parkland
A33	Parkland/scattered trees,Mixed,	Trees / Parkland	Trees / Parkland
A33/A2	Parkland/scattered trees,Mixed,/ with scrub	Trees / Parkland	Trees / Parkland
A3p	Parkland/scattered trees,,(parkland)	Trees / Parkland	Trees / Parkland
Au_SN	Woods/Trees/Scrub,,with semi-natural habitats	Mixed habitats	Mixed / other / uncertain
B11	Grassland,Acid,Unimproved	Grassland, semi-natural	Semi-natural grassland
B12	Grassland,Acid,Semi-improved	Grassland, semi-natural	Semi-natural grassland
B21	Grassland,Neutral,Unimproved	Grassland, semi-natural	Semi-natural grassland
B22	Grassland,Neutral,Semi-improved	Grassland, semi-natural	Semi-natural grassland
B31	Grassland,Calcareous,Unimproved	Grassland, semi-natural	Semi-natural grassland
B32	Grassland,Calcareous,Semi-improved	Grassland, semi-natural	Semi-natural grassland
B4	Grassland,Improved,	Grassland, improved	Improved grassland
B4/J11	Grassland,Improved,/arable (probable)	Uncertain agriculture (improved grass or arable)	Uncertain agriculture
B4f	Grassland,Improved,(floodplain/grazing marsh)	Grassland, improved	Marshy grassland
B5	Grassland,Marshy,	Grassland, marshy	Marshy grassland
B5/E3/F/H2	Grassland,Marshy,/Fen/Swamp/Saltmarsh	Grassland, marshy	Marshy grassland
B5/E3/F/H2_Bu	Grassland,Marshy,/Fen/Swamp/Saltmarsh and Rough grassland	Grassland, marshy	Marshy grassland
B6	Grassland,Poor,Semi-improved	Grassland, semi-natural	Semi-natural grassland
B6/J3	Grassland,Poor,Semi-improved/Ephemeral/short perennial	Uncertain	Mixed / other / uncertain
Bu	Grassland, rough (probable semi-improved)	Grassland, semi-natural	Semi-natural grassland
Bu_A11	Grassland,(unknown),/ with broadleaved trees	Grassland, semi-natural	Semi-natural grassland
Bu_A2/A3	Grassland,(unknown),/ with scrub,trees	Grassland, semi-natural	Semi-natural grassland
Bu_Au	Grassland,(unknown),(unknown type)_ with wood, scrub or trees	Grassland, semi-natural	Semi-natural grassland
Bu1/A11,A2	Grassland,(unknown),Unimproved/ with broadleaved trees or scrub	Grassland, semi-natural	Semi-natural grassland
Bu1/A2	Grassland,(unknown),Unimproved/ with scrub	Grassland, semi-natural	Semi-natural grassland
Bu1/A2,A3	Grassland,Unimproved/ with scrub,trees	Grassland, semi-natural	Semi-natural grassland
Bu1/Bu2	Grassland,,(semi-improved or unimproved) (rough grassland)	Grassland, semi-natural	Semi-natural grassland
Bu2	Grassland,(unknown),semi-improved (GQ)	Grassland, semi-natural	Semi-natural grassland
Bui	Grassland,(unknown),(probably improved)	Grassland, unknown	Mixed / other / uncertain
BuiUrb	Grassland,(unknown),(probably improved)[urban]	Grassland, unknown	Mixed / other / uncertain
Buu	Grassland,(unknown),(unknown type)[urban]	Grassland, unknown	Mixed / other / uncertain

Buu/C31	Grassland,(unknown),(unknown type)/ with Tall ruderal (rail verge)[urban]	Grassland, unknown	Mixed / other / uncertain
C11	Bracken,Continuous,	Other	Mixed / other / uncertain
C12	Bracken,Scattered,	Other	Mixed / other / uncertain
C31	Other,Tall ruderal,	Other	Mixed / other / uncertain
C32	Other,Non-ruderal,	Other	Mixed / other / uncertain
D2	Wet dwarf shrub heath,,	Heathland	Heathland
D5	Mosaic: acid grassland, dry heath,,	Heathland	Heathland
D5_Au	Mosaic: acid grassland, dry heath,,/ with Woods, Trees, Scrub	Heathland	Heathland
D5_Bu_Au	Mosaic: acid grassland, dry heath,,/ rough grassland / with Woods, Trees, Scrub	Heathland	Heathland
Du	Heath (unknown type),,	Heathland	Heathland
E22	Flush and spring,Basic,	Mire	Fen, marsh and swamp
E3/F1	Fen,,(lowland fen)	Mire	Fen, marsh and swamp
E31	Fen,Valley mire,	Mire	Fen, marsh and swamp
E33	Fen,Flood-plain mire,	Mire	Fen, marsh and swamp
F1	Swamp,,	Swamp and marginal	Fen, marsh and swamp
Fu	Swamp/marginal/inundation,(unknown),	Swamp and marginal	Fen, marsh and swamp
G	Water (inland),(unknown),	Water, fresh	Water
G26	Running water,Brackish / Tidal,	Water, brackish	Water
H11	Intertidal,Mud/sand,	Intertidal	Mixed / other / uncertain
I1	Natural rock,(unknown),	Natural rock	Rock, exposure and waste
I14b	Natural rock,Other exposure,boulders	Natural rock	Rock, exposure and waste
I21	Artificial rock/exposure/waste,Quarry,	Artificial exposure / waste	Rock, exposure and waste
I22	Artificial rock/exposure/waste,Spoil,	Artificial exposure / waste	Rock, exposure and waste
I24	Artificial rock/exposure/waste,Refuse-tip,	Artificial exposure / waste	Rock, exposure and waste
J11	Cultivated/disturbed land,Arable,	Cultivated / disturbed land	Cultivated / disturbed land
J11t	Cultivated/disturbed land,Arable,(Allotments)	Cultivated / disturbed land	Cultivated / disturbed land
J12	Cultivated/disturbed land,Amenity grassland,	Grassland, amenity	Improved grassland
J12v	Cultivated/disturbed land,Amenity grassland,(road verge)	Grassland, amenity	Improved grassland
J13	Cultivated/disturbed land,Ephemeral/short perennial,	Cultivated / disturbed land	Cultivated / disturbed land
J360	Built-up area,Buildings,(domestic)	Built up areas	Built-up areas and infrastructure
J361	Built-up area,Buildings,Business or Industry	Built up areas	Built-up areas and infrastructure
J362	Built-up area,Buildings,Shed/Garage/Farm building	Built up areas	Built-up areas and infrastructure
J363	Built-up area,Buildings,Structure	Built up areas	Built-up areas and infrastructure
J364	Built-up area,Buildings,Glasshouse	Built up areas	Built-up areas and infrastructure
J36u	Built-up area,Buildings,(unknown type)	Built up areas	Built-up areas and infrastructure
J37	Built-up area,Sealed surface,	Built up areas	Built-up areas and infrastructure
J4	Bare ground,,	Other	Mixed / other / uncertain
J5	Other habitat,,	Other	Mixed / other / uncertain
J511	Other habitat,Road,Surfaced	Roads	Built-up areas and infrastructure
J512	Other habitat,Road,Unsurfaced	Roads	Built-up areas and infrastructure
J52	Other habitat,Roadside/Pavement,	Pavement	Built-up areas and infrastructure
J53	Other habitat,Railway,	Railway	Built-up areas and infrastructure
J54	Other habitat,Path, sealed,	Path	Built-up areas and infrastructure
J55	Other habitat,Probable garden/brownfield or park,	Gardens / Parks / Brownfield	Garden
J55Urb	Other habitat,Probable garden/brownfield or park,[urban]	Gardens / Parks / Brownfield	Garden
J56	Other habitat,Private garden,	Garden	Garden
J56Urb	Other habitat,Private garden,[urban]	Garden	Garden
Linear	Linear habitats	Other	Mixed / other / uncertain
Unclassified	Unclassified	Unclassified	Unclassified