Unique Object Identifiers within the Digital National Framework (DNF)

– Consultation on proposed changes to identifier management.

14th November 2005

1. Purpose of this paper

In March 2000 Ordnance Survey issued several consultation papers regarding the Digital National Framework (DNF). These papers covered the concept of the Digital National Framework and supporting aspects of Metadata, Data Association and Unique Identifiers [all Ordnance Survey, 2000]. Since that time unique identifiers have been implemented by various organisations, including Ordnance Survey through the OS MasterMap layers. In the light of five years of experience the need for modest changes has been recognised for some time. This paper will describe why these are necessary, what they are and seek your views on the proposals.

This paper has been prepared by the DNF Expert Group (DNF Expert Group, 2005) in conjunction with Ordnance Survey (as publishers of the original consultation paper on unique identifiers in 2000). It has been jointly published and distributed by the DNF Expert Group and Ordnance Survey. Policy on this issue will be guided by the DNF Expert Group in future.

2. Definitions

For the purpose of this paper the following terms are used to describe two different kinds of object which are assigned unique identifiers:

a) Those reference database objects that represent identifiable features in the real world. These objects model topographic features (or their equivalents in the marine environment) that most people would recognise - such as a walls, hedges, buildings, fields etc. These objects will be represented in a database as points, lines or areas. In DNF such objects will be found in the reference data (and may be supported by segments of closing geometry e.g. inferred links).

b) Those objects that are abstract representations defined by a user's application. Potentially and collectively, this is a much greater set of objects formed by points, lines or areas into an abstraction representing the user's application "view" of the world. For example a local authority land terrier would define all property extents in terms of land and buildings occupied and then used for taxation, land charges searches etc.

Similarly land ownership title extents, farm registers, river networks, road networks, planning zones, statistical output areas also demonstrate other examples of user

defined objects. Application information can then be attached to any of these objects (e.g. traffic flow counts, surface condition reports, owners name and mortgage details etc).

Further information will be found in the DNF Model (DNF, 2005).

A user-defined object might itself be collected into higher levels of aggregation for example local river stretches into main river stretches between junctions and thereon to the entire river e.g. River Thames from source to sea.

Each of the above individual objects, either at reference base level or at aggregation level would be assigned a unique identifier. From this perspective and from experience to date, it is evident that while some of these objects have already been created today either as text records or geographic database objects and considerably more have yet to be created.

3. Identifiers

The purpose of the unique identifier is to unambiguously identify a reference object, a user-defined object or an object at some level of aggregation.

Therefore the aim is to support:

- users who create data objects in the base reference data
- users who create data objects referenced to the reference data e.g. collections of reference data objects to form a property extent or river network.
- cross-referencing of different user views using data association
- data sharing of application data by using identifiers as a reference.

4. Why there is a need to make some changes

There are several aspects that require attention.

4.1 The DNF Identifiers paper, issued in 2000, proposed that users could be allocated a block of identifiers from the total population of 16-digit TOIDs. This was subsequently agreed and several organisations did this. From 2001 Ordnance Survey agreed blocks of unique identifiers with several organisations. A record has been maintained of the identifier block ranges of those organisations. For example a range of identifiers was agreed with the Countryside Agency to support the Countryside and Rights of Way Act 2000. All access land objects are now referenced by a user-defined identifier representing collections of topographic reference objects.

4.2 The adoption of Geography Mark-up Language (GML), (OGC, 2005), based on XML, required that the first character of any identifier using that standard must be non-numeric. Since TOIDs are numeric identifiers, when publishing OS MasterMap, Ordnance Survey chose to prefix them all with “osgb” on publication in GML in order to meet this requirement, resulting in a prefix and identifier e.g. <osgb><1234567890123456>.

4.3 Formatting issues have since arisen as software vendor's tools have been developed to support the unique identifier. Some systems hold the unique identifier as a numeric integer value, and others as a character string. In some systems the prefixes are omitted and in others they are included, and some pad the numeric portion of the identifier with leading zeroes. Collectively these inconsistencies undermine the ability to link information and share data.

The proposals that follow seek to remedy these issues and take advantage of this opportunity to simplify the allocation method as well as promote greater consistency.

In 2005 Ordnance Survey issued a statement regarding the royalty-free use of the TOID [Ordnance Survey, 2005]. The proposals outlined in this paper have no material effect on that policy paper. Indeed it is hoped that anyone responsible for unique identifiers will also make their use available free of charge and without restriction as cross references to support greater cross-business interoperability and data sharing.

5. The Proposed Changes

5.1. It is proposed that the allocation of blocks of TOIDs should be discontinued, and replaced instead with the registration of 4-character prefixes by different organisations or departments such that the combination <prefix> + <identifier> will always be unique within the Digital National Framework.

5.2. In order to support this change, we would recommend that systems dealing with unique identifiers should store and exchange them in full (i.e. as a character string including the prefix[1]).

6. Implications of the proposals

The implications of the proposals are:

6.1 Existing allocations

The current allocation of blocks of numeric identifiers somewhere within the range 0000 0000 0000 0001 – 9999,9999,9999,9999 will no longer be relevant and those organisations holding such blocks will be free to use all other values within the full range as well, so long as they register a prefix and incorporate this in future.

6.2 New users and identifiers

Any organisations who require an identifier range in future would simply register their four character prefix with the DNF Registry. This would automatically check that their preferred four character code is not already registered and confirm registration. The user would then be free to use the full range of identifiers as they wish.

6.3 Numbers and characters

[1] This maintains compatibility with the Open Geospatial Consortium's "FID" architecture.

While the numeric form has attractions in terms of storage and comparison, it is more difficult to enforce a common method of formatting and storage across software, systems and many users, which is clearly a major impediment to data linking. For this reason the preferred method in future is to hold the value as a character string. Clearly some systems will need to be modified to enable this.

6.4 Use of existing identifiers

In many cases, as noted above, users may already hold a unique identifier in their records referring to a geographic object e.g. via an UPRN, field parcel reference, USRN or a Title Number etc. This may or may not match a 1:1 correspondence to the application object (see 2.b above) when it acts as a identifier linking together several features in the reference base (or associated reference data). It is therefore impractical to provide advice here that covers all circumstances. The main options are to:

- use the existing identifier, and include the prefix on publication.
- create an object identifier as outlined in this paper and cross reference it with the identifier in their existing records.

6.5 Recommended form of DNF identifier

The form will be:

<prefix><identifier>

Where the <prefix> is four characters.

Whilst any form of <identifier> is theoretically possible, comparison, indexing and searching becomes exponentially more complex as the number of varieties increase. It is therefore *strongly recommended to support consistency* that the form of the full object identifier is a 4-character prefix followed by 16 “numeric” values, (held as characters) i.e. 0-9 including leading zeros. Numeric ranges of less than 16 could also be adapted by adding leading zeros (as in an existing identifier as described 6.4 above).

6.6 Versioning and Object Lifecycles

Whilst identity is important and the focus of this paper, the object version number and formally managed lifecycle are also key aspects of object management to support data linking and data sharing at any level to which third party information may be attached. These aspects will be covered in other papers on the DNF website the near future and interested parties should contact info@dnf.org for more information.

6.7 Intelligence in identifiers

As in earlier papers the arguments regarding the dangers of building intelligence into an identifier remain strong and this position is retained in this paper. The recommendation is that the <identifier> segment should continue to contain no intelligence or internal coding.

7. Timing

On closure of this consultation the policy will be confirmed on the DNF website before the end of January 2006.

Given a positive response it is planned that the changes will applied from April 1st 2006. However the registration of the prefix would be introduced at least one month prior to this.

It is recommended that any system or software changes would be introduced as soon as possible following the confirmation of the new policy in early 2006 and ideally within 12 months of the policy being applied in April 2006.

All those who have been issued blocks of identifiers and are currently registered with Ordnance Survey will be contacted directly as part of this consultation.

Ordnance Survey will then discharge this responsibility to the Expert Group through an automated process to be incorporated in the DNF website.

8. Items for consultation

8.1 Do you agree that the registration of organisation prefix is preferable to the allocation of blocks of identifiers?

8.2 Do you agree that the use of a character string to handle the full identifier will support greater consistency of storage and exchange across organisations?

8.3 Do you have any other comments?

Responses will be gratefully received and acknowledged to info@dnf.org, or either of the names below, by 13th December and by 31st December at the latest.

If you have any queries or need to discuss any items please contact DNF Expert Group members:

- Jamie Justham on 01527 556920 or Jamie.Justham@dottedeyes.com
- Jonathan Simmons on 02380 792597 or Jonathan.Simmons@ordnancesurvey.co.uk

Prepared by the DNF Expert Group and Ordnance Survey.
November 2005.

References:

DNF 2005. Website: www.dnf.org

DNF Expert Group, 2005: http://www.dnf.org/Introduction/Involvement.htm

DNF Model. 2005. http://www.dnf.org/Introduction/Roadmap.htm

Open Geospatial Consortium, 2005,

http://www.opengeospatial.org/specs/?page=specs

Ordnance Survey Consultation Papers 2000:

- Digital National Framework, Consultation Paper 1/2000
- Unique Identifiers. Consultation Paper 3/2000
- Metadata, Consultation paper 4/2000
- Data Association. Consultation Paper 5/2000

N*ote: these papers are no longer available but copies may be provided if required.*

Ordnance Survey 2005. Royalty Free use http://www.dnf.org/Publications/Papers.htm

<end v2.00>