Office for National Statistics

Coding and Naming Policy for UK Statistical Geographies

June 2009

Edition No.: **v 1.0**

ONS Geography

Office for National Statistics

A National Statistics publication

National Statistics are produced to high professional standards set out in the Code of Practice for Official Statistics. They are produced free from political influence.

About us

The Office for National Statistics

The Office for National Statistics (ONS) is the executive office of the UK Statistics Authority, a non-ministerial department which reports directly to Parliament. ONS is the UK government's single largest statistical producer. It compiles information about the UK's society and economy, and provides the evidence-base for policy and decision-making, the allocation of resources, and public accountability. The Director-General of ONS reports directly to the National Statistician who is the Authority's Chief Executive and the Head of the Government Statistical Service.

The Government Statistical Service

The Government Statistical Service (GSS) is a network of professional statisticians and their staff operating both within the Office for National Statistics and across more than 30 other government departments and agencies.

Contacts

This publication

For information about the content of this publication, contact ONS Geography
Tel: +44 (0) 1329 444971
Email: ons.geography@ons.gsi.gov.uk

Other customer enquiries

ONS Customer Contact Centre
Tel: 0845 601 3034
International: +44 (0)845 601 3034
Minicom: 01633 815044
Email: info@statistics.gsi.gov.uk
Fax: 01633 652747
Post: Room 1.101, Government Buildings,
Cardiff Road, Newport, South Wales NP10 8XG
http://www.ons.gov.uk

Media enquiries

Tel: 0845 604 1858
Email: press.office@ons.gsi.gov.uk

Copyright and reproduction

Contents

1. Executive Summary

Introduction

1.1 Standards for coding and naming UK statistical geographies are inconsistent across the National Statistics (NS) community. Multiple schemes exist with a wide variety of code formats, structures and lengths. There is also a variable mix of intelligence built into codes. Some embedded intelligence works and allows relationships between geographies to be identified, but only where the reader has a prior understanding of the code structure. But there are also many cases where the built-in intelligence has corrupted when geographies change over time. Multiple names exist for the same place and different naming conventions are used by organisations. Poor coding practice at operational level also presents a risk to statistical quality.

1.2 These unnecessarily confused circumstances inhibit the exchange and integration of geographical data, particularly at the UK national level. There are clear, high level drivers that indicate this situation should be rectified. The UK Statistics Authority's (UKSA) Code of Practice[1] refers to "using common geographic referencing and coding standards". The UK Location Strategy[2] identifies within its strategic actions that 'we use common reference data so we know we are talking about the same places', and that 'we can share location-related information easily through a common infrastructure of standards...' The statistical geography coding and naming policy fits squarely with these desired outcomes.

[1] UK Statistics Authority Code of Practice

[2] The UK Location Strategy

New coding structure and practice

1.3 Given the above, ONS will implement a new coding and naming policy for some statistical geographies on 1 January 2011 **(Annex A)**. This policy has already been separately and independently implemented since 2004 by Scotland for all new or amended statistical geographies. A simple nine character alpha-numeric coding structure is at the centre of the new system. It is made up of two parts within the format ANNNNNNNN. The first three characters (ANN) represent the area **entity** (ie an area 'type', such as 'county'), with the first alpha value representing the country within which the entity lies. The second part of the code comprises six number characters (NNNNNN) and this represents the specific area **instance** (eg Hampshire).

1.4 The new code structure is already in use for over forty-one thousand instances of lower and middle layer Super Output Areas in England and Wales. In Scotland, it has also been implemented for a number of their entities, covering over eight thousand instances, including Data Zones and Intermediate Geographies.

1.5 Other salient features of the new NS statistical geography coding policy are:

• The coding system is not hierarchical and does not contain embedded intelligence (other than the first country character).

• The owner of the entity (or those to whom this responsibility is formally devolved) is responsible for the application of the new coding standards.

• Codes cannot be re-used.

• Codes will not change solely because of a name change.

• ONS will provide a central metadata service and administrative support for the new coding methodology by introducing a Register of Geographic Codes (RGC) and a clear set of maintenance rules **(Annex A)**.

• For ONS' geographies, the Change History Database (CHD) will ensure that relationships can be maintained between the old and new codes, and old and new statistical geographies (where this applies), and other aspects of geographical intelligence (eg hierarchies).

• For Scottish geographies, the Scottish Government (SG) will separately implement a RGC and a CHD, which will feed into ONS' RGC. The SG databases will be managed by the Office of the Chief Statistician (OCS) and they will liaise with ONS when cross border instances occur or when UK-wide geographies are created.

• ONS will provide guidance for the presentation of codes in outputs, as is currently the case.

Implementation

1.6 ONS will 'go-live' with the new codes (**Annex A**) in its public outputs on 1 January 2011 - the 'declaration date'. Code owners are welcomed to adopt the new codes prior to declaration date if able to do so. Advance awareness of the new format should also guide those presently implementing any geographic reorganisations or new information systems. There is no requirement to wait for formal adoption of the policy in order to utilise the new nine character structure. Those entities not recoded will keep their old codes. New entities may be added to the new coding standard at any time, as notified to ONS or the SG by other owners.

1.7 After the declaration date, all transfers of area coded data between NS organisations, and codes used in publications, must be in the new coding standard only (where they exist for the entities and instances in question). For those organisations remaining solely on the old code basis, the RGC and the CHD will provide the necessary conversion/lookup functions and metadata.

Naming policy

1.8 With regard to statistical area naming policy a number of core principles of good practice have been identified that should be applied wherever possible. Flexibility in the implementation of the principles is clearly required because of the historical inertia and emotion attached to place names. There are also legislative requirements around the Welsh & Gaelic languages to consider. Thus, it is simply not practical to implement absolute standards and wholesale change. In fact, the adoption and operation of the new coding policy, in some way, 'allows' for the continued existence of inconsistent naming conventions.

Antecedence

1.9 There has been a relatively long lead-in to the formulation of this policy. In early 2005, ONS commissioned a review by an independent consultant. The Department of Health, the (then) Office for the Deputy PM, NISRA, GROS, the Scottish Executive and the Welsh Office were consulted. The first version of the proposal was presented to the National Statistics Geography Group later in 2005 and the ONS Geography Co-ordination Board in 2007. Taking onboard the positive steer from these groups and other feedback, further work was undertaken. In April 2008 the GSS Regional and Geography group (GSSRG) received completely new and full documentation that gave details of the issues that existed with the coding and naming of statistical geographies within the GSS, as well as the suggested policy solution. A request was made to circulate the material as widely as possible, with comments called for and received by the end of December 2008. Bilaterals also took place during the same period. Further feedback since this time has been considered and incorporated where pertinent. The draft policy was then circulated to over 100 interested parties and put on the ONS website in the spring 2009. In June 2009, the GSS Regional and Geography Committee ratified the policy.

2. Coding policy for UK statistical Geographies

Key Definitions

2.1 **Geographic entity**: a specific 'group' or 'category' of statistical geographies in a UK constituent country; for example, electoral wards, or unitary authorities in England.

2.2 **Area instance**: a specific area within a geographic entity; for example, a named electoral ward, or a named unitary authority.

2.3 In most cases, the areal units will form complete coverage of the relevant constituent nation. Some, though, are stand-alone units that do not have total coverage, such as National Parks. Also, there may be cases where new entities are derived by overlap, such as the part of a local authority that falls within a national park.

2.4 **Data owner**: an organisation responsible for a geographic entity with authority to create and terminate area instances in consultation with ONS, who will manage the process through the Register of Geographic Codes. The data owner is responsible for keeping historical relationship and look-up information for their geographies, and for maintaining items of detailed information (such as instance names and name changes).

2.5 **Register of Geographic Codes (RGC)**: this is a 'light' but key document **(Annex A)** that summarises the range of area instances within each geographic entity for the UK, including those archived, with sufficient information to manage the allocation of new codes in cooperation with the entity owners. The RGC will be updated as close to real time as possible and available from the ONS website**.** For Scottish geographies, this process will be managed by the Office of the Chief Statistician (OCS), who will liaise with ONS as appropriate.

2.6 **Change History Database (CHD):** a database of every current and archived ONS owned or managed entity and instance, available on hard media or via a web link, which replaces and adds to the functionality of current ONS support products, such as the Standard Names and Codes (SNAC) and the Ward History (WH) Database. The CHD enables users to identify such things as higher and lower hierarchical relationships that are exact one-to-one or one-to-many, predecessor instances and identifiers, temporal delimiters and historical names.

Coding policy - exceptions

2.7 There are either (a) some statistical digital boundary sets for which it is not sensible to move to the new coding system; or (b) there are a few circumstances where only a proportion of an organisation's geographies will be recoded into the new system (by ONS) to run alongside the owning organisations' code ranges.

2.8 First, there is the case of the UK NUTS/LAU geographies, which form part of an established pan-European coding system administered by Eurostat. The NUTS/LAU codes are already ubiquitously used and accord to good coding practice. It would, therefore, be completely illogical and unproductive to have a mirror set of codes.

2.9 Postcodes are undoubtedly used as a statistical geography, but there was no pressure from stakeholders on the inclusion or otherwise of postcode sectors. At the moment, postcode sectors will not be coded.

2.10 For health geographies in England and Wales, the Organisational Data Service (ODS), as part of NHS Connecting for Health, is responsible for the publication of all NHS organisation and practitioner codes. Published ODS code standards are *de facto* part of the official NHS data standards. The new NS coding policy presented in this paper, however, is concerned solely with spatial (digital) boundary codes and not NHS *organisational* codes. Of course there is often a legal relationship between the two, but this is not necessarily always the case. So, an area belonging to a health organisation may change, but the organisational code may remain the same. Thus, to ensure best practice, ONS will create and maintain the new coding system for a number of statistical health geographies ie for the purpose of presenting and managing NS health statistics. These statistical area codes will not necessarily carry any currency within the NHS data standards, systems, services and central data repositories that have latterly been developed through the National Programme for Information Technology (NPfIT).

Coding policy - code structure

2.11 The code consists of a 9 character alpha-numeric code (ANNNNNNNN) consisting of two parts; the entity and the instance.

2.12 The first part (ANN) identifies the geographical entity (area type). The first alpha character of the entity code indicates the country (or Isle of Man, or Channel Islands or 'cross-border' status) within which the entity is bounded. The 2nd and 3rd numeric characters will represent a specific geographic entity, but with no other intelligence in the code. Where similar geographic entities exist across the UK, they will be considered as different geographic entities within each constituent country. Consideration has been given to a UK-wide entity code but rejected on the basis of political factors, user requirements and an extremely limited number of truly UK-wide, consistent geographies. 'I' and 'O' will not be used to avoid possible confusion with numeric characters. **Table 1** below illustrates the 'country' allocations. It demonstrates the degree to which this situation is ‘future-proofed’ in terms of capacity for the creation of new entities.

Table 1

Country/Part	Approximate number of entities	'A' in ANN (and those available)	Number of entities possible
England	70	E (ABCDFGH)	8*99 = 792
Wales	35	W (XYZ)	4*99 = 396
Scotland	35	S (TUV)	4*99 = 396
Northern Ireland	25	N (PQR)	4*99 = 396
Cross-border	few	K	1*99 = 99
Channel Islands	few	L	
Isle of Man	few	M	1*99 = 99
Unassigned	n/a	J	1*99 = 99

2.13 The cross-border designation (K) indicates that one or more of the instances within the entity is not completely contained within one of the constituent countries. There are presently few of these geographies, but numbers might increase in the future (eg with environmental statistical geographies, where national boundaries may not be respected).

2.14 Each area instance will have a 6 digit numeric code. This instance will be uniquely identified within the UK by its combination with the entity code. Instances must not be coded with, and/or be based on, inbuilt intelligence (eg alphabetically or hierarchically), as any later change (like renaming) that inevitably occurs might upset this inbuilt intelligence. The nature of hierarchical relationships and any other classification-type of intelligence must be recorded and managed elsewhere. For ONS managed geographies, this will be within their Change History Database (CHD) product. For Scottish statistical geographies, this will also be within a similar product/tool.

Coding policy - code maintenance rules

2.15 Codes cannot be re-used. The re-use of codes is bad practice and presents risks to the analyses of time-series geographic data.

2.16 Except for a few special circumstances detailed below, a new instance of the area will be created with a new code only when a change occurs to the boundary of an area instance. For example, name changes alone do not initiate a code change for the area instance. However, the name change will be recorded elsewhere in the CHD (or other organisations' equivalent).

2.17 There is not a 'minimal change rule' whereby small changes preclude the creation of a new instance. This is because the interpretation of such a rule would be inconsistent. To make it consistent, it would have to be based on a threshold value of change, such as given percentage of area or some such quantifiable variable. It follows that this would require onerous analyses, which may not easily be possible in some business environments.

2.18 Boundary revisions by Ordnance Survey surveyors that form part of the ongoing quality assurance process in constructing Boundary Line™ will not be interpreted as a change to a boundary that requires the creation of a new area instance.

2.19 When an area instance in a particular entity becomes an instance in another type of entity, then the instance will be allocated a new instance code in the new entity. For example, this could occur when a local authority district changes status to become a unitary authority.

2.20 Where a significant revision takes place to the majority of instances for a particular entity, then it might be logical to create a new entity with new instance codes, as required. The decision for this action lies with the owner of the entity, although ONS will of course liaise and provide guidance on such matters in its role as maintaining the RGC and the associated supporting CHD.

2.21 Finally, there will be some special cases where new codes will have to be applied to new instances within an entity that are spatially, at least, an exact match to their predecessors. Such situations arise where legally binding processes formally terminate an instance at a given date, but then recreate the same instance (at least in terms of its spatial extent) on a following date. This is the case with electoral wards following Boundary and Electoral Commission reviews and recommendations, where legally mandated changes are detailed within Statutory Instruments/Orders.

Coding policy - presentation

2.22 This section is primarily concerned with standards in publications (paper and electronic). There are three elements. First there is the matter of area presentational *order.* Second, there is the question of whether or not to use codes in publications. Third, if used, there is the question of their format.

2.23 Full guidance on the order of presentation is available in **Annex B**. The guidance covers:

- administrative areas at regional level
- electoral areas at European electoral region and Westminster parliamentary constituency levels; and
- health areas, health authority/health board and primary care levels

2.24 The principle is to present areas alphabetically by name in regional groups, usually by higher level geographical entities, for example local authority districts by counties. This principle applies to all levels and all geographical breakdowns of areas. As far as possible, areas should not be presented in code order or by some other method, unless there are extremely valid reasons for doing so. An example, for instance, might be ranking areas according to their value within a statistical data field, or presenting in geographic order to ensure consistency in order of presentation of areas that are shown in two different languages.

2.25 The inclusion of codes in publications should be avoided, if possible. The use of names alone is preferable. It is possible, however, that codes are required in publications, for example by reason of 'house' rules.

2.26 If circumstances dictate that codes must be included within a publication then they should be presented in full format, without any corruptions or shorthand adjustments, such as separating the entity and instance parts with a symbol. Undertaking any of the latter, or for example reducing the number of leading zeros in the instance codes, would lead to confusion and inconsistency in presentation.

2.27 In operational spreadsheet or database scenarios, it is imperative that the full code is always used. Matching, joining and other analyses are far more easily and rigorously undertaken when using codes. This requires that the full code is present. Some users might again be tempted to introduce a special character to split the code into its component parts. However, this might cause difficulty because special characters may contain computational meaning. Special characters would also inhibit the efficient exchange of data.

Coding policy - governance, support products and services

2.28 ONS will maintain a UK Register of Geographic Codes (RGC) (**Annex A**), working closely with owners where we have been given the remit to manage other owners' codes. The RGC's function is to allow the coordinated issue of new codes and the management and maintenance of code changes, and to maintain the relationship between active and archived code ranges (**Annex C**). In Scotland, a separate register of geographic codes will be held by the OCS, which will allow them to operationally allocate and manage codes for Scottish geographies. Appropriate information from these activities will be passed onto ONS to be held on the RGC. For cross-border or UK wide geographies a formal mechanism will be required to ensure co-ordination.

2.29 The RGC contains the following information for those entities coded under the new system:

Country/Part	Approximate number of entities	'A' in ANN (and those available)	Number of entities possible
England	70	E (ABCDFGH)	8*99 = 792
Wales	35	W (XYZ)	4*99 = 396
Scotland	35	S (TUV)	4*99 = 396
Northern Ireland	25	N (PQR)	4*99 = 396

Cross-border	few	K	1*99 = 99
Channel Islands	few	L	1*99 = 99
Isle of Man	few	M	1*99 = 99
Unassigned	n/a	J	1*99 = 99

- Entity code
- Entity name
- Entity acronym
- Entity theme
- Entity country of coverage
- Cross-border instance indicator
- Related entity codes (in other countries)
- Entity status
- The number of live instances
- The number of archived instances
- Date of last instance change
- Instance code range within the entity
- Entity owner
- The date the entity was introduced on the RGC
- The date the entity started (where known)

2.30 For entities and instances owned by ONS, the Change History Database will meet the additional requirements that are not met by the RGC. Such things might be the standard names, the previous names, previous code identifiers, creation and cessation dates, centroid co-ordinates, and lower and higher hierarchical relationships where these are in exact one-to-one or exact one-to-many/many-to-one situations.

2.31 The ONS CHD will record for ONS owned or managed geographies:

- the unique identifier code range (the new codes as a key field)
- the policy department responsible

• the policy department contact

• the higher (built to) hierarchy instance code(s)

• the lower (built from) hierarchy instance code(s)

• the predecessor instance (as close as possible, such as the corresponding area instance(s) that existed prior to the boundary change

• the revision cycle

• the standard name (see naming proposals)

• the historical name(s)

• alternative identifiers (codes used in preceding schemes)

• the temporal extents ie creation and cessation dates

• position eg centroid co-ordinate

2.32 Redundant codes cannot be re-used. They will be archived in the CHD where their effective dates of operation and cessation, along with their spatial relationships with their replacement geography will be recorded. In time, live area instance codes within each geographic entity range will not be consecutive since redundant codes will be preserved for historic analyses. This will avoid the situation sometimes faced in the past where the same code exists for different spatial units.

2.33 The process by which the entity and instance codes are allocated is outlined in **Annex C**. ONS, or the OCS if it is a Scottish statistical geography, will use the RGC to allocate the next available code or that specifically requested by the owner.

2.34 In all but a few exceptions, only the owner will originate coding for entities and instances. Other organisations cannot operate different coding schemes for entities and instances they do not own.

3. Naming policy for UK statistical geographies

3.1 It is difficult to devise and implement any hard and fast policy that affects place names. Names often have a lot of history and emotion attached to them and any force for change has to combat the inertia that builds from these factors. There are also some legislative requirements at the UK level around Welsh & Gaelic languages that need to be considered. Therefore, in contrast to the mechanistic processes of coding, naming requires more flexibility.

3.2 In this light, it is probably only practicable at this point in time to identify core principles, or conventions, and aspire to their full implementation. There will undoubtedly have to be exceptions in existence to the ideal model, with users working as best they can with the guidance.

3.3 Those responsible for naming new instances in statutory instruments (SI) have a special responsibility for ensuring conformity to the naming principles, because any exceptions or anomalies then have to be reproduced until they are repealed by another SI.

3.4 The principles on naming should be applied to all new statistical geographies; they are not expected to be applied retrospectively.

3.5 There are a number of naming conventions that should be used:

• Names for area instances within a given entity should be, as far as possible, unique across the UK.

• The use of the same names at different levels in a hierarchy should be avoided (eg previously, Durham was both a county and a district).

• The type of area should only be part of a name in order to provide uniqueness (where for instance the same name cannot be avoided). For consistency, this type-name should come at the end and be applied to the higher level geography (eg Durham County)

• Names can change without any other change to the attributes of the instance; a name change alone does not result in a code change.

• The use of abbreviations should be minimised and standardised. 'St', without the inclusion of a full stop, is only allowable as an abbreviation for 'Saint' where it is part of the formal name; 'S' must not be used for 'South'.

• Where a name is shortened, there should be a single NS version established by the entity owner. Such data must be recorded in a CHD database or local gazetteer by the owner.

• A standard set of characters should be used in names and the use of non-alphanumeric characters such as ampersands, slashes and hyphens should be minimised.

• Possessive apostrophes should be used.

• As is currently the case, ONS will provide rules for the presentation of names in outputs (**Annex B**).

• The CHD will record appropriate time stamped metadata about changes in the names of entities and instances for which ONS is the owner, and other owners should do similar.

3.6 Areas, their boundaries and their names are determined by the owner of the entity. In general, the owners work independently – there are, for instance, separate Boundary Commissions for England, Wales, Northern Ireland and Scotland. Thus, any initiative that seeks to improve co-operation and consistency in the naming of geographical areas should be welcomed and worked towards. The assumption is that agreement and proper practice for the naming of new instances can be achieved in the medium to longer term.

Annex A

The Register of Geographic Codes

https://geoportal.statistics.gov.uk/Docs/Products/Register_of_geographic_codes_(May_2014)_UK.zip

Annex B

ONSG presentational guidance

http://www.ons.gov.uk/ons/guide-method/geography/geographic-policy/best-practice-guidance/presentation-order-guidance/index.html

(This has been reviewed to take account of the impact of local government re-organisation of April 2009)

Annex C

The allocation of entity and instance codes in England and Wales

For Scottish geographies, this process will be managed by the Office of the Chief Statistician (OCS), who will liaise with ONS as appropriate, in order that the ONS RGC maintains a UK scope.