Guidelines for BS 7666:2006

Section 2. How to create a gazetteer of a new type of geographic object

Version 1

January 2007

This Section of the Guidelines covers how to create a gazetteer of a new class of geographic object. Part 0 of BS 7666 defines a generic gazetteer model that can be used for any class of geographic object, and this section describes how to implement this. It describes the planning and initiation of such an implementation and general principles of gazetteer construction. Finally the process of creating a new part of BS 7666 is outlined.

Contents

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

BS7666 has become synonymous with streets and land and property units. However, since its inception, it has been the long-term intention to extend the Standard with further parts specifying gazetteers of other types of geographic object. This has been done with the new Part 5, for delivery points. The purpose of Part 0 is to define a generic gazetteer model that can be used for any type of geographic object. It can be applied to any type of geographic referencing object, i.e. ones that are used for referencing and locating other business information.

A gazetteer can be produced for any type of geographic object. What will differ between them is the form of spatial reference used.

Examples of geographic objects that are within scope:

- Country
- Administrative area
- Census output area
- Town
- Locality
- Watercourse (rivers, lakes, etc)
- Harbour
- Industrial site
- Military site
- Hospital
- Public buildings
- National park
- Environmental protection area
- Contaminated land
- Planning zone
- Development land
- Estate land

Note that these do not have to be national but may be restricted to a specified territory and may be restricted by ownership, i.e. either public or private.

---

2. Planning and initiation

2.1 Definition of scope

Before starting any gazetteer implementation, it is necessary to define the scope of the gazetteer. This should describe the types of object to be included in the gazetteer, with the rules for inclusion or exclusion. In defining the gazetteer scope, it is necessary to take into account why the gazetteer is required, what purpose it will serve, and the type of information that will be linked to geographic objects contained in it.

Example of gazetteer scope
Localities in Great Britain, defined as named neighbourhoods, suburbs, districts, villages, estate or settlements (but not towns) that are used to reference several properties, whose name is recognised by the local authority for purposes of addressing.

2.2 Data specification

BS 7666 defines the general structure of the gazetteer. It does not define the details of content. As part of any implementation, it is necessary to specify in detail the data to be included. This should include the following:

- **Description of the details of the implementation of the Standard**, defining the data structures to be used (use of additional attributes, field lengths, domains, classification systems and codes etc);

- **Rules for inclusion of instances of the geographic object** and their identification;

- **Identification of the source of the data**: a process needs to be established to recognise, identify and label instances of the geographic object. This is likely to be from existing data sources, for example records of properties recorded for a particular purpose. It should include new instances being created, where a well-defined business process needs to be established, including how the new instances are referenced, for example a process of naming or address creation;

- **Identification of the attributes of each class of geographic object**: this may include ones additional to the requirements of the Standard.

2.3 Object classifications

Objects in the gazetteer may not all be the same type. Two levels of classification scheme are allowed. The primary classification scheme should be used for high-level, external purposes, to roughly divide the objects.

Example classification of localities:
Localities to be captured in a gazetteer could be classified by:

1. Rural villages or settlements

2. Suburbs of towns

3. Industrial or trading estates

4. Other

A secondary classification scheme can be used for more detailed classifications, such as internal or application-specific classifications. Note that the secondary classification scheme need not be a refinement of the primary classification scheme, and may be independent of it.

2.4 Data maintenance

A gazetteer is not a static dataset, but a continually changing description of a set of real-world objects. Consequently, it is essential that a maintenance regime is established. There are three main stages in the life-cycle of an object record in the gazetteer, creation, change and closure. Different procedures are required for each.

- **Creation**: a business process needs to be devised to identify new instances of the geographic object in the real-world, and to collect the necessary data about them. This will involve some level of interaction with the life-cycle of the object, for example notification by some regulatory body.

- **Change**: change to a gazetteer entry can occur for many reasons. They essentially fall into two categories, those representing real-world change and those due to correction of current data or insertion of missing data..

- **Closure**: a business process needs to be devised to identify when instances of the object cease to exist in their recorded form. This may involve some level of interaction with a regulatory body. The gazetteer record for this instance is then amended to change its state, and to input a value for the end date. Historic records should not be deleted, as they may be still be of interest, but may be archived.

2.5 Data quality

Quality levels and the processes required to control and assure that these levels are maintained need to be established at the outset of gazetteer creation. How this can be done is described in more detail in Section 3.

3. General principles of gazetteer construction

3.1 Mandatory, optional, conditional and additional attributes

In the Standard, attributes are identified as being mandatory, optional, or in some cases conditional.

Mandatory attributes should always be provided. **Optional** attributes should be provided, except where they are not applicable, e.g. for additional languages in a non-English Gazetteer, or where what is referred to does not exist or is not required, e.g. some additional fields in an address. In some cases, an attribute is recorded in the Standard as being **conditional**, and a condition is stated when it should be provided. In this case the attribute is mandatory when the condition is satisfied.

In certain instances it may not be possible to determine the value of an attribute. Null values for mandatory attributes should be avoided since it is ambiguous whether the value has been omitted in error or is unknown.

Additional attributes may be provided in an individual implementation. These should be specified in the data specification.

3.2 Spatial references

A spatial reference is a description of a real-world place which can then be used to reference other information. A coordinate reference is a particular type of spatial reference, but here we are primarily concerned with non-coordinate references, based upon names of real-world places. The best example of this is an address. A general address structure is given in Annex C of Part 0, and is used in Parts 1 and 2. Depending on the type of object included in the gazetteer, and the territory of use, variations on this may be required.

Examples of spatial referencing system
- Geographic address: object name or number, street name, locality name, town name, administrative area name (see section on **Gazetteers and Addressing**);
- Postal address;
- Postcode;
- County name;
- Country name or code.

Where a new spatial referencing system is required for a gazetteer, it should be defined as specified in **4.2.4** of Part 0 of the Standard.

3.3 Geographic extents

Geographic extent is a detailed description of the "footprint" of the geographic object, recorded either as a collection of one or more geographic objects or as one or more boundary polygons. Where the extent is described by a collection of smaller geographic objects, the identifiers of these are recorded. An example of where this might be used, is Regions defined as collections of counties and other local authority areas, where the extent is described by the identifiers of these local authorities. Where the extent is described by boundary polygons, these may be recorded by either a set of coordinates, or by polygon identifiers such as TOIDs (identifiers for topographic objects in Ordnance Survey's MasterMap® product). Multiple polygons may be used for non-contiguous areas, and "cut-outs" may be used to exclude inner polygons, for example for some current counties. These are illustrated in Figure 1.
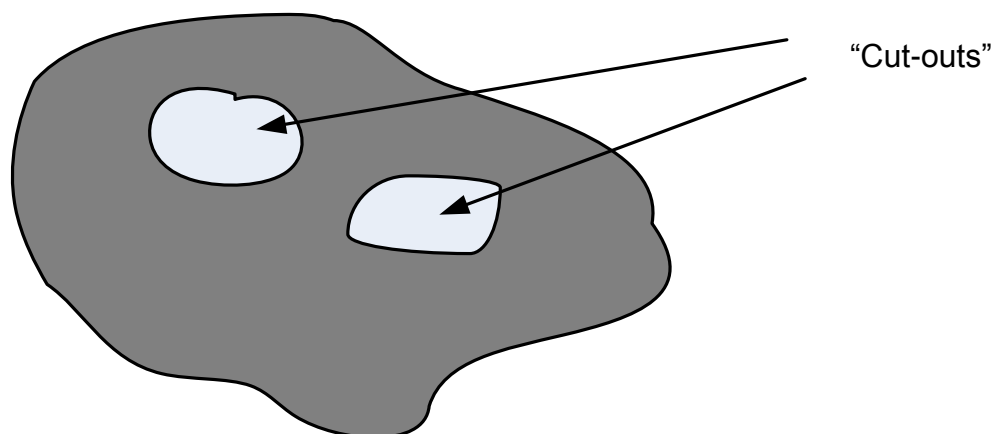

Figure 1: "Cut-outs" within a polygon

3.4 Dates

Dates should be recorded at an appropriate level of resolution. Normally this will be a day, but where this is not known, it may be only a month (e.g. '2006-08') or a year (e.g. '1900'). The dates should be recorded consistently either in the basic format (YYYYMMDD) or extended format (YYYY-MM-DD, YYYY-MM or YYYY), where YYYY is the year, MM the month and DD the day. The two formats should not be mixed, and for each implementation of the Standard it will have to be decided which format is to be used.

Care needs to be taken to distinguish between actual dates, when something happened either in the real world or to a source of information, and capture and update dates, when changes are made to the data.

Actual dates:

- **start date:** the date at when the geographic object came into existence;

- **end date:** the date when the geographic object ceased to exist;

- **current state date:** the date when the geographic object came into its current state;

- **current date:** the date at which the gazetteer is considered to be current.

Capture and update dates:

- **entry date:** the date when the geographic object record was entered into the gazetteer;

- **update date:** the date when the geographic object record was last updated;

- **metadata date:** the date when the metadata was last updated.

Where dates are not known exactly, a notional date at which the date criterion was known to be correct should be used. This applies in particular to start date. Where update date is not known, or the record has not been updated, the update date should be same as the start date.

3.5 Links to other objects and gazetteers

In common with the other Parts of the Standard, Part 0 specifies a facility for linking to other objects and gazetteers. How this is implemented will depend on the nature of the object and the corresponding related objects and also the nature of the linkage. The linkage is always geographic in nature but the objects do not have to be coterminous or exactly coincident spatially. The relationship may involve time as well as space.

In all cases, the link will take the form of the identifier or identifiers of the related object or objects in the other dataset or gazetteer. The nature of the relationship and the dataset to which the data is cross-referenced should be identified in the metadata (see section on **metadata**). This relationship need not be one-to-one (i.e. an object in one dataset may relate to more than one object in another dataset. The correspondence need not be exact as described above).

Examples of possible types of relationship:

simple cross-referencing: objects such as land parcels correspond to those in another dataset. The cross-reference is then between the respective identifiers of the same instance of the object in the gazetteer and the other dataset.

Temporal: an object now occupies the same location as that occupied historically by another object, for example, a plot of land which has now been built on. The cross-reference is between the identifier of the historic object and that of the object which has replaced it (wholly or in part).

Complex cross-referencing: the objects within scope of the gazetteer are similar but not identical to those recorded in another dataset, for example the objects in a land and property gazetteer and the topographic features on a digital map. The cross-reference is then between the identifier of the gazetteer object and those of the instances of the topographic features that most closely represent the object.

Parent-child: one object (the child) cannot exist without a corresponding higher level object (the parent). The child is often formed through subdivision of the parent object. The parent and child may be the same class of object, for example Primary Addressable Objects and Secondary Addressable Objects such as a flat within a larger property or different classes of object, for example streets and elementary street units. Explicit provision is made in the standard for parent-child relationships.

Other geographic objects: for example objects in the same gazetteer whose geographic footprint overlaps, for example buildings on different vertical levels.

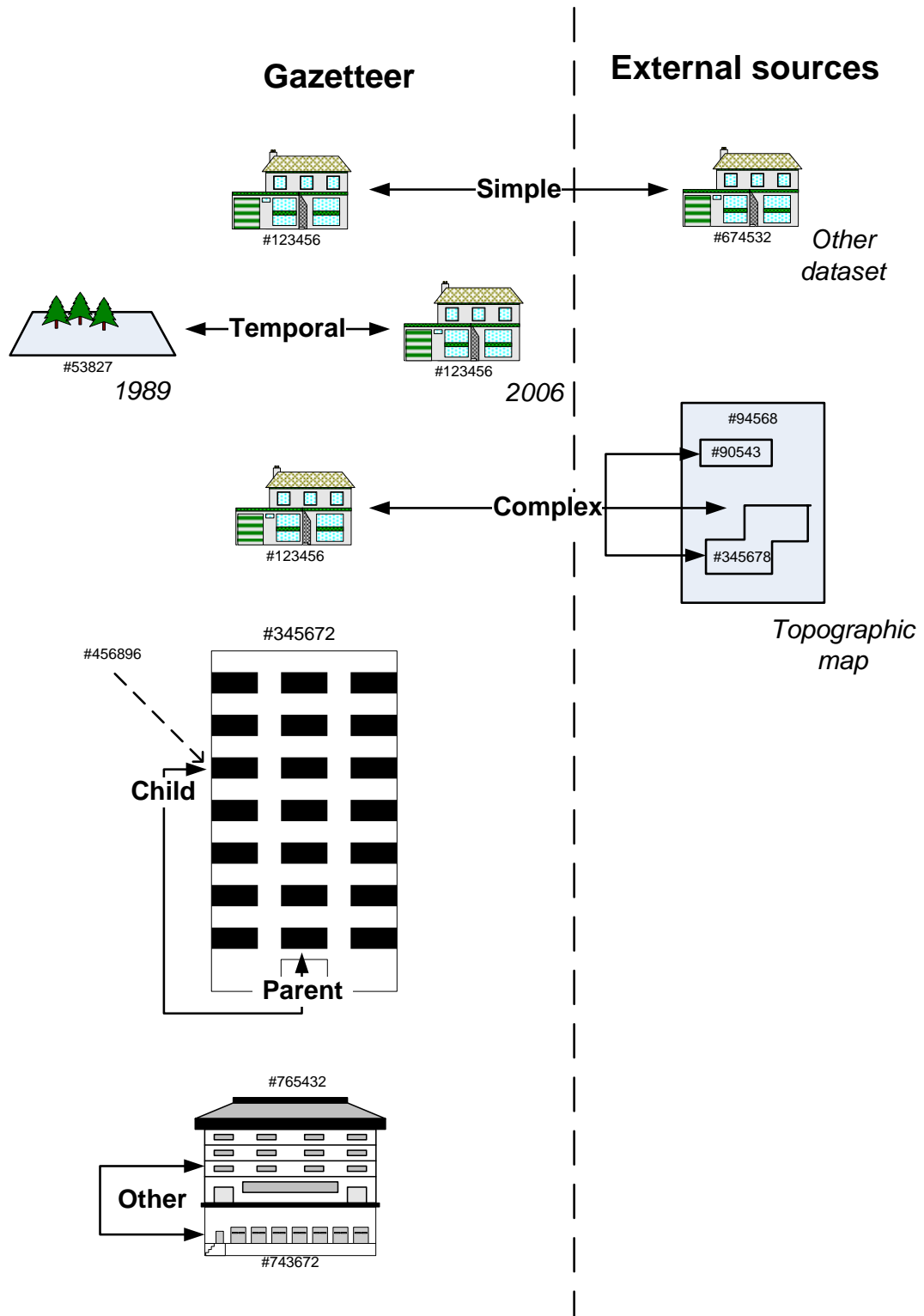These relationships are illustrated in Figure 2.

Gazetteer   **External sources**

Simple

#123456   #674532   *Other dataset*

#53827   ←Temporal→   #123456

1989   *2006*

#123456   ←Complex→   #94568
                                      #90543
                                      #345678

                                      Topographic map

#456896   #345672

Child

Parent

#765432

Other

#743672

Figure 2: Possible types of relationship which can be recorded in a gazetteer

3.6 Metadata

The Standard mandates the provision of metadata for all gazetteers. The main purpose of the metadata is to describe the dataset. It not only provides basic information for operational purposes, but can be fed into a metadata service such as *gigateway*[1] for general data resource enquiries.

Clause **4.2** of Part 0 of the Standard specifies various metadata items:

- **Name**: the formal name of the gazetteer;

- **Scope**: a description of the types of objects to be included in the gazetteer e.g. trunk roads;

- **Territory of use**: the area covered by the gazetteer, e.g. 'England and Wales', 'Great Britain', 'Scotland', 'Northern Ireland';

- **Language**: any non-English languages used in the gazetteer, for example Welsh. Note that this refers to the language used for making the entry not the words themselves, as for example an English language gazetteer may contain words in other languages. If only the English language is used, then this need not be stated;

- **Character set**: any non-English character sets used, e.g. Gaelic. If only the standard character set is used (as for English), then this need not be stated;

- **Gazetteer owner**: the organisation with overall responsibility for the gazetteer. Note that this may not be the same as the gazetteer custodian;

- **Gazetteer custodian**: the organisation or officer responsible for the compilation and maintenance of the gazetteer;

- **Coordinate system**: the system used to define coordinates of points in the gazetteer. This is likely to be the National Grid of Great Britain for Great Britain, and ITM (Irish Transverse Mercator) for Northern Ireland. Note that these are land based systems, and gazetteers that cover wider areas are likely to use some form of latitude and longitude (e.g. ETRF 89);

- **Coordinate axis units**: the unit of measurement for the coordinates, for example metres;

- **Metadata date**: the date of the last update to the metadata;

- **Spatial referencing system**: this is likely to be specific to the type of object included, but in general will be some sort of address or set of references to higher level units such as countries;

- **Primary and secondary classification schemes**: the classification schemes used;

- **State coding scheme**: any scheme used to describe the operational state of objects in the gazetteer, for example the stage reached in its lifecycle;

---

[1] The UK GI metadata service - see gigateway.org.uk

- **Current date**: the date at which the gazetteer is considered to be current. Note that this is different from the date of last update of the data;

- **External cross-referencing schemes**: any external cross-referencing schemes used.

Some of these metadata elements are mandatory, and some are optional. Optional items should be provided whenever they are applicable and known. Multiple values may be given for language, character set, spatial referencing system and external cross-referencing scheme, where appropriate.

3.7 Data quality

The Standard specifies a requirement for a data quality report. This is an assessment of the data in terms of the following:

- Lineage

- Currency

- Positional accuracy

- Attribute accuracy

- Completeness

- Logical consistency

These are discussed in more detail in Section 3 of these Guidelines.

4. How to create a new part of BS 7666

4.1 The need for further Parts

BS 7666 defines gazetteers for various types of geographic objects. There are specific Parts covering gazetteers of streets (Part 1), land and property (Part 2) and delivery points (Part 5). Further Parts could be created for gazetteers of other types of geographic objects if required.

Part 0 can be used to define a gazetteer of a different type of geographic object. As with Part 5, it might be considered desirable to make a specific additional Part of the standard to cover this. This is likely to be done when:

- There is a clear business case;

- There is a widespread requirement in a specific user community;

- There will be implementation in a range of organisations;

- There are specific spatial referencing mechanisms that need to be clearly defined;

- There is a need to collect data on a local basis for creation of a national dataset.

4.2 The standards creation process

BS 7666 comes under the auspices of BSI Committee IST/36 *Geographic information.* This committee proposes and provides technical approval of British standards. It is sponsored by the Association for Geographic Information (AGI).

The stages in the creation of a standard are as follows:

- Business case approval: a formal proposal is made to BSI, to ensure that the proposed standard meets their requirements, and on approval leads to formal set-up of a Standard development project;

- Obtain resources: because of the formal nature of a standard, professional drafting is required, and funding for this may be obtained through DTI, which can take some time;

- Set-up Steering Group: this will contain representatives of the major stakeholders and will be responsible for the overall conduct of the project;

- A series of drafts are produced for review by a Working Group, culminating in a Draft for Public Comment (DPC);

- Public consultation: the DPC is published by BSI, and formal comments are invited from the public;

- Review of comments: a Review Group goes through all the comments received, produces a response to them, and makes any required changes to the draft standard;

- BSI conformance: BSI Editorial Department ensure that the standard conforms to the rules for standards, and arranges the presentation of the document;

- Technical approval: IST/36 reviews and approves the standard for its technical content;

- Publication: this is done by BSI who determine the price of the publication, and provide some marketing.

After five years, all standards are subject to a formal review. This may recommend one of the following:

- To ratify the standard for a further five years;

- To issue technical amendments;

- To create a new edition of the standard;

- To withdraw the standard.