

---

# Guidelines for BS 7666:2006

## Section 4. How to create a 'national' gazetteer

Version 1

January 2007

*This Section of the Guidelines deals with how to create a 'national' gazetteer. It describes how to amalgamate 'local' gazetteers into a 'national' one. It identifies quality issues between the two types of gazetteer and in particular issues relating to how a national gazetteer should be maintained.*

### Contents

	<u>Page</u>
Preface .....	2
1. Introduction .....	2
2. How to amalgamate 'local' gazetteers into a 'national' one .....	4
3. Maintenance .....	8

## **Preface**

These Guidelines are intended for use with BS 7666: 2006 *Spatial datasets for geographical referencing*. They complement the Standard with more detailed explanation of the content and a general approach to creation of gazetteers. They are not specific to any particular implementation, for which it is expected that specific data specifications and capture and maintenance rules will be produced.

The Guidelines are aimed at:

- gazetteer owners – those with ultimate responsibility for the gazetteer;
- gazetteer custodians – those responsible for the creation, maintenance and quality of gazetteers;
- suppliers of gazetteer software;
- those developing and implementing gazetteer systems
- providers of gazetteer data;
- others who are responsible for aspects of the quality management of gazetteers.

The Guidelines are currently in four Sections:

1. Introduction to BS 7666;
2. How to create a gazetteer of a new type of geographic object;
3. Quality assessment and reporting;
4. How to create a national gazetteer.

Further Sections will cover specific implementation issues:

- How to create a street gazetteer;
- How to create a land and property gazetteer;
- How to create a delivery point gazetteer.

No guidelines for public rights of way which form an informative annex to Part 1 of the Standard are planned at present.

These Guidelines have been produced under the auspices of BSI IST/36 geographic information who are responsible for BS 7666. They were written by Rob Walker and Les Rackham working under the guidance of a Steering Group comprising representatives of major stakeholders in the Standard. The work is sponsored by the Department for Communities and Local Government (DCLG), Ordnance Survey, Office for National Statistics and Mayrise Ltd.

This publication may be reproduced free of charge in any format or medium provided that it is reproduced accurately and not used in any misleading context or in a derogatory manner. The material must be acknowledged and the publication cited when being reproduced as part of another publication or service.

## **1. Introduction**

### **1.1 Local and national gazetteers**

Gazetteers, especially street gazetteers and land and property gazetteers, are often described as being 'local' or 'national'. These terms were chosen to correspond to levels of government involved in the creation and usage of such gazetteers. Land and property gazetteers for example, are being created in local authorities and merged to form the National Land and Property Gazetteer (NLPG). However, here we will use the terms 'local' to mean of restricted coverage, and 'national' to mean of more widespread coverage encompassing several 'local' gazetteers.

### **1.2 Creation of a national gazetteer by amalgamation of local gazetteers**

There is a further assumption that 'national' gazetteers are created by amalgamation of local gazetteers. This need not necessarily be the case. For example, a national gazetteer of streets could be created directly from an existing source such as Ordnance Survey data. In this section we will concentrate on how to amalgamate local gazetteers into 'national' ones, although many of the issues identified are also relevant to creation of national gazetteers directly.

### **1.3 Purpose of a national gazetteer**

The purpose of a national gazetteer and a local gazetteer could well be different. For example, objects of local interest may be of no significance nationally. Indeed the local priorities might vary in different areas, resulting in inconsistencies nationally. Similarly, the degree of detail on which locations (particularly streets and BLPUs) are recorded could be quite different, resulting to issues of granularity. Typically, local gazetteers require all used locations to be included, whilst national gazetteers only require locations that are significant on a wider basis.

### **1.4 Issues of national consistency**

When local gazetteers are created separately by different independent organisations, it is inevitable that there will be variations in content, consistency and quality, due to different capabilities, resources and commitment amongst those creating and maintaining the data. Many of these issues are to do with people, management, organisation and copyright, which are outside the scope of these Guidelines.

A nationally agreed and implemented specification is essential; the existence of the Standard alone cannot ensure consistency. Whereas local specifications may be extended according to local requirements there must be conformance with the national specification. The nature of national data specifications is described below.

An exchange format for the transfer of data to the national level will be highly desirable. This combined with a software validation process to make basic checks on data formats and domains will considerably facilitate the creation of a national gazetteer.

It will also be a big help to have tools for capture, validation and output of data which conform to the national specification and can be shown to do so. Ideally, this should be through a national certification process whereby tools are taken by a neutral party and subjected to an agreed set of controlled tests.

Those creating the local gazetteers need to understand the national requirements. Thus, training and adequate instructional material and guidance produced at the national level are likely to be needed.

It is important to ensure user feedback from any usage at the national level and a means of rapid correction at the local level.

## ***2. How to amalgamate 'local' gazetteers into a 'national' one***

### **2.1 Definition of national scope**

The scope for a national gazetteer will be different from that for a local gazetteer. These differences could include extended territory of use, more restricted rules for inclusion, and reduced currency. These should be identified in the scope statement for the national gazetteer.

### **2.2 National data specification**

As for any gazetteer implementation, a detailed data specification is required, conforming to the standard. At the national level, this should include the following:

- Description of the details of the implementation, defining the data structures to be used;
- Rules for inclusion of instances of the location and their identification;
- Implementation-specific classification schemes to be used, for example standard lists of codes used in the description and naming of geographic objects e.g. DfT road classifications, ONS local authority code lists;
- National referencing schemes to be used, establishing consistent supporting geographies where possible for a national system of spatial units, for example:
  - standard lists of localities and towns to be used,
  - inclusion of prefixes on identifiers, to make them nationally unique,
  - inclusion of higher-level fields in an address;
- Identification of the attributes of each class of location to be provided;
- Specification of cross-references required;
- Target levels for gazetteer currency with frequency of submission of updates.

### **2.3 Capture and maintenance rules**

As for any gazetteer implementation, practical guidance is required for such things as:

- Procedures for initial capture of data;
- National agreement and mechanisms for the assignment of unique identifiers;
- Procedures for maintenance of data, including frequency of update;

- Statement of what objects may be held at a local level, but is not required at the national level (which may be identified by means of ‘flags’ in the implementation). These excessive items should be filtered out before submission to the national dataset;
- Establishment of nationally recognised “standards”, against which data can be tested (e.g. Ordnance Survey large-scale mapping for coordinates);
- The format for transferring the data to the national level;
- National quality standards, specifying Acceptable Quality Levels, which cover not only logical consistency in terms of format and domain, but also other aspects set out in the Standard, including associations;
- Details of quality control and quality checks to be carried out locally, including methods, measures, reports and Acceptable Quality Levels.

#### **2.4 Data differences at the national level**

Some objects may be different at the national level compared with their occurrence in local gazetteer. Some may need to be amalgamated into true ‘national’ objects, for example trunk roads, motorways, rivers and BLPUs that cross local boundaries.

Descriptive identifiers may provide a unique spatial reference within the local territory of use but they may not be unique nationally. For example a locality or town name may be unique in a particular county but not nationally, this should be covered in the national data specification but checks will still need to be made.

#### **2.5 Quality control and quality assurance**

The quality of any national gazetteer compiled from local gazetteer is going to be heavily dependent on quality management at the local level. Acceptable quality levels for local gazetteer submissions will have to be determined and agreed together with the method of testing to establish that levels are being met.

On initial submission, each local gazetteer should have a quality report. This will need to be updated from time to time as agreed nationally. A protocol will have to be in place for dealing with data that does not meet nationally agreed quality levels.

Even with these processes established, it will be necessary to assure the quality of the data submitted to achieve as much consistency as practical. This will have some limitations. Checks on the ground or even of local data sources are unlikely to be feasible. Periodic audits of both local quality management processes and data could be considered if the value of the national gazetteer justifies it.

It will not be possible to replicate all the tests carried out at the local level or to establish that all the acceptable quality levels have been met. A more limited set of acceptable quality levels will need to be established which tie-in to those at the local level.

Most tests at the national level are going to have to rely on the gazetteer data itself, typically tests of logical consistency. For content correctness, nationally available datasets such as PostcodeAddress File (PAF) and Ordnance Survey MasterMap or

Land-Line can be used but these have currency and scope limitations. It is suggested that the following at least should be checked at the national level:

- **Logical consistency**
  - **Format:** is the data in the correct format, i.e. are all fields of the specified data type (this can be checked by software processes);
  - **Domain:** do the attribute values fall within expected ranges (this can be checked by software processes), these might include:
    - Reference numbers in pre-defined ranges,
    - Codes from the national pre-defined lists,
  - **Temporal consistency:** date not before a certain date, and not in the future (this can be checked by software processes),
  - **Association consistency:** are cross references valid (it should be possible to check these by software processes). For example:
    - Streets referenced by BLPUs,
    - Primary addressable objects referenced by secondary addressable objects.
- **Completeness** – this is likely to require manual checking of a sample and comparison with other national datasets:
  - **Commission** : are there entries for objects or locations that are out-of-scope such as lamp posts in a land and property gazetteer, or duplicated (particularly due to overlaps of data scopes between adjoining local gazetteers);
  - **Omission**
    - are the total number of entries in the gazetteer at the level expected for the current date. This will difficult to determine precisely but alternative sources should indicate the approximate number expected;
    - are all mandatory attributes present (this can be checked by software processes);
- **Attribute accuracy** - for many attributes, it is possible to define a set of valid values (some software checking or software tools may assist with this). These might include:
  - names, by checking single occurrences of names used in spatial references;
- **Positional accuracy** - are the individual coordinate references for the locations within the territory of use of the local gazetteer (software could be used for this if the polygon defining the territory of use of the local gazetteer is known).
- **Granularity:** are the objects at the right level of granularity, for example are streets excessively subdivided (this will require manual checking).

However, a great deal can be gained from an initial overall assessment. A visual examination of the raw data in the local gazetteer will give confidence, or otherwise, in the quality of the data.

Further details of tests are described in the section on Quality Assessment and Reporting. At the conclusion of all testing of a local gazetteer, a test report should be sent back to the data originator. This should identify the tests performed, who performed them and when, and the results. For further details of test reports, see the section on Quality Assessment and Reporting. Any errors found, either specific or general, should be identified, requesting that they be investigated and corrected. Any datasets failing to reach the Acceptable Quality Level should be rejected.

Where tests can be performed by software, these could be carried out by the supplier of the data as a pre-condition for submission of the data. This requires nationally available software tools for data validation.

To conform to BS 7666 it will be necessary to issue a quality report for the national gazetteer. This needs to summarise the quality that can be expected at the national level. The accuracy of this report will depend upon:

- the rigour of the quality controls and quality assurance in the production and maintenance of the local gazetteers;
- the achievement of nationally established acceptable quality levels locally;
- the accuracy of the quality reports accompanying the local gazetteer submissions;
- the checks applied to the local gazetteers on receipt at the national hub.

National users will expect national consistency in the quality of the data. Reporting average values for quality measures which mask large variations at local level will not endear users to the national gazetteer.

### **3. Maintenance**

#### **3.1 Introduction**

Data maintenance is important in all gazetteers. For a national gazetteer, the issues are more than just processing updates from the local gazetteers. There are additional issues of consistency, currency and frequency of update. There is likely to be great variability in the maintenance regimes for the different local gazetteers. Whilst it is relatively straightforward to obtain updates on a regular basis, it is likely that the currency of these updates (when the real-world change occurred as opposed to when the local gazetteer was updated) will vary considerably. This will lead to variations in quality across the national gazetteer.

#### **3.2 Processing change from local gazetteers**

Changes resulting from changes in the real world will be of three basic types:

- **Changes to existing instances:** the changes will need to be checked, including validity of the identifiers;
- **New instances:** as well as checking the individual entry, its relationship to any other location will need to be checked, to ensure referential integrity;
- **Closure of instances:** any references to the location in other gazetteer entries will need to be checked, to ensure that no orphan records are created, for example a street should not be deleted while there are BLPUs that reference it. Deleted entries should be archived to provide a historic record.

There will also be changes resulting from the correction of errors. These can be any of the above.

A record of history should be maintained through periodic archiving of versions of the gazetteer.

#### **3.3 Change at the national level**

Some objects in the national gazetteer may not occur as such in a local gazetteer, or may occur in part in different gazetteers. These may include motorways, high-speed rail links etc. Separate process will need to be set up to deal with changes to these. Notice of some of these changes may come from sources other than local gazetteers.

#### **3.4 Quality control**

As well as checking change data on receipt, additional testing of the data against that already held will be required. This should particularly check for issues of consistency, for example duplicate entries being introduced, deletions of entries not held nationally, conflicts with other entries and changes being reversed. The methods outlined earlier will be applicable here. Most queries will have to be referred back to the data source for resolution.

From time to time, the national quality report will have to be re-issued, as indicated above. This will rely heavily on local quality management and reporting. See the Section on Quality Assessment and Reporting for more details.