

t2.5

The National Land Use Database: building new national baseline data of urban and rural land use

Andrew Harrison, consultant to DETR/NLUD
Bob Garland, statistician, DETR

Introduction

The vision of the National Land Use Database (NLUD) is to establish a complete, consistent and detailed geographical record of land use in England, which is kept up to date and delivers information to users that meet their business requirements (NLUD, 2001).

To achieve this vision the NLUD project is working to establish an integrated framework for land use information at the national, regional and local level. The framework will consist of a series of core data specifications for land use, including toolkits and good practice guidance, underpinned by a data infrastructure comprising national data standards. The key data standards for NLUD are: the Ordnance Survey Digital National Framework (DNF), the National Land and Property Gazetteer (NLPG) and the NLUD classification (Harrison, 2000).

Work to establish this integrated framework and data infrastructure is developing in two main directions. First, methods are being investigated to automatically derive baseline land use information from OS DNF and existing spatial data sources which are capable of rapid extension to the national (England in the first instance) level. Second, research is commencing to extend the operational NLUD Previously-Developed Land (PDL) project into a broader framework for Local Authority (LA) monitoring of all development land.

NLUD Land Use Classification

Development of the NLUD Land Use Classification is central to these developments and is the key data standard for NLUD. The classification was primarily designed to serve identified user requirements across central government and related agencies. Land use classes and definitions were based on the Land Use Change Statistics (LUCS) classification (GSS, 2000) and were developed around the modes of data collection proposed for NLUD in the mid-1990's (Dunn & Harrison, 1994).

The current version of the NLUD Classification (version 3.2) is the outcome of a formal expert review and field-testing in conjunction with the OS Ashford Trial (Project'93) of structured data. It has been implemented within the Microsoft ACCESS™ NLUD PDL Data Entry Monitoring Tool (DEMT) to help LA users assign land use to identified sites. The DEMT provides translation between NLUD and the statutory Town and Country Planning Use Classes Order (UCO) and a glossary of over 400 detailed land use descriptions linked to NLUD classes (NLUD, 2000a).

Earlier this year, in preparation for the NLUD baseline research, this classification has been revised to provide a more consistent approach to classifying recreational land and buildings and to extend the definition of 'Field crops' to include the former 'Ploughed fields' class. The resulting classification (version 3.3) is presented in Table 1. Links to other classifications have been established through correspondence tables that translate native codes to the NLUD land use classes. Correspondence tables have been set up to translate business use codes from the Valuation Office (VO) National Non-Domestic Rating (NDR) lists and land cover codes from satellite-derived Land Cover Map 2000 (LCM2000) data.

Greater standardisation of the classification will be achieved by building on these developments. In particular, further refinement of the hierarchical reporting structure of the classification with a ‘thesaurus’ of terms and use of correspondence tables.

Using OS DNF data, linked to the NLUD Classification, as a ‘generic’ tool for defining basic spatial units (BSUs) for land use will further extend these developments. It will enable the construction of land use data sets through data association as demonstrated by the NLUD Baseline research. More importantly it provides the basis for modelling different BSUs in relation to different dimensions of land (e.g. activity, function, ownership, cover) and the possibility of developing a multidimensional approach to land classification (APA, 1999). This approach allows multiple characteristics to be classified simultaneously, as each classification dimension is modelled as a separate attribute, and can therefore support heterogeneous applications in a flexible and consistent manner.

NLUD Baseline

The NLUD project was established in the early 1990’s by the Department of the Environment (DOE, now Department for Transport, Local Government and the Regions) in research designed to address the paucity of information about the extent and distribution of land use in England. This research identified the need for an integrated system of information on land use to establish baseline data from which to monitor changes in land use in both the rural and urban environment (Dunn & Harrison, 1994). Subsequent work tested a variety of methods for capturing land use information and attribution of polygons in conjunction with experimental OS structured digital mapping.

The current NLUD Baseline research builds on this previous work with the aim of exploiting the new OS DNF data to deliver a complete data specification capable of providing national data. Land use attributes are initially derived from OS DNF feature codes and text association and supplemented by analysis of a number of existing third-party data sources (Harrison et al., 2001). To date a series of prototype land use ‘products’ have been created for a representative set of test sites. The research is led and funded by DTLR in collaboration with OS. Infoterra Limited is the main contractor for the work.

Data Sources

A variety of methods for associating land use data with DNF are being investigated. There are two underlying approaches:

1. data association based on ‘internal’ DNF feature types and classifications, and
2. data association based on ‘external’ data sources.

Three potential sources of land use information are contained in the ‘internal’ DNF feature types and classifications:

- the polygon feature code and associated description,
- multiple vegetation cover descriptions, and
- the DNF text layer.

The use of ‘external’ data sources was investigated with a view to assessing the trade-offs between cost and accuracy in providing additional land use information. Information from four maintained data sources with national coverage was used.

The key ‘external’ data sources for ‘urban’ areas are:

- residential land use inferred from OS Address-Point and Code-Point, and
- business codes from VO NNDR data (linked to Address-Point).

The key 'external' data sources for 'rural' areas are:

- land use categories extracted from LCM2000, and
- woodland types from Forestry Commission (FC) Woodland Inventory data.

Test Areas

Twelve test areas were selected (in general 2 x 2 km squares) to cover as wide a variety of land uses and geographical areas as possible. For example, Victoria was chosen as an inner city area, Dudley as part of a Metropolitan Borough, Chesterfield as a county town, part of Basingstoke to provide an urban/rural boundary site, sites in West Sussex and Kent to provide enclosed rural landscapes and sites in the Peak District and Dartmoor to provide upland unenclosed areas. In practice, choice of specific locations was constrained by the availability of OS DNF and LCM2000 data sets both of which were under construction at the time of data supply.

Methods

The key objective of the research was to develop and test an automated process for assigning land use attributes to BSUs defined using DNF polygon features. In practice, two separate semi-automatic processes have been defined for 'urban' and 'rural' areas.

The methods for 'urban' and 'rural' areas are presented in Tables 2 and 3. Land use attributes are initially derived from OS DNF feature codes and associated text. This information is refined for urban areas using OS Address-Point and Code-Point data in conjunction with VO NNDR data. In rural areas, land use attributes are extracted from LCM2000 in conjunction with FC Woodland Inventory data. Figures 1 and 2 show a sequence of images from the analysis of the Aylesbury and Essex test sites to illustrate key data sources and the performance of the methodology.

Validation of prototype data sets

Data validation is performed by analysing the correspondence between prototype data, derived semi-automatically, and reference data, derived by air-photo interpretation (API) and ground checking.

Table 4 presents results for the urban test sites. Initial results, based on the Aylesbury test site, suggested that it was possible to assign a land use code to 87% of the total area with an overall correspondence of 74%. This correspondence increased to 86% if the data were generalised to the NLUD Divisional level. However, overall results based on the other urban test sites are lower both in terms of coverage (50-87%) and correspondence (32-57%). These other urban sites have higher concentrations of industrial and commercial land, which are poorly classified by the methodology at present. In addition, the new OS Code-Point boundaries derived by tessellation rather than from OS DNF features, were only partially successful in correctly assigning the use of land adjacent to buildings. Simulated Code-Point boundaries used for the Aylesbury test site had been digitised to follow OS DNF features.

Results for rural areas are presented in Table 5. Use of LCM2000 data enables a higher completion to be achieved with 85-96% of the total area being allocated a land use code. However, assessment of correspondence and likely accuracy has produced more variable results. The low correspondence for the Dartmoor test site is primarily due to the spatial mismatch between OS DNF polygons and LCM2000 polygons in unenclosed upland areas. Here LCM2000 identifies areas of semi-natural grassland, bracken and heath, however, the boundaries of these areas are not defined within OS DNF data. This raises questions as to how these areas should be represented within OS DNF and what level of mapping detail different users of NLUD baseline data will require. One possibility is to leave these unenclosed areas unclassified and instead to provide the user with an orthorectified image layer that would enable direct interpretation of the required land cover boundaries.

Re-assignment of areas of bare soil misclassified as 'urban' to field crops and use of the FC Woodland Survey to supplement woodland identification by LCM2000 increases overall correspondence. Results for the Milford, Sussex and Essex test sites suggest that an overall correspondence of 70% is probably a better figure of merit for the current configuration of the 'rural' methodology at the NLUD Class level. Higher levels of correspondence (80%) are achieved by generalising the classification to the NLUD Divisional level.

Validation of the first prototype data will be used to identify ways in which the current methods can be improved with a view to improve current levels of completion and correspondence and to remove the need for manual intervention.

NLUD PDL and Development Land Monitoring

NLUD Previously-Developed Land (PDL)

The NLUD Previously-Developed Land (PDL) project commenced in 1998 in response to the policy need for information on 'brownfield' land (Harrison, 1999). The Planning for the Communities of the Future policy statement in February 1998 set the target that, by 2008, 60 per cent of additional housing should be provided on previously-developed land and through the conversion of existing buildings. The statement also launched NLUD to provide a consistent assessment of previously-developed land. Planning Policy Guidance Note 3 – Housing, published in March 2000, provided guidance on how local authorities should give priority to the release of previously-developed land (DETR, 2000a). Supplementary guides provided details of the urban capacity studies required to assess how much new housing can be accommodate within existing urban areas and the monitoring required to evaluate the delivery of the policy (DETR, 2000b; DETR, 2000c). The NLUD PDL work is a key part of this information.

The project has been developed as a partnership between DTLR, OS, English Partnerships and the Improvement and Development Agency (IDeA). Local authorities contributed much to the success of the first round of data collection in 1998, with 95 per cent of them making a return. This enabled the publication of estimates of the total area previously-developed vacant and derelict land and other land that may be available for redevelopment. Site level data was also disseminated for 170 local authorities as a trial to assess the potential market.

Consultation with the local authorities has been a key feature of NLUD PDL. The specification for the 1998 data collection was developed through discussions with a cross section of local authorities. Likewise, the specification for the 2001 update evolved through regional workshops (NLUD, 2000b). The update specification has some notable additional features:

- definition of 'previously-developed' linked directly to the 'policy' definition in Planning Policy Guidance Note 3,
- collection of site boundaries (polygons), and
- recording of site history through various stages from previously-developed to redeveloped.

A new version of the DEMA was distributed to local authorities. This contains data validation that should improve data quality and limit the need to check data items with local authorities. It was recognised that the collection of polygons is an extra demand on local authorities. The project has sought to alleviate this in the following ways:

- development of a version of the DEMA with a 'polygon picking' function for distribution to local planning authorities which do not have access to proprietary GIS software,
- an accreditation scheme for GIS suppliers to develop tools which assist in polygon capture and linking the polygon and attribute data, and
- an extension to the NLUD Help Desk to provide support on GIS issues.

The update is progressing slowly. Local authorities have to fit the work in with other priorities such as completion of the Development Plan and urban capacity studies. Some of the information is derived from ad hoc surveys which are scheduled at different times of the year. The project needs to set out more clearly how NLUD PDL is an integral part the urban capacity studies and sequential approach to the release of land for housing which were introduced by PPG3. NLUD PDL needs time to become established. The specification will, therefore, remain constant for the next few update rounds.

Development Land Monitoring

Consultation with local authorities has revealed that in order to have a complete picture for planning information is required on 'greenfield' sites. This provides, for example, housing information on the balance between 'brownfield' and 'greenfield' sites. There is also a need to monitor land for industrial and commercial development. The NLUD PDL model could be extended to include all sites with planning permission and to follow them through to completion. The NLUD Development Monitoring project is taking forward in consultation with the Greater London Authority (GLA) and possibly other local authorities. The GLA are introducing a new London Development Database, which will collect information from Boroughs on planning permission and completions. It is planned that research will investigate how the DEMT could be extended to support this work.

The overall aim of this work is to find out how the NLUD PDL approach can support the monitoring requirements of local planning. The process will only work if the information collected is useful for each local planning authority. It is envisaged that a 'core' set of items will be collected to a common specification. This will enable the data to be compiled to provide regional and national pictures. Local authorities may include additional items around the 'core' to meet particular local needs.

Conclusions and Future Work

NLUD Baseline

The NLUD Baseline results presented here demonstrate the potential for automatically deriving land use information from OS DNF data in conjunction with a limited set of nationally available data sources. While the results are variable at the NLUD Class level, at the more generalised NLUD Division (13 land use categories) the overall accuracy is ~80% for 'rural' areas and ~70% for 'urban' areas.

During the remainder of 2001 the research will be extended to produce a county-level data set (~1,000 km²) for demonstration and user evaluation purposes. This extension will allow further refinement of the methodology and investigate the effects of scaling-up the methodology from test sites to a large tract of land including urban, rural and upland areas. During 2002 DTLR and OS will carry out a joint assessment of the business and technical feasibility of this approach for incorporating land use in future releases of OS DNF.

NLUD Development Land Monitoring

The NLUD PDL project is a success but there is need to imbed the data collection more firmly in monitoring for planning required by PPG3. The links between NLUD, urban capacity studies and the sequential release of land need to be made more explicit.

There is potential to extend to the NLUD PDL methodology to cover all development land including 'greenfield' sites. Research will explore this in the context of the GLAs 'London Development Database' and possibly with other authorities.

Project Plan 2001-2003

A two-year project plan to achieve the framework for NLUD Baseline and Development Land Monitoring is presented in Figure 3.

Research and development for creation of baseline data at the national level is underway. Work on the county-level demonstrator will commence following release of OS DNF data in November 2001. The demonstrator will be used for market research and business planning during 2002. Assuming a successful business case can be established then production of data can commence with data becoming available in early 2003.

NLUD PDL data collection is planned to continue without major changes until 2003 while research and development starts on development of LA Development Land Monitoring. Collaborative research with LAs during 2002 will be developing integrated specifications for core data sets. These specifications will be piloted before final specifications are released to LAs and software suppliers during 2003. Note that the availability of Baseline data at the national level will feed into these specifications. As the new LA monitoring approach is rolled out during 2004 it will over time supersede NLUD PDL data collection.

Acknowledgements

The authors wish to acknowledge the contribution of Robin Fuller and Geoff Smith at CEH (NERC) for advising on use and interpretation of LCM2000 data. Thanks also to Phillipa Swanton, David Askew and Jonathan Rhind (DEFRA), Harry Bruhns (The Bartlett, UCL) and Graham Bull (FC) for help in acquiring data sets for the Baseline research.

References

APA (1999) *Land-Based Classification Standards*, Chicago: American Planning Association (available from <http://www.planning.org/lbcs>).

DETR (2000a) *Planning Policy Guidance Note 3: Housing*, London: The Stationary Office.

DETR (2000b) *Tapping the Potential – Assessing Urban Housing Capacity: Towards Better Practice*, London: The Stationary Office.

DETR (2000c) *Monitoring Provision of Housing through the Planning System: Towards Better Practice*, London: The Stationary Office.

Dunn R & Harrison A R (1994) *Feasibility study for deriving information about land use stock*, Final Report to the Department of the Environment, 106p + appendices.

GSS (2000) *Land Use Change in England No 14*, London: Government Statistical Service (available from <http://www.planning.detr.gov.uk/luc14>).

Harrison A R (1999) *The National Land Use Database: from concept to implementation*, Proceedings of Eleventh National Conference of the AGI, Olympia, London, 3.7.1 – 3.7.4.

Harrison A R (2000) *The National Land Use Database: developing a framework for spatial referencing and classification of land use features*, Proceedings of the AGI Conference at GIS 2000, Olympia, London, 5.6.1 – 5.6.8.

Harrison A R, D'Souza G & Smith G M (2001) *Integrated analysis of spatial data sets to create baseline urban and rural land use data*, Proceedings of First Annual Conference of Remote Sensing and Photogrammetry Society, DTI, London.

NLUD (2001) The National Land Use Database <http://www.nlud.org.uk/>.

NLUD (2000a) NLUD Previously Developed Land (PDL) Technical Specification (available from *Key Published Documents*: <http://www.nlud.org.uk/>).

NLUD (2000b) NLUD Previously Developed Land (PDL) Data Specification (available from *Key Published Documents*: <http://www.nlud.org.uk/>).

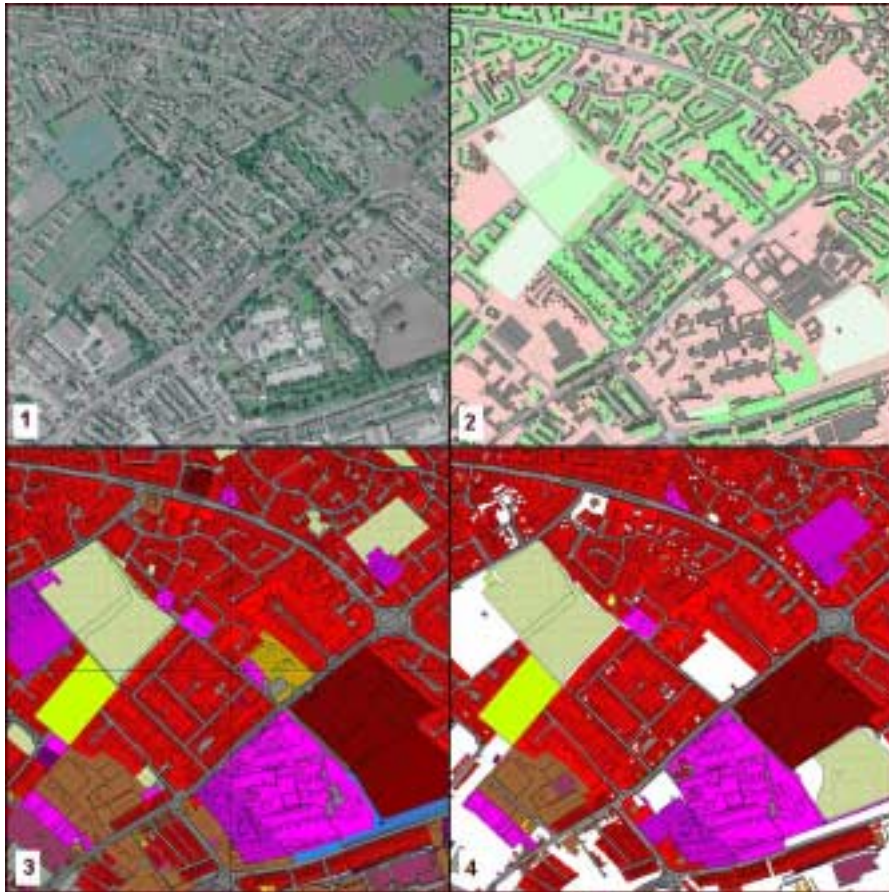


Figure 1 Processing sequence for the Aylesbury test site. 1: True-colour orthophotograph, 2: OS DNF data, 3: Reference data derived from API and ground data collection, 4: Prototype results – steps 1-10)

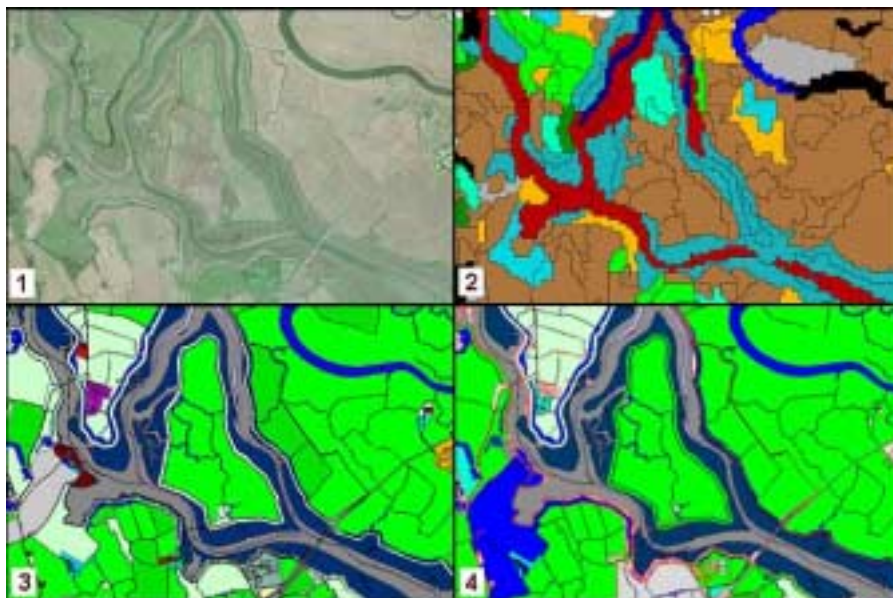


Figure 2 Processing sequence for the Essex test site. 1: True-colour orthophotograph, 2: LCM2000 data, 3: Reference data derived from API and ground data collection, 4: Prototype results – steps 1-10). Note that different legends and colours are used in Figures 2(2) and 2(3&4).

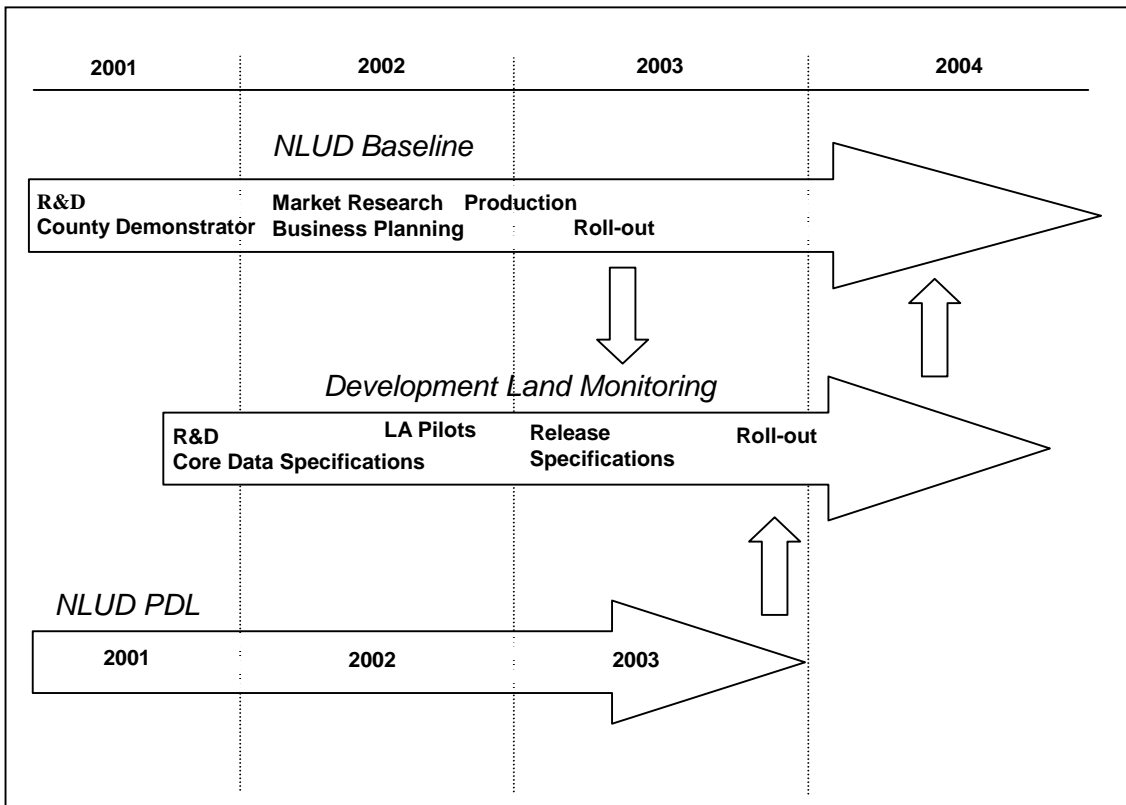


Figure 3 Project plan to achieve framework for NLUD Baseline and Development Land Monitoring

NLUD Land Use Classification v3.3	
1 Agricultural	6 Minerals and Landfill
1.1 Field crops	6.1 Mineral workings and quarries
1.2 Fallow land	6.2 Landfill waste disposal
1.3 Horticulture and orchards	7 Recreation
1.4 Improved pasture	7.1 Leisure and recreational buildings
1.5 Field margins	7.2 Outdoor recreation
2 Woodland	7.3 Allotments
2.1 Conifer woodland	8 Transport
2.2 Mixed woodland	8.1 Roads
2.3 Broadleaved woodland	8.2 Public car parks
2.4 Undifferentiated young woodland	8.3 Railways
2.5 Scrub	8.4 Airports
2.6 Felled woodland	8.5 Docks
2.7 Land cultivated for afforestation	9 Residential
3 Unimproved Grassland and Heathland	9.1 Residential
3.1 Unimproved grassland	9.2 Institutional and communal accommodation
3.2 Heathland	10 Community Buildings
3.3 Bracken	10.1 Institutional buildings
3.4 Upland mosaics	10.2 Educational buildings
4 Water and Wetland	10.3 Religious buildings
4.1 Sea/Estuary	11 Industrial and Commercial
4.2 Standing water	11.1 Industry
4.3 Running water	11.2 Offices
4.4 Freshwater marsh	11.3 Retailing
4.5 Salt marsh	11.4 Storage and warehousing
4.6 Bog	11.5 Utilities
5 Rock and Coastal Land	11.6 Agricultural buildings
5.1 Inland rock	12 Vacant Land and Buildings
5.2 Coastal rocks and cliffs	12.1 Vacant land previously developed
5.3 Inter-tidal sand and mud	12.2 Vacant buildings
5.4 Dunes	12.3 Derelict land and buildings
	13 Defence Land and Buildings

Table 1 NLUD Land Use Classification (version 3.3) comprising 13 Divisions and 51 Classes

#	Method	Rules
1	Recode OS DNF features to NLUD Classification v3.3	Derive classes: 8.1 Roads, 8.3 Railways, 4.0 Water and Wetland
2	Process OS DNF associated text field	Search for recognised text strings in “associated” field
3	Process OS DNF point text layer	Search for recognised text strings in point text layer
4	Apply adjacency rule to Outdoor Recreation polygons	Select all unclassified polygons adjacent to Outdoor Recreation where feature_de = “Defined Natural Land Cover”
5	Process OS Address Point for Residential land	(a) assign address-points within ‘Residential’ OS Code-Point polygons to 9.1 Residential (b) assign DNF polygons with ‘Residential’ address points to 9.1 Residential
6	Process VO NNDR data	Match VO data to OS Address-Point and then to DNF
7	Process VO NNDR data (manual)	Manually match VO NNDR to OS Address-Point
8	Process OS Address Point (manual)	Attempt to assign Business Name of address-point to unclassified OS DNF polygons
9	Process OS Code-Point polygons	Apply rules to determine curtilage within unit postcode polygons
10	Apply adjacency rule to Residential polygons	Assign residential curtilage for those buildings not assigned by postcode method

Table 2 Methodology for ‘urban’ test sites

#	Method	Rules
1	Recode OS DNF features to NLUD Classification v3.3	Derive classes: 8.1 Roads, 8.3 Railways, 4.0 Water and Wetland
2	Process OS DNF associated text field	Search for recognised text strings (e.g. ‘Farm’, ‘Pond’) in “associated” field
3	Process OS Address-Point for Residential land	(a) assign address-points within ‘Residential’ OS Code-Point polygons to 9.1 Residential (b) assign DNF polygons with ‘Residential’ address-points to 9.1 Residential
4	Process OS Address-Point for Farm	Assign OS DNF Building polygons containing ‘Residential Farm’ address-point to 9.1 Residential
6	Process VO NNDR data	Match VO data to OS Address-Point and then to DNF
7	Process LCM2000 polygons	Compare selected set of DNF polygons with LCM2000 polygons and translate land cover codes to NLUD v3.3
8	Process FC Woodland Inventory data	Compare FC Woodland Inventory and LCM2000 classified areas - IF areas are incorrectly classified by LCM2000 as woodland THEN assign to ‘Unclassified’
9	Apply adjacency rule to Farm polygons	Assign all unclassified buildings without an address-point and within 50m of a farm to 11.6 Agricultural Buildings
10	Apply adjacency rule to Residential polygons	Assign all unclassified polygons adjacent to residential buildings where feature type = “Multiple Surface Land” and area > 2000 m ² to 9.1 Residential

Table 3 Methodology for ‘rural’ test sites

	Completion %		Correspondence %		
	By area assigned	By polygons assigned	For polygons assigned	Overall NLUD Class	Overall NLUD Division
Aylesbury	87.5	92.2	84.9	74.3	85.9
Chesterfield	50.3	57.2	65.0	32.9	70.0
Dudley	53.6	83.0	65.5	36.6	67.3
Ashford	55.37	73.19	66.8	37.4	71.7
Victoria	70.4	67.3	75.2	53.0	76.5
Basingstoke	87.2	90.8	66.0	57.5	74.4

Table 4 Summary of Research Pilot results for 'urban' sites

	Completion %		Correspondence %		
	By area assigned	By polygons assigned	For polygons assigned	Overall NLUD Class	Overall NLUD Division
Milford	85.3	89.7	78.4	70.0	86.3
Essex	96.0	40.7	71.9	70.8	96.0
Sussex	91.8	43.9	49.3	44.4	76.8
Dartmoor	85.6	25.4	29.0	24.6	80.4
Peak District	96.1	72.4	<i>LCM2000 data not available</i>		
Tonbridge	88.4	48.2	53.4	46.2	82.3

Table 5 Summary of Research Pilot results for 'rural' sites