OxCam LNCP Natural Capital Baseline Assessment: Comparing habitat basemaps for Oxfordshire

Authors:

Jim Rouquette[1], Alison Holt[1] & Alison Smith[2]

[1]Natural Capital Solutions and [2]University of Oxford

Reviewed by:

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

Introduction

The first and most fundamental step in creating a local natural capital plan is to understand the natural capital assets of an area. Information is required on the type, extent, location and condition of natural capital assets and mapping the habitats is the first key step in this process. This is an important output in its own right but is also a necessary prerequisite to assessing the benefits that arise from natural capital, their value, and opportunities to enhance natural capital assets and place them at the heart of decision making.

Data on habitats and land use is becoming increasingly available form a wide range of sources, but varies in quality, age, spatial resolution and cost. OS MasterMap Topography Layer provides the most detailed mapping available across the whole of Great Britain, with every parcel of land mapped, but includes only limited data on habitats. The most promising approaches to habitat mapping developed over the last few years have therefore taken this highly detailed underlaying basemap (OS MasterMap) and used a series of additional data layers to classify each polygon into an accurate habitat type. This relies on a series of rules developed to assign habitats and can utilise different data sources. However, as yet there has not been an assessment of the differences that may occur when creating maps using different rules, and the impact of using different data sources on the final output.

This report compares three habitat basemaps of Oxfordshire that have been compiled using slightly different techniques and combinations of spatial data, to explore how similar they are, and where the differences lie. The three basemaps are the Natural Capital Solutions (NCS) basemap, the Habitats and Land Use (HLU) basemap created by Alison Smith at the University of Oxford, and another version of the HLU basemap that does not use one of the key data sources (Phase 1 habitat and land use survey) that we refer to here as the LCM basemap. Although it is not possible to use one map as the definitive “correct” basemap with which to compare the others, it is possible to determine how closely the maps match, where the differences lie, and make some reasoned assumptions about processes to increase accuracy.

Basemapping approaches

The basemapping procedure layers different spatial data sets together that give information on habitat types across a specified landscape. Using one layer as a template, each polygon within it is assigned the most appropriate habitat from the spatial data layers, using a set of rules prioritising which is the most up to date, accurate or reliable source. All three basemaps have been created in this way and use OS MasterMap as the base layer polygons to which habitats are assigned. The NCS basemapping process uses a different set of rules to the HLU and LCM maps, and they all use slightly different combinations of spatial data (see Tables 1 & 2). The HLU (and LCM) procedure was developed specifically for Oxfordshire before being applied elsewhere, so is set up to match the habitat classification and data input available in Oxfordshire, whereas the NCS approach was developed outside of Oxfordshire but was designed to be used anywhere in GB.

Locally derived data were available from the Thames Valley Environmental Records Centre (TVERC) in the form of a Phase 1 habitat and land use survey. This mapped most fields in the county based on the JNCC Phase One Habitat Classification Scheme, as well as NERC Act Section 41 Habitats of Principal Importance (previously known as BAP habitats), and was derived from a mixture of field survey data and aerial photograph interpretation. The NCS and HLU maps used this data set, but the LCM map did not. All maps also utilised a (slightly different) range of national data sets as outlined in Table 2.

Table 1. Summary of approach used to create the three basemaps

Basemap	Approach
NCS	Modified version of EcoServ-GIS V3.3, with semi-manual addition of Phase 1 habitats. Habitats assigned to OS MasterMap polygons.
HLU	Bespoke automated process. Habitats assigned to OS MasterMap polygon boundaries, but split to create new shapes where Phase 1, Priority Habitats and Designations followed different boundaries.
LCM	Bespoke automated process. Habitats assigned to OS MasterMap polygon boundaries, but split to create new shapes where LCM, Priority Habitats and Designations followed different boundaries. i.e.) same as for HLU, but replacing TVERC Phase 1 habitat and land use survey with CEH Land Cover Map and Priority Habitats, and omitting local designations.

Table 2. Data layers used to create each of the three basemaps

Data layers	NCS	HLU	LCM
OS MasterMap Topography Area Layer	✓	✓	✓
OS MasterMap Greenspace	✓	✓	✓
OS Open Greenspace	✓	✓	✓
CEH Land Cover Map 2015	✓		✓
Natural England Priority Habitats Inventory	✓		✓
TVERC Phase 1 habitat and land use survey	✓	✓	
National Forest Inventory	✓		
Ancient Woodland Inventory	✓		
Built-up Area Boundaries	✓		
Digital Terrain Model	✓		
Agricultural Land Class		✓	✓
Conservation designations (e.g. SSSI), including local designations from TVERC (Local Wildlife Sites, Local Geological Sites and Road Verge Nature Reserves)		✓	
Conservation designations (not including local designations)			✓
Hedges		✓	✓
Ancient trees		✓	✓

The total number of polygons in the three maps were as follows: NCS map had 1,375,329 polygons, the HLU map had a total of 1,425,288 polygons, and the LCM map contained 1,409,428 polygons. The differences were partly due to polygon splitting in the HLU map (and to a lesser extent in the LCM version), but was also linked to an issue with missing data in the MasterMap Area input data for the NCS map[1]. This makes it difficult to compare the number of polygons between the NCS map and the other two versions. However, the HLU map contained an additional 31,417 polygons compared to the underlying OS MasterMap, as a result of splitting OS MasterMap polygons to incorporate the TVERC Phase 1 habitat data.

Note that local habitat and site designation boundaries are often drawn from aerial photos and do not match OS Mastermap boundaries, so that if these datasets are simply intersected, hundreds of thousands of tiny 'slivers' appear along polygon edges. The procedure for splitting polygons developed at the University of Oxford[2] almost eliminates these slivers, but a few thousand still remain. Therefore, some of the 31,417 extra habitat polygons created when adding TVERC Phase 1 habitat data to OS Mastermap in the HLU dataset will be edge slivers rather than genuine additional habitat features. A further 13,291 polygons were created in the HLU dataset from merging in Agricultural Land Class, and 20,684 more from merging in habitat conservation designations.

Harmonising habitat categories

The habitat categories used in each of the basemaps differ, even though all use primary Phase 1 Habitat Survey categories. The NCS basemap for Oxfordshire has 79 different habitat categories. The HLU map has more categories at 95, and the LCM has the least at 64. There is therefore a difference between all of the basemaps in the detail to which habitats are categorised, due to the different spatial information used to create the maps, but also differences in the rules that assign the habitats to polygons. The HLU classification appears to break habitats down into more detailed categories in some cases. For example, the NCS basemap has one category of freshwater, whereas the HLU has standing water and running water (both from Phase 1 data), reservoir, drain and canal (from OS Mastermap) and water, where no habitat data is available to split between standing and running water. The HLU has more detail for habitats such as semi-natural grasslands with trees, orchards, and open mosaic habitats. On the other hand, the NCS basemap includes additional detail for woodland with scrub, parkland with scrub, fen, ruderal vegetation, it separates marshy grasslands from floodplain grazing marsh, and includes six different categories of buildings (e.g. separating domestic buildings from business or industry and so on) instead of the one in the HLU basemap.

For some types of green infrastructure taken from the OS Greenspace datasets, e.g. golf courses, cemeteries, allotments and recreation grounds, the main habitat interpretation column in the HLU basemap retains specific habitats (e.g. buildings, water, woodland) but replaces grass or unknown natural

[1] During checking of the NCS version of the basemap, it was discovered that a large number of small gaps existed, spread out in different parts of the map. This was traced back to the version of the OS Mastermap Topography Layer supplied to the project by Defra. Defra were unable to supply a new copy as their master version was also corrupted with these gaps, but an alternative version of the MasterMap Area layer was found without these gaps. The majority of the missing polygons were extracted, then classified as described above and merged with the previous version. The extraction process was not straightforward, so very small gaps do remain, but in total 34,001 polygons were added during this process.

[2] The University of Oxford team would like to acknowledge the valuable work of Martin Besnier, a visiting student from the Université Paris Sud, who designed the first version of the polygon splitting method.

surfaces with the green infrastructure type. Thus the grassed areas of golf courses will be classed as 'golf course', while any areas of woodland or water will be mapped separately. All the underlying habitat classifications (from OS MasterMap, Phase 1 habitat, OS Open Greenspace etc) contributing to the final interpreted habitat type are retained in separate columns. In contrast, the NCS map records the actual habitats of these green space areas (e.g. amenity grassland) and records the type of green infrastructure in a separate column.

The LCM map uses a simpler list of habitats than either of the other maps, for example using the term neutral grassland where the NCS map divides these into unimproved neutral grassland and semi-improved neutral grassland, and the HLU map uses all three terms (because the underlying Phase 1 habitat dataset includes some indeterminate neutral grassland as well as unimproved and semi-improved).

How well do the habitat basemaps match?

The three basemap habitat categories were combined and coded. To test how well the NCS, HLU and LCM habitat basemaps agree, a confusion matrix (also known as an error matrix) approach was used. This is the standard way to assess the accuracy of maps (e.g. Foody 2004[3], FAO 2016[4]) or, more widely, the level of agreement between any two different assessments of the same thing. Confusion matrices were run for each combination of basemap (NCS / HLU, NCS / LCM, and HLU / LCM) in ArcGIS. The confusion matrix process allocated c.5,000 stratified random assessment points across the basemap extent, each point was then assigned a habitat from map 1 and map 2 (e.g. NCS and HLU). A matrix was then created that compared the habitat assigned by one map to that of the second map. Scores were given for each habitat type and the overall proportion of points that match exactly, along with a Kappa statistic. The Kappa statistic is a measure of the agreement, taking into account the amount of agreement that would be expected by chance alone. Kappa is standardized on a -1 to 1 scale, where 1 is perfect agreement, 0 is exactly what would be expected by chance, and negative values indicate agreement less than chance.

When comparing the NCS basemap with the HLU, the overall proportion of points that matched was 0.75, and the Kappa statistic was 0.67. A 75% agreement is considered quite good between the NCS and HLU basemaps, and a Kappa statistic between 0.61 and 0.8 is considered to indicate "substantial agreement"[5]. Most of the differences that do occur can be attributed to where the habitat categories used in the basemaps do not align. For example, 91% of the points given the broad category of agricultural land in the HLU map were classified as either arable or improved grassland in the NCS map. Running water and standing water in HLU have been classified as a broader category of water in NCS, so they do not align simply due to the different classifications used. Habitat categories showing very high agreement (>90%) between the two maps included arable, building, road, road verge, manmade path, track, garden, marshy grassland, heath with scattered broadleaved trees, and semi-improved acid grassland. Several of these habitats are identified based solely on MasterMap, although some require additional information from other data sources. Habitats that showed intermediate levels of agreement and which occur regularly include the woodland categories, scrub categories and some of the grassland categories. The habitats

[3] Foody (2004) Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. Photogrammetric Engineering & Remote Sensing, 70: 627–633.
[4] FAO (2016) Map Accuracy Assessment and Area Estimation: A Practical Guide. National forest monitoring assessment working paper No.46/E, Food and Agriculture Organization of the United Nations, Rome.
[5] Viera A.J. and Garrett J.M. (2005) Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine, 37: 360-363.

showing the lowest agreement tend to be categories that occur rarely, along with the "natural surface" category. This is a catch-all habitat category used in the HLU map, which is not used at all in the NCS map, so there is no agreement. It covers 2,500 ha (1% of Oxfordshire) in the HLU map, for areas classed just as 'Natural surface' in OS Mastermap and with no Phase 1 or greenspace data. In the majority of cases (84%) this habitat is classified as improved grassland in the NCS map. Related to this, almost all of the points classified as improved grassland or amenity grassland in the HLU map are classified the same in the NCS, but the NCS map has a large number of additional improved grassland and amenity grassland points, which are classified differently (most often as natural surface) in the HLU map. Another area of significant difference occurs in the four "parkland and scattered trees" categories, which are used in both the NCS and HLU maps but to a greater extent in the NCS map, and the three "scattered trees" categories which are only used in the HLU map (for areas where OS Mastermap identifies 'scattered trees' but there is no information on what type of surface these trees are on). There are clear differences in how these habitats are assigned and how mixed habitats in general are classified.

When comparing the NCS basemap with the LCM map the matching proportion was considerably lower at 0.59, as was the Kappa statistic (0.48). This means that 59% of the points tested were classified the same when comparing the two maps and taking chance agreement into account, the Kappa statistic indicates only "moderate agreement". Habitat categories showing very high agreement (>90%) between the two maps included building, road, road verge, manmade path, track, garden, reedbed, and heath with scattered broadleaved trees. Most of these habitats had also been in agreement between the HLU and NCS maps (above) and are based on MasterMap data. Differences were more apparent in the natural habitats. For example, of the points classified as semi-improved neutral grassland in the NCS basemap, 33% were classified as improved grassland in the LCM map, 17% as amenity grassland, 14% as natural surface, and 36% were classified into 10 other habitat categories. Note that when these points were compared to the HLU map in the previous comparison, 61% were classified as semi-improved neutral grassland (an exact match), 38% as neutral grassland (a broader but appropriate category) and only 1% were classified in a completely different habitat. This is because both the HLU and NCS maps use TVERC Phase 1 habitat data for this habitat type.

When comparing the HLU map with the LCM version, the matching proportion was intermediate compared to the previous two comparisons, at 0.67, as was the Kappa statistic (0.58). Habitat categories that showed very high agreement (>90%) between the two maps included, building, sealed surface, road, road verge, manmade path, track, garden, and heath with scattered broadleaved trees, similar to the previous comparisons, but also running water, standing water, drain, natural sports facility, and allotment. Errors are larger in most of the more natural habitat types, partly caused by classification differences (e.g. semi-natural broadleaved woodland classified as the broader category of broadleaved woodland in 34% of occurrences), but many more significant differences are also apparent (e.g. 35% of these semi-natural broadleaved woodlands are classified as habitats that are not broadleaved woodland, most of which are not woodland at all). The match between grassland habitat types was considerably worse. For example, habitats classified as semi-improved neutral grassland by the HLU map were classified as improved grassland on 34% of occurrences in the LCM map, arable 18%, natural surface 14%, amenity grassland 14%, and a further seven different habitats for the remaining occurrences.

Overall, the NCS basemap matches with the HLU basemap to a greater extent than either do with the LCM basemap. Where the NCS and HLU maps do not match in their classification of habitat and land cover, it is largely due to the different habitat categories used, and for the most part it comes down to broad versus

more detailed categorisations of the same habitat type, although this is not always the case, as outlined above. When the LCM is compared to the other two maps, these classification differences again account for some of the differences, but there are far more occurrences where completely different habitats have been assigned. This analysis accounts for how well the habitat classifications match, but it does not take into account how the basemaps may differ in their estimation of the area of each habitat. This is important as it has an impact on the estimates of ecosystem service provision.

Comparing habitat type and extents

The three basemap habitat categories were combined and coded, so the frequency that each habitat was assigned to a polygon, the area of each habitat, and the % of the total area of Oxfordshire covered by each habitat could be compared for each basemap (Table 3). Once again, comparisons are difficult due to differences in the habitat classifications used. Comparing the NCS and HLU maps first, the habitats where the basemaps predict almost the same area and percentage cover include semi-improved acid grassland, reedbed, allotments, landfill, manmade path, rail, tracks, sealed surface, and roadside – manmade (pavements). Most (but not all) of these habitats are classified directly from MasterMap. There are 24 habitats / land covers where the HLU basemap predicts an area that is at least 100 ha (0.05% of the total area of Oxfordshire) greater than the NCS map and 8 of these are more than 1000 ha higher. Examples of these categories are: broadleaved, coniferous and mixed woodland, orchard, dense scrub, scattered tree categories, neutral grassland, marshy grassland, agricultural land, golf course, and sports facilities. The largest difference is for neutral grassland, however this is because all neutral grasslands in the NCS basemap are classified as unimproved or semi-improved and the differences are smaller (although still present) if the categories are merged. Similarly, many of the difference in the other categories are significantly reduced when considering broader categories (e.g. merging all mixed woodland categories together results in almost identical overall amounts of mixed woodland recorded in the two basemaps). The NCS map does not classify function such as golf courses and sport facilities within the main basemap, but in a separate column, and would record these as habitats such as amenity grassland or scattered trees. There are 21 habitats / land covers where the NCS basemap predicts a higher area than the HLU map by greater than 100 ha. Examples of these categories are: broadleaved, coniferous and mixed plantations, scrub (unknown), improved, unimproved calcareous, and amenity grasslands, and arable. The largest difference between the two maps occurs for amenity grassland, where the NCS map records an additional 4,344 ha (or 1.67% of total landcover) compared to the HLU map. This is because much of this is classified as greenspace categories such as ‘golf course’ or ‘recreation ground’ in the HLU map.

There are close matches between NCS and LCM basemaps for habitats including man-made path, rail, sealed surface, track, man-made roadside (pavement), landfill, and allotments, all habitats that are identified based on MasterMap and MasterMap Greenspace layers. When comparing the range of mismatch in area or percentage cover estimation between the NCS and the LCM basemaps there is a higher range of error than for the comparison with the HLU map. The LCM estimates a higher area by at least 100 ha for 24 habitat / land cover types, with 12 of these different by at least 1000 ha (>0.45% of total landcover). Examples of these categories are: semi-natural broadleaved woodland (the largest difference), coniferous and mixed woodland, scattered trees of all types, broadleaved parkland, semi-natural, neutral, and marshy grasslands, fen/swamp, standing and running water. Conversely, the NCS map estimates a higher area by >100 ha for 26 habitat / land use types, 10 of which differ by more than 1000 ha. Examples of these categories are: broadleaved woodland, broadleaved and coniferous plantation, mixed woodland,

scrub, parkland and scattered trees of each type, neutral semi and unimproved, calcareous and improved (largest difference) grassland, water, and road side features. Again, the different classifications used exaggerates the differences, although differences do remain and are larger than for the comparison between NCS and HLU maps.

The differences between the HLU and LCM basemaps are less than for the NCS and LCM basemaps. The HLU predicts a higher area for 13 habitats (>100 ha), with six greater than 1000 ha. Examples of these habitats and land covers are coniferous woodland, mixed woodland, semi-natural broadleaved woodland (largest difference), broadleaved parkland and scattered trees, semi-natural grassland, amenity grassland and agricultural land. The LCM basemap predicts a higher area and percentage cover for 19 habitats and land covers by over 100 ha, with six of these greater by over 1000 ha. These habitats include broadleaved, mixed and coniferous woodland plantations, semi-improved neutral grassland, and improved grassland (largest difference). Much of this is due to differences in the accuracy of the woodland classifications used.

Whilst many of the percentage differences between basemaps are low (the vast majority of habitats differ by less than 0.1% of the area of Oxfordshire), the larger differences such as 1.67%, 2.77%, 3.30% and 3.38% are significant in terms of the area of Oxfordshire (4,351, 7,217, 8,598, and 8,807 ha). The most significant discrepancies in terms of total area between the NCS basemap, the HLU and LCM basemaps lie in amenity, improved and neutral grasslands, and semi-natural broadleaved woodland.

Table 3. Combined habitat categories across the NCS, HLU and LCM basemaps, frequency each habitat code is assigned, area of each habitat and the % of the total area of Oxfordshire covered by each habitat.

		NCS			HLU			LCM		
Habitat	**Code**	**Frequency**	**Ha**	**% cover**	**Frequency**	**Ha**	**% cover**	**Frequency**	**Ha**	**% cover**
Broadleaved woodland	3	8252	6341.7	2.43	25033	7864.6	3.02	16209	4069.1	1.56
Broadleaved woodland - plantation	4	2051	1805.7	0.69	2608	1589.6	0.61			0.00
Broadleaved woodland - semi-natural	35	4817	6749.6	2.59	10056	6545.0	2.51	27198	15352.7	5.89
Coniferous woodland	21	1421	1773.1	0.68	3062	1899.0	0.73	3290	2396.5	0.92
Coniferous woodland - plantation	60	997	1564.6	0.60	1412	1400.4	0.54			0.00
Mixed woodland	37	950	1026.7	0.39	3591	1499.4	0.58	4726	3451.6	1.32
Mixed woodland - plantation	43	1436	1926.9	0.74	1697	1413.5	0.54			0.00
Unknown woodland - plantation	42			0.00	120	72.6	0.03			0.00
Orchard	63	777	281.3	0.11	134	34.4	0.01	211	113.5	0.04
Traditional orchards	30			0.00	1220	265.8	0.10	851	189.0	0.07
Unknown young woodland	41			0.00	906	745.4	0.29			0.00
Felled woodland	74			0.00	45	19.2	0.01			0.00
Scattered scrub	55	425	268.2	0.10	830	315.1	0.12			0.00
Scattered trees - broadleaved	26			0.00	2043	571.8	0.22	1691	719.0	0.28
Scattered trees - coniferous	67			0.00	189	39.9	0.02	170	49.1	0.02
Scattered trees - mixed	70			0.00	145	34.2	0.01	126	29.0	0.01
Scrub on semi-natural grassland	29	459	140.0	0.05	208	82.4	0.03			0.00
Scrub,(unknown),	105	1409	267.6	0.10			0.00			0.00
Dense scrub	28	1868	867.5	0.33	3933	985.1	0.38	1787	371.0	0.14
Parkland and scattered trees	39	1366	2086.5	0.80	86	83.7	0.03			0.00
Parkland and scattered trees - broadleaved	40	12563	2274.8	0.87	2391	2418.7	0.93	4288	5296.6	2.03
Parkland and scattered trees - coniferous	79	1089	173.2	0.07	9	4.5	0.00			0.00
Parkland and scattered trees - mixed	81	1686	281.7	0.11	33	9.6	0.00			0.00
Grassland,(unknown),(probably improved)	104	79	104.4	0.04			0.00			0.00
Grassland,(unknown),(unknown type)/ with Tall eral (rail verge)[urban]	106	670	66.7	0.03			0.00			0.00
Semi-natural grassland	36	67	143.3	0.05	546	133.5	0.05	1476	1760.8	0.68
Semi-natural grassland and scattered scrub	49	2066	1505.8	0.58	701	322.6	0.12	843	541.6	0.21
Semi-natural grassland with scattered trees – broadleaved	56			0.00	59	16.2	0.01	105	85.7	0.03

Semi-natural grassland with scattered trees – coniferous	109			0.00				4	1.4	0.00
Semi-natural grassland with scattered trees – mixed	93			0.00	3	0.2	0.00	3	2.2	0.00
Acid grassland	76			0.00	203	55.6	0.02			0.00
Acid grassland - semi-improved	90	6	0.7	0.00	4	0.7	0.00			0.00
Acid grassland - unimproved	77	39	29.0	0.01	15	1.7	0.00			0.00
Lowland dry acid grassland				0.00			0.00	159	163.5	0.06
Neutral grassland	19			0.00	2772	2646.4	1.02	780	1297.2	0.50
Neutral grassland - semi-improved	34	3880	3967.9	1.52	5930	5211.2	2.00			0.00
Neutral grassland - unimproved	44	548	1295.9	0.50	197	107.6	0.04			0.00
Neutral grassland and scattered scrub	88			0.00	2	3.8	0.00			0.00
Calcareous grassland	50			0.00	1578	859.3	0.33	1265	1210.8	0.46
Calcareous grassland - semi-improved	82	149	207.4	0.08	222	180.1	0.07			0.00
Calcareous grassland - unimproved	73	553	681.9	0.26	192	190.8	0.07			0.00
Improved grassland	6	92669	56331.7	21.62	60110	54670.0	20.98	33472	49115.6	18.85
Improved grassland and scattered scrub	57			0.00	297	117.5	0.05			0.00
Grassland, Marshy,	103	33	20.1	0.01			0.00			0.00
Marshy grassland	66	1555	3763.2	1.44	2827	3925.6	1.51	2516	4275.8	1.64
Marsh with scattered trees - broadleaved	15	109	75.6	0.03	71	21.2	0.01	41	13.5	0.01
Marsh with scattered trees - mixed	91			0.00	1	0.2	0.00			0.00
Poor semi-improved grassland	51	181	159.6	0.06	93	117.3	0.05			0.00
Bracken - continuous	62	4	1.9	0.00	34	19.0	0.01			0.00
Tall herb and fern	71	18	18.8	0.01	56	22.7	0.01			0.00
Heath with scattered trees - broadleaved	83	1	0.2	0.00	1	0.2	0.00	1	0.2	0.00
Heathland	89			0.00	21	5.5	0.00	11	1.8	0.00
Fen, marsh and swamp	16			0.00	252	27.0	0.01	190	59.5	0.02
Lowland fens	46	115	112.7	0.04	468	192.8	0.07	298	135.5	0.05
Swamp	45			0.00	10	6.4	0.00			0.00
Reedbed	53	150	56.8	0.02	190	54.1	0.02	143	31.7	0.01
Marginal vegetation	92			0.00	2	0.1	0.00			0.00
Wet woodland	52			0.00	552	137.0	0.05	5	1.1	0.00
Standing water	17			0.00	6838	1458.9	0.56	6934	1462.4	0.56
Running water	0			0.00	9100	1166.2	0.45	9806	1168.9	0.45
Water	65	18590	3103.2	1.19	185	8.5	0.00	179	9.0	0.00

Weir	38			0.00	80	0.4	0.00	83	0.4	0.00
Drain	18			0.00	4118	188.2	0.07	4169	188.4	0.07
Canal	25			0.00	366	88.1	0.03	378	87.8	0.03
Reservoir	64			0.00	36	178.3	0.07	37	178.2	0.07
Ephemeral vegetation	75	43	23.9	0.01	64	38.6	0.01			0.00
Hedge with trees	85			0.00	2	0.2	0.00			0.00
Line of trees	84			0.00	6	0.4	0.00			0.00
Linear habitats	101	4898	766.1	0.29			0.00			0.00
Allotments, city farm, community garden	33	426	228.5	0.09	574	240.7	0.09	520	257.9	0.10
Amenity grassland	54	44813	5309.3	2.04	17929	965.2	0.37	51013	5446.1	2.09
Amenity grassland and scattered scrub	72			0.00	32	7.8	0.00			0.00
Arable	5	19874	125706.6	48.25	29428	125009.0	47.97	44620	124275.6	47.69
Arable and scattered trees	86			0.00	10	31.9	0.01			0.00
Arable field margins	94			0.00	1	0.0	0.00			0.00
Agricultural land	1	243	240.9	0.09	9582	562.5	0.22	3864	2786.2	1.07
Cultivated/disturbed land	47			0.00	341	99.3	0.04			0.00
Open mosaic habitats	61			0.00	212	236.8	0.09	399	241.3	0.09
Other habitat	108	78	7.5	0.00			0.00			0.00
Other habitat,probable garden/brownfield or park	107	501	117.2	0.04			0.00			0.00
Garden	24	372066	10428.7	4.00	371353	10479.1	4.02	372001	10552.5	4.05
Natural surface	2			0.00	73179	2565.7	0.98	38157	2314.3	0.89
Sand	59			0.00	30	4.7	0.00	20	5.7	0.00
Cemeteries and churchyards	31			0.00	1568	119.8	0.05	1873	188.9	0.07
Golf course	48			0.00	1056	1342.4	0.52	635	1364.5	0.52
Natural sports facility, recreation ground or playground	32			0.00	2975	1352.0	0.52	2681	1372.4	0.53
Fountain	80			0.00	21	0.1	0.00	21	0.1	0.00
Natural rock, other exposure, boulders	102	2	0.0	0.00			0.00	2	0.0	0.00
Landfill	78	32	80.9	0.03	36	79.3	0.03	24	68.8	0.03
Quarry or spoil	68	286	725.1	0.28	516	664.2	0.25	122	443.3	0.17
Quarry or spoil (disused)	69			0.00	83	12.8	0.00	78	45.2	0.02
Bare ground	22	171	38.6	0.01	268	50.6	0.02			0.00
Track	20	11003	1107.3	0.42	12751	1128.4	0.43	13197	1127.5	0.43
Building	7	525232	4012.3	1.54	513330	3882.1	1.49	518928	3910.6	1.50

Swimming pool	58			0.00	87	1.1	0.00	89	1.1	0.00
Path - manmade	14	22949	314.6	0.12	22311	306.4	0.12	23415	310.8	0.12
Rail	12	1162	283.2	0.11	928	280.4	0.11	1016	281.5	0.11
Road	9	67639	3976.5	1.53	66340	3913.5	1.50	67590	3935.3	1.51
Road island / verge	13	40024	2595.9	1.00	41316	2548.3	0.98	41024	2564.6	0.98
Roadside - manmade	11	22717	733.9	0.28	20639	708.1	0.27	21146	714.0	0.27
Roadside - unknown surface	87			0.00	763	10.5	0.00	753	10.5	0.00
Sealed surface	8	74992	3731.5	1.43	75859	3765.6	1.45	77153	3767.1	1.45
Bridge	10			0.00	2307	24.0	0.01	2448	24.3	0.01
Bridge - natural	27			0.00	624	1.3	0.00	587	1.2	0.00
Undefined	23	3124	649.0	0.25	884	131.6	0.05	2607	761.6	0.29
Total		**1375329**	**260554.4**		**1425288**	**260592.7**		**1409428**	**260602.8**	

Table 4 shows a simplified habitat classification, grouping the different habitat classes into broader groups (woodland, arable, water etc), and grouping HLU green space (golf courses etc) together with the underlying amenity grassland and 'natural surface' (which is mainly amenity grassland). This table illustrates that the overall habitat classifications in the different datasets are broadly similar at this higher level. Some of the key differences are:

- The LCM dataset relies on CEH Land Cover Map 2015, derived from remote sensing data, which omits many small polygons of semi-natural grassland, and has some difficulty in distinguishing semi-natural from improved grassland. It is expected that the Phase 1 habitat data should be more accurate for small areas of semi-natural habitats. This explains why the LCM dataset has more amenity grassland and less semi-natural grassland than the NCS and HLU datasets, which use Phase 1 data.
- The LCM dataset does not distinguish scrub as a separate habitat, so the 300 ha scrub in this dataset is derived from OS MasterMap. The NCS and HLU datasets contain around 1300 ha of scrub, mainly from the Phase 1 dataset.
- The LCM has around 5000 ha less improved grassland and 2-3000 ha more 'Parkland and scattered trees' than the other datasets. This may be because the LCM dataset uses the Natural England Wood Pasture and Parkland priority habitat dataset, which tends to class entire parks or estates as this habitat, whereas the more detailed datasets show that some of these parks are actually improved grassland or arable land without trees.

Table 4. Area (ha) of simplified habitat groups

Simplified habitats	NCS	HLU	LCM	Comments
Allotments	228	241	258	
Amenity grassland, natural surface and other green space	5,309	6,354	10,687	HLU includes 2500 ha 'natural surface' (some of which may be improved grass) and 2814 ha green space e.g. golf courses (comprised of amenity grass and improved grass). Much of the LCM amenity grass is classed as semi-natural grassland in Phase 1.
Arable	125,947	125,703	127,062	
Building	4,012	3,882	3,911	
Fen, marsh and swamp	245	302	240	
Garden	10,429	10,479	10,552	
Heathland	0	6	2	HLU heathland is mainly from small polygons in Phase 1 and BAP habitats
Improved grassland	56,436	54,787	49,116	Some improved grassland in TVERC data could have been converted to arable.
OMHD		237	241	
Orchard	281	300	303	
Parkland with scattered trees	4,816	3,179	6,179	LCM uses Wood Pasture & Parkland priority habitat dataset, which covers whole parks /estates, not all of which is actually parkland with scattered trees.

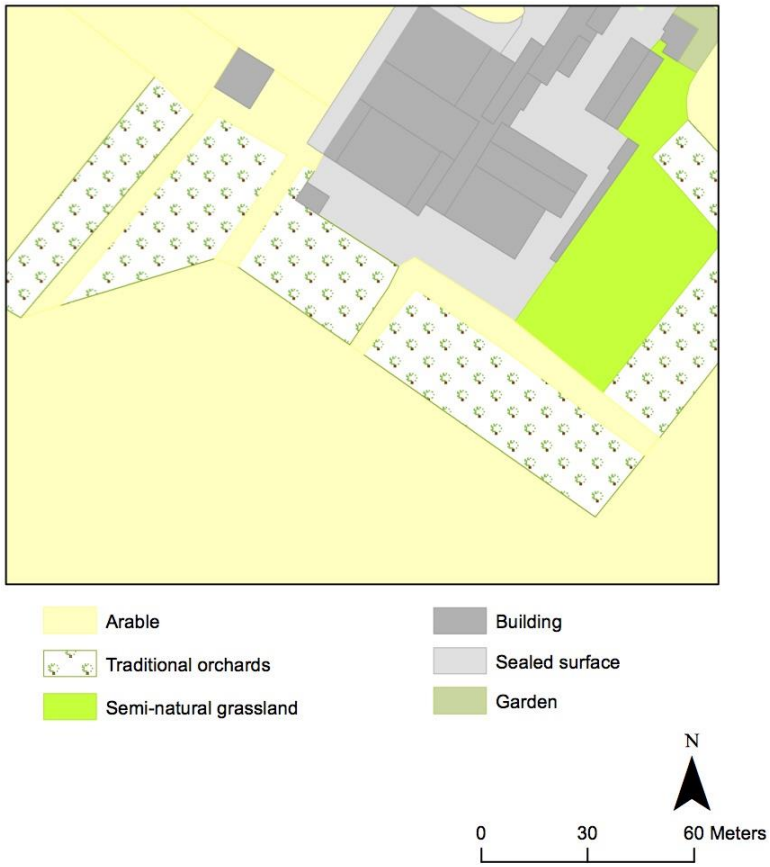
Quarry, landfill, bare ground, rock	806	761	563	
Scrub	1,403	1,300	371	There is 1315 ha of scrub in Phase 1, missed by LCM.
Sealed surface	9,040	8,998	9,033	
Semi-natural grassland	12,026	13,919	9,253	11,632 ha of semi-natural grassland in Phase 1; small patches not picked up by LCM
Track	1,107	1,128	1,127	
Undefined	1,540	132	762	NCS includes 700 ha of 'linear habitats' (e.g. grassy strips, embankments) which are often 'natural surface' in HLU.
Verge	2,596	2,559	2,575	
Water	3,103	3,090	3,096	
Woodland	21,188	23,167	25,271	
Total	**260,515**	**260,523**	**260,603**	

Comparing basemap extracts

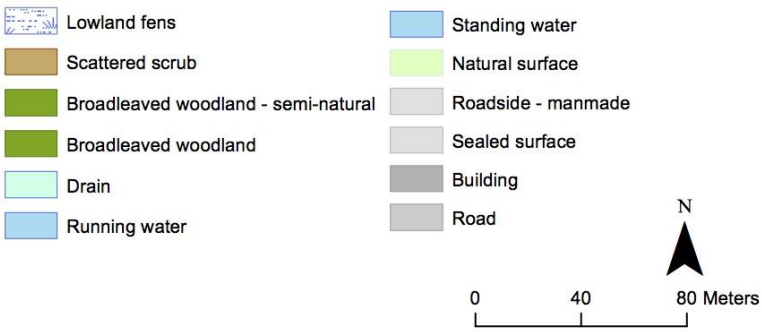
Extracts from the NCS and HLU basemaps have been chosen to illustrate visually the consequences of the differences and similarities in the basemaps. Comparison A shows habitats that have been classified in a similar way, but with some small differences. This is the most common occurrence. Both maps identify areas of orchard in a predominantly arable area, adjacent to a building surrounded by sealed surface. One small difference is that the grassland has been identified as semi-natural in the HLU basemap and an unknown type of grassland with some scrub or trees in the NCS map. This may be a consequence of extra data in the NCS mapping process showing the presence of other habitat types. These is also a small patch of improved grassland identified at the top of the NCS map, which is classified as arable in the HLU map. Checking against satellite images in Google Maps confirms that this is a wide strip of grassland between arable fields.

Comparison B is an example of where OS MasterMap polygons have been split in the HLU basemap along Phase 1, BAP habitat or designations boundaries. This results in two distinct areas in the middle of the extract of lowland fen and scattered scrub, whereas the NCS basemap classifies the whole of this area as parkland with scattered trees. The NCS basemapping procedure does not split polygons and as a result some detail can be missed where the OS MasterMap polygon boundaries do not match up with other data sources.

Comparison A. This extract shows broadly similar habitat classifications between the basemaps, but with some small differences. Top HLU, bottom NCS basemap. Note that different symbologies are used to display the habitats.



Comparison B. This demonstrates how habitat boundaries can differ due to the splitting of OS MasterMap polygons in the HLU basemap. Top HLU, bottom NCS basemap.



Conclusions

The overall agreement between the NCS and HLU basemaps is quite good, and the discrepancies are mostly due to the different classification systems used when creating the maps, where some habitats will be classified only broadly in one basemap, but in more detailed categories in another. However, there are habitats that are classified completely differently between the two maps, with the most common difference being found in mixed habitats and amenity grassland. Manmade habitats are recorded with very high consistency between the different maps.

There are bigger differences between the LCM map and the other two versions, with significant differences occurring in the habitats assigned as well as in the detail of the classifications used. The biggest difference in terms of area occurs for semi-natural broadleaved woodland, where the LCM estimates a much higher coverage.

That differences exist between the NCS and HLU maps may be surprising given that both use Phase 1 Habitat Survey information, but they also use slightly different combinations of additional data, and a different set of rules to assign habitats. As a result, there are bound to be some differences in the overall basemaps produced. Some of the key difference that arise between the maps are as follows:

Mixed habitats – these present the greatest difficulty when classifying habitats and result in many of the differences recorded. This occurs particularly for grassland habitats that also contain trees or scrub. Phase 1 habitat survey does not easily allow for mixed habitats, so new habitat categories have to be created and rules set when assigning these habitats, which differ between maps according to the approach taken. Rules are applied in both the NCS and HLU maps so that if trees or scrub are mentioned in the OS MasterMap Descriptive Term, one of the mixed categories mentioned will be applied. However, differences occur both in terms of the habitat categories created and when these are used. The NCS map contains 12 different categories with scattered trees/scrub or where these are a secondary feature, whereas the HLU map contains 17 such categories and many of these do not match those in the NCS classification. In the HLU rules, if the trees are scattered and the grassland is semi-natural, the habitat will be classed as 'Parkland with scattered trees' (either broadleaved, coniferous or mixed). If there are scattered trees but no mention of grass, the classification is just 'Scattered trees'. If the trees are not scattered, the habitat is assumed to be former semi-natural grassland now taken over by trees so it is marked as, for example, 'Broadleaved woodland on neutral grassland'. This leads to differences with the NCS dataset. For example, a polygon defined as unimproved neutral grassland in the TVERC Phase 1 dataset, but with scattered broadleaved trees mentioned in MasterMap will be assigned to "unimproved grassland with scattered trees" in the NCS map, and "Parkland with scattered trees - broadleaved" in the HLU version.

Amenity grassland – the NCS map classifies a much greater area as amenity grassland than the other two maps. All three maps apply a rule that natural surfaces in urban areas that aren't gardens and have no trees or scrub, will be amenity grassland. It is likely that this is generally true, but will not always be the case. However, as mentioned above, the HLU and LCM maps also reclassify the grassed parts of certain greenspace categories according to the functional use, e.g. golf course, etc.

Functional spaces – information on land use and function such as golf courses, natural sports facilities and so on can either be used to help determine habitats (as for the NCS map) or can be used as a final category in its own right (HLU Map), with an associated decrease in amenity grassland area as mentioned above.

Splitting or lumping – in a number of categories, differences occur depending on whether habitats are split or lumped together. The biggest examples for this are water where the NCS map lumps all of these together whereas the HLU map divides them into 6 categories, and buildings, where NCS splits them into 6 categories, whereas HLU lumps then all together.

Splitting polygons – the HLU map splits the underlying MasterMap polygons to create smaller polygons where habitats mapped in the TVERC habitat data follow different boundaries to the MasterMap polygons. This will be most relevant for picking up small patches of habitat (such as small marshes within larger fields or small patches of heathland within woodland) or picking up transitions between habitats across wider areas. In theory this will create more accurate maps. The impact of splitting It is likely to be more significant in upland areas where there are less boundaries and large polygons in MasterMap.

Importance of local habitat data

The comparison with the LCM map highlights the importance of local habitat data. The LCM map was produced with the same data as the HLU map, with the exception of the local habitat data provided by TVERC and so a comparison between these two maps provides a direct assessment of the difference that this data makes. There were significant differences between the LCM map and the HLU map both in terms of the detail of classification possible, but also in terms of habitats being classified completely differently between the two. Approximately a third of polygons were classified into a different or more detailed habitat type with the additional TVERC data and almost all of the differences occurred in the non-built habitat types.

Oxfordshire is unusual in having a detailed and up-to-date Phase 1 habitat and land use survey available across most of the county. Across the rest of the Ox-Cam Arc, Phase 1 data is only available in Cambridgeshire and this dates from the 1990s so has to be used with care. For Bedfordshire, Buckinghamshire and Northamptonshire there is no equivalent data available, although these counties all have some local data on the more important habitats for biodiversity. Across the country the picture is equally mixed, although up-to-date Phase 1 or equivalent surveys are rare. Although it is not possible to extrapolate in detail to other areas with different data availability, it is clear that the better the input data the better the quality of the maps that will be produced and that the effect can be significant.

Overall impact

Despite highlighting the differences and their causes in the above sections, it is important to note that the NCS map and the HLU map, which both used locally collected and comprehensive habitat data, are quite similar once different details of classifications are taken into account. Wholesale differences in habitat are relatively rare. This is reassuring given that the approaches rely on a whole series of rules that were set up independently as well as slightly different input data. There are steps that can be taken to ensure closer alignment of such maps in the future and these are set out in the recommendations section (below) and in the separate *Lessons Learnt* report that is an additional output of the Ox-Cm LNCP basemapping project.

Overall, man-made and urban features were mapped with great accuracy and consistency between the different maps. Water features were also captured very consistently, although varied in how they were classified. The more biodiverse habitats were captured relatively well by both the NCS and HLU maps, but much less well by the LCM map which did not have locally derived habitat data. Bigger differences occurred in the assignment of mixed habitats, especially where trees or scrub grew on other habitats, resulting in

some noticeable differences. The largest difference in terms of area was for amenity grassland, with the HLU map assigning some of these habitats to functional categories such as golf course or natural sports facility, or to "natural surface" for the 1% of Oxfordshire where no specific habitat data was present.

Differences in basemap classification will have an impact on the estimation of ecosystem service provision. However, these differences in the basemap are not likely to be large enough to make significant differences to overall provision estimations.

Recommendations

Locally derived habitat information – the quality and reliability of basemaps would be improved if local habitat information was available and up-to-date. The optimum would be to have a habitat survey that included all land use types (e.g. a Phase 1 type survey) and was updated regularly, but a compromise position would be for up-to-date surveys of areas of semi-natural habitats. Other data and licensing issues are discussed in the *Lessons Learnt* report.

Overall approach – the approach adopted by both the NCS and HLU involves using OS MasterMap as the underlying mapping unit and then using a series of rules to classify each polygon to the most appropriate habitat, based on a series of additional data layers (which can vary from place to place or depending on license restrictions). We consider this general approach to be much the best way of creating natural capital (habitat) basemaps and would recommend that this should be the approach used elsewhere. The approach is highly flexible as it can use any combination of data available and can incorporate new data sources as they become available (e.g. Sentinel derived maps). Overlaying data onto an OS MasterMap basemap remains better than using maps without the underlying basemap as OS MasterMap remains a rich source of mapping information, capturing details that would not be captured by satellite data alone and relating it to on-the-ground reality. OS MasterMap is going to become freely available in part in the near future, so this will enable the roll out of this approach to all areas across GB.

Habitat classifications – there is a need to align the habitat classification schemes so that natural capital basemaps can be compared more easily. The Phase 1 classification scheme has been the default until recently, but is now being superseded by the UK Habitat Classification Scheme. This has the major advantage over Phase 1 that secondary habitats, such as trees and scrub can be added easily as a secondary code and functional characteristics such as use as a playing field can also be added. This would really help to address the issue of mixed habitats highlighted here. It is recommended, therefore that classifications move towards using UK Hab.

Habitats used – this will be largely addressed by using the UK Habitat Classification Scheme (above), but it would be sensible to adopt the best bits of the two different approaches described in this report (the NCS and HLU versions). It would be beneficial to include open mosaic habitats which are not currently included in the NCS version and to separate traditional orchards from standard orchards where data allows, as both of these habitats are of biodiversity conservation value and are not adequately captured by the NCS map. Within the HLU map, habitats such as floodplain grazing marsh (floodplain wetland mosaic) should be captured as a distinct habitat type (these are currently captured in a separate column (BAP habitat) but the main interpreted habitat is defined as 'marshy grassland'). We considered whether it would be best if function / land use (e.g. golf course, playing field) was not conflated with habitats but remained as a separate code or column of data. However, the University of Oxford team need to incorporate these additional green space categories into the main habitat field, because this is used as the key to join the

base map to an ecosystem service scoring matrix. Ecosystem service scores for green space categories differ in several important respects, e.g. golf courses are assumed to be less publicly accessible than recreation grounds, and cemeteries have a higher 'sense of place' than golf courses, so this is fundamental to the ecosystem service scoring approach.

Standardised approach – using the UK Habitat Classification Scheme would address some of the issues highlighted here, but there will remain differences in the rules applied and therefore in the way that habitats are assigned. It would be highly beneficial therefore, if a standardised approach was developed and used throughout the country. To that end the project team are currently working with others to produce an updated basemapping approach that builds on both of the methods described here. This will be automated and free to use and it is hoped that this will become the default way to produce baseline natural capital (habitat) maps. That would ensure consistency of approach and outcomes and enable maps to be compared directly from different areas.