Encoding Semantics in the DNF

Aled Greenhalgh, Philip James, David Fairbairn
School of Civil Engineering and Geoscience,
Newcastle University
Newcastle upon Tyne, UK
{aled.greenhalgh; philip.james; dave.fairbairn}@ncl.ac.uk,
http://www.students.ncl.ac.uk/aled.greenhalgh

SUMMARY
The UK's Digital National Framework (DNF) provides for linking business and application data to a common spatial reference. This paper examines means to encapsulate richer semantics in these links and goes on to report on early work investigating ways that these relationship models, as well as representing more complex relationships within and between concepts embodied in the information can be used at the data level to manage identity and ensure its integrity.

KEYWORDS: *Semantics, databases, DNF, OWL-DL, interoperability*

The UK's Digital National Framework (DNF) is "an industry standard for integrating and sharing business and geographic information from multiple sources" (Anon, 2005). The Framework encapsulates a step away from cartography towards more database-inspired geographic information management, placing increasing importance on linking attributes from external databases to spatial objects. The DNF was conceived by Ordnance Survey (OS), Great Britain's national mapping agency (Ordnance Survey, 2000) and has since been taken forward by the DNF Expert Group representing a number of stakeholders as a means to enable and support integration of business and spatial information. Likewise, the OS's research into semantics is concerned with stepping outside purely cartographic representation of concepts and enabling its geographic information to be automatically exchanged, understood and integrated (Ordnance Survey, 2005c). Both initiatives aim to enable interoperability though providing a base for customers to build on and extend their use of geographic information assets, with a significant central role for OS products.

The goals of the DNF are to:

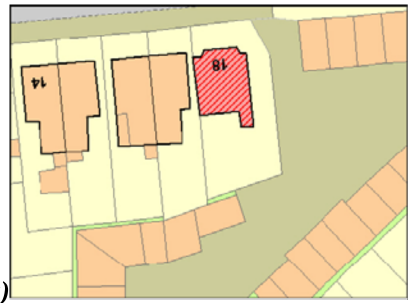
- "Establish a coherent structural model of national reference datasets and the relationship with application information.
- Evolve and maintain a national framework based on this model to support consistent integration of geographic information and thereon enable the easy and increasingly automated exchange and analysis of application information from different sources.
- Establish and evolve a consistent approach to georeferencing and the inter-relationship of application information/data with reference information/data.
- Establish and evolve a consistent approach in the modelling, integrity and connectivity of geographic information." (Ordnance Survey, 2004)

It is proposed within this paper that modelling and encoding semantics within this Framework is wholly within and important to the scope of this aim, and yet that this is an area which has been insufficiently explored, let alone standardised.

SEMANTIC DATABASE RELATIONSHIPS

The UK leads the world in many aspects of Spatial Data Infrastructure (SDI) and this is not least through the availability of a seamless national, spatial, digital database, marketed as MasterMap® (OSMM), that uniquely identifies all its features using the Topographic Object ID(TOID®) (Bamps and Beusen, 2005). The availability of this topographic base has formed the initial enabler of the DNF, which is premised upon linking business and application data to a reference base of spatial features through a system of unique feature references.

The DNF currently sets out only a very simplistic set of linking relationships between geometries. It describes 1:1, 1:n and partial equality between MasterMap features and application database objects, illustrated in Figure 1, below. It allows the creation of 'local TOIDS' for geometries which are not available in MasterMap. The details of this as a mechanism for production and referencing of local application-specific features are currently being consulted on, pending revision (DNF Expert Group and Ordnance Survey, 2005).

a)


b)


c)


Figure 1: Illustration of:
a)1:1 equality
b)1:n partonomy
c)1:partial equality/mereology

from (Ordnance Survey, 2005a)

More advanced relationships than this currently exist within MasterMap:

- Multiple similar representations of a single feature exist where a road is represented as a set of `RoadLink` features, which along with a name make up a compound `Road` feature in the Integrated Transport Network (ITN) layer and also as a `TopographicArea` in the Topography layer

- `Premise` features are linked to a delivery point in the forthcoming Address Layer 2 product

Work within OS on more advanced internal data models (Dolbear et al., forthcoming, Hart and Greenwood, 2003) has created a more complex schema which is capable of referencing varied feature granularity. This enables storage of metadata at sub-feature level, where the name of a feature may have a different source and providence from its geometry, but also allows for production and referencing of aggregate features. Example composite features are `School`s made up of `Building` features.

As stated above, the DNF is focussed on business applications, i.e. live systems where changes and updates are frequent if not constant, as opposed to ‘snapshot’ datasets. With this in mind, an immediate concern is that features represented with greater complexity and denser relationships will create greater maintenance overhead when modifications are made. While the intransigence of a TOID within the MasterMap base and other object identifiers within ‘DNF-compliant’ datasets are one of the key features of the Framework the fact of a dynamic reference base remains a chief complexity. MasterMap TOIDS follow a set of specific lifecycles set out in the specification (Ordnance Survey, 2005b) as should other compliant identifiers. However, this is not to say that these lifecycle concepts are compatible on a semantic level.

APPROACH

The following section reports on initial stages of work on enriching the Framework with a means of representing more complex geographies. The specific focus is on using ontologies and description logic (DL) within the databases, where the same mechanisms which allow us to represent relationships also allow us to maintain them.

The combination of ontology and DL has been used to ensure consistency in databases (Frank, 2001), and to provide semantic integration of multiple representations in spatial databases (Stoter et al). The tools will be used here for both of these purposes and in addition will be used to describe and maintain the specific nature of links between objects in databases.

Ontologies can be constructed at a number of different levels of abstraction, ranging for example, from representing physical reality to representing subjective knowledge and ideas (Frank, 2001). The application here is interested in concepts’ representation in data rather than in the concepts themselves. This requires modelling of concepts where they are either represented in data, they express links between data or where they might lead to changes in the data. This level of focus means we need neither model the concepts with respect to a foundational or standard upper ontology, such as DOLCE or SUMO (Niles and Pease, 2001, Masolo et al., 2003), nor take a wholly data-centric approach, such as provided in (Laborda and Conrad, 2005).

MODELLING

The modelling process initially follows the methodology for formulating an ontology set out in (Mizen et al., 2005) and more fully in (Mizen, 2005). The first stage builds a conceptual ontology as a knowledge glossary of formalised human-readable sentences. This draws upon expert knowledge and texts from the domain, which in this particular purpose may additionally encompass database schemas. This stage of the methodology is modified in this work in placing most focus on those concepts that are expressed as verbs, which by identifying actions and processes contribute to a lifecycle for each concept. Nouns are only captured where they exist in the data or are necessary to fully express a relationship.

The second stage, that of representing these concepts in a machine-readable ontology, will concentrate on representing each relationship as a class. These classes must then be axiomised with spatial and temporal relationships. The need to reason across real-world spatial and temporal changes invites investigation of the dynamic process and event models in (Cole and Hornsby, 2005, Galton and Warboys, 2005) and the vocabulary set out in the SPAN and SPAN ontologies of Grenon and Smith (2004). While it is expected that this body of work will provide suitable concepts for modelling events and processes bringing about these changes this is something which has thus far proven difficult to express satisfactorily in Web Ontology Language (OWL) (Dean and Schreiber, 2004). OWL is taken as the language of choice here as syntactic interoperability, which this choice maximises, is an associated aim.

APPLICATION

The prototypical application and motivating example given here is one linking OSMM to two local authority systems which are described briefly below: the street gazetteer and the highway engineering and asset management system. Local authorities are selected here as an example problem domain for a number of reasons: firstly that they, along with other government agencies, are key Ordnance Survey customers and DNF stakeholders; they have a current mandate to 'join up' datasets and service delivery; and they have disparate and heterogeneous data, although the overall scope is fairly well defined and is strictly so in the geographic sense.

As the street naming authority, local authorities are mandated to hold and maintain a street gazetteer, identifying each street in its area and allocating each a Unique Street Reference Number (USRN). This is also used as the basis for their land and property gazetteer in which all properties must be bound to a street and are allocated a Unique Property Reference Number (UPRN). Both are used for a variety of purposes and systems within individual authorities. The data model for these gazetteers is defined in the BS7666 standard (BS7666-1:2000; BS7666-2:2000), through which both gazetteers are currently being compiled into national-level gazetteers (Intelligent Addressing, 1999). Due to the enforced necessity of binding properties to streets a range of complex semantics ensues in a number of situations such as: in rural areas where a property parcel may not have road access; houseboat and jetties; advertising boards and hoardings; and fish farms. This gazetteer includes private roads and does not cover roadside assets such as street-signs, and emergency call boxes. Motorways and trunk roads are included only where required by attached properties (NLPG Technical Working Party, 2001).

The property gazetteer is the primary dataset used in the authority's business transactions, for instance in collecting council tax and business rates, in managing planning applications and for managing housing tenants. Efforts have been made in individual authorities towards linking gazetteers to OSMM. However, these efforts are rare to the degree that where they have occurred they have been seized upon excitedly as canonical examples of the DNF approach (Simmons, 2005, Higgs and Malcomson, 2005).

Local authorities also generally have highway engineering and asset management systems which coordinate highway works, including with utilities and other government agencies. These systems primarily manage the paved assets of a highway, recording structure and condition of footway, kerb and carriageway. These systems are also standards-based, compliant with and outputting to the UK Pavement Management System (UKPMS) standard (Spong, 2005). This referencing system is wholly network-based, with locations defined relative to longitudinal distance from junction nodes and cross sectional position. The scope of these systems includes footways, cycle paths and individual road lanes. They do not include those motorways and trunk roads managed by the Highways Agency. Topographic area and road length within this dataset are the most significant attributes and are calculated quite differently to either the street gazetteer or MasterMap.

As highlighted above, OSMM contains a number of different representations of roads, as well as footways, kerbs and pathways contained across the Topographic Area, Topographic Line and Integrated Transport Network (ITN) layers.

The three key concepts here then are:

- `Highway`, as in a pavement management system
- `Street`, as in the street gazetteer
- `Road`, as in OSMM

Clearly these are all similar, but still different concepts. Interoperability between each is desirable for a number of reasons including viewing and using highway engineering data in systems based on MasterMap or the local gazetteers. These two systems are used as the basis of an increasing majority of local authority business systems, including planning and development control, electoral registration, public health and trading standards (Greenhalgh, forthcoming). Improved service delivery outcomes from this might include the ability to communicate with properties attached to the highway before streetworks begin, for instance.

The modelling process builds a set of encodings by which these concepts:

1. are represented in data
2. relate to one another
3. can be changed and updated, such that 1 & 2 still hold

Conversion between the two location referencing systems—network-based and grid-system-based—would also be specified in the model.

EXAMPLE

A `Highway` is made up of a network link of assets in the pavement management system and that this can be represented in OSMM as a set of Topographic Areas representing the carriageway and footways either side and a set of Topographic Lines on the boundary between these as kerbs.

$\text{HighwaySection} \subseteq \exists\text{hasAsset.(Kerb} \cup \text{Carriageway} \cup \text{Footway)}$

$\text{HighwaySection} \subseteq \exists\text{connects.(}\forall\text{Node)} \cap \exists\text{connects.} \leq 2$

$\text{Highway} \subseteq \exists\text{hasSection.} \geq 1$

$\text{Carriageway} \cup \exists\text{partOf.Carriageway} \equiv$
$\text{RoadArea.SpatiallyIntersecting(}\forall\text{HighwaySection)}$

$\text{Footway} \cup \exists\text{partOf.Footway} \equiv \text{PavedArea.SpatiallyTouching.Carriageway}$

$\text{Kerb} \equiv \text{Boundary.SpatiallyTouching(Carriageway} \cap \text{Footway)}$

By describing the possible changes to a highway, such as closure, pedestrianisation, and traffic calming, then other datasets, such as OSMM in this example, can be used to represent the same concepts, including consistency across changes.

$\text{PedestrianisedSection} \subseteq \forall\text{hasAsset.(}\neg\text{Carriageway)} \cap \exists\text{hasVersion.(}$
$\exists\text{hasAsset(Carriageway))}$

$\text{LocalReplacementFeature} \equiv \forall\text{PedestrianisedSection.SpatiallyIntersecting.RoadArea}$

$\text{LocalReplacementFeature} \subseteq \text{hasRestriction} \in \text{'No Vehicles'}$

$\text{LocalReplacementFeature} \subseteq \text{hasOrgId} \in \text{'locl'}$

$\text{ReplacementFeature} \equiv \forall\text{LocalReplacementFeature}$

If a section of a street is pedestrianised this would be initially recorded in the highway engineering system. A reasoner using the axioms above could then update the OSMM topographies and the OSMM Integrated Transport Network (ITN) with the changed feature type (‘road’ to ‘general surface’) and potentially with new geometries. It would also remove the ITN *RoadLink* while keeping the definition of the *Street* in the gazetteer whole. This would all the while retain the referential integrity of the database, particularly with respect to the TOID, as shown in the state diagram and database table fragment below.

Identification of change is reflected in MasterMap through addition of new TOIDs and associated geometry, and changes in or removal of geometries along with an increment of the version held against the corresponding TOID. These concepts of change in MasterMap are already set out in a suitably structured way, with four main sources of change described within the OS MasterMap standard: creation, deletion or modification due to real world change; error correction; indirect change for those features which participate in topology rules and those for which a change is made in OS’s internal data but not delivered in the product data. Consideration also has to be paid to the source of change: ‘real-world change’ is inevitably reflected at some point later in OS data, within six months for significant ‘Category A’ change such as new buildings. When the OSMM is then later updated by OS’s processes, the system would then reason that the changed or replaced OSMM feature was a new representation of the feature which had been previously updated locally. Previously formed links, made to the local TOID representing the pedestrianised area, now additionally represented by an OSMM feature area, would need to continue to remain valid as these are representations of the same thing. As links to locally created TOIDS may have been created externally to the database, and hence outside the reasoning universe, each mapping to a replacement must be persistent.

toid_org_id	**toid**	**toid_ver**	**usrn**
osgb	A	1	X
locl	B	1	X
osgb	A	2	X
locl	B	1	X

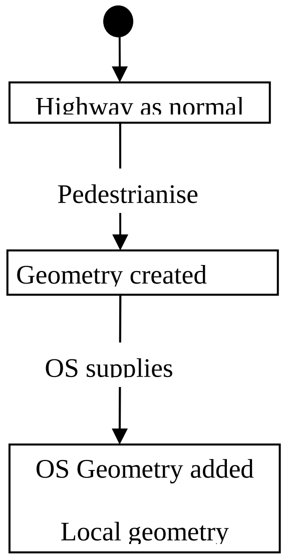


Figure 2: Database table fragment and state model

Setting up these models would enable applications such as defect logging by a member of the public through a call centre. A complaint of a 'missing kerb on Dean Street' would utilise the Street Gazetteer and OSMM used in the call centre application to locate the defect, but would store it natively in the highway engineering system.

CONCLUSION AND DIRECTIONS FOR FURTHER WORK

A key attraction of the DNF is its relative simplicity and potential for more efficient storage and use of some part of the massive body of GI available. Within this there remain innumerable application areas and ever more opportunity to take the approach set out here as more datasets and systems are converted to DNF compliance. However semantic issues frequently appear dauntingly complicated to those information managers tasked with implementation, therefore it is hoped that the approach set out here is one which can clearly be seen to offer greater integrity of information alongside richer relationship description.

In terms of referencing and encoding the semantics identified within the relationships within and between databases, it is suggested that there may be unrecognised potential in constructions like the TOID to reference not just a particular geometry, but also a particular view of that object, a specific *concept* as termed here. This might be attached to Werner Kuhn's envisioned Semantic Reference System (Kuhn, 2003), which is fittingly and closely analogous to a spatial reference system, and as such describes operations of projection and transformation between systems of differing semantic references. It is this model that OS has taken up as an aim for their research (Ordnance Survey, 2005c). Notwithstanding an implementation in (Kuhn and Raubal, 2003) these operations remain abstract, and do not specify the consequences on the transformed/reprojected feature's identity, and hence identifier: specifically whether a reprojected or generalised concept has the same identity as its original or more specific form and as such merits the same identifier.

Current tools provide a limitation to simple implementation of this approach to managing change. Reasoners such as FaCT++ and RacerPro (Racer Systems, 2005, Tsarkov and Horrocks, 2005) follow a request-response model, in which a static-state graph representation of the ontology including instances is passed to the inference engine: the DIG Interface (Bechhofer, 2003) has become the *de facto* standard for this exchange. The response then usually follows some significant time later where the ontology is of any non-trivial complexity. As such the power of currently available algorithms and computer hardware in combination with this communication mechanism means that real-time or near real-time systems cannot be implemented with these tools. Additionally this communication model does not suit dynamic systems on the whole: albeit the OWL-DL open-world logic model is designed to allow reasoning and inference where the domain is not fully expressed there does not exist, at either the logic level nor at the communication level, a method of retracting or modifying assertions passed to the engine. This means that that a new request must be constructed whenever a single change occurs after the initial request.

BIBLIOGRAPHY

Anon (2005) *What is DNF?* Available at: http://www.dnf.org/Introduction/WhatIsDNF.htm, (Accessed: 10th December 2005).

Bamps, C. and Beusen, P. (2005) *Spatial Data Infrastructures in the United Kingdom: State of play 2005.* K.U.Leuven: Leuven, Belgium.

Bechhofer, S. (2003) *The DIG Description Logic Interface: DIG/1.1: DL2003 Workshop.* Rome, June 2003.

Cole, S. and Hornsby, K. (2005) *Modeling noteworthy events in a geospatial domain': GeoSpatial Semantics: Proceedings of First International Conference, GeoS 2005.* Mexico City, Mexico, November 29-30, 2005. Springer.

Dean, M. and Schreiber, G. (Eds.) (2004) *OWL Web Ontology Language Reference.*

DNF Expert Group and Ordnance Survey (2005) *Unique Object Identifiers within the Digital National Framework (DNF) – Consultation on proposed changes to identifier management.* Available at: http://www.dnf.org/Publications/Papers/20051114%20Unique%20Object%20Identifiers%20v200.pdf (Accessed: 12th December 2005).

Dolbear, C., Goodwin, J., Mizen, H. and Ritchie, J. (forthcoming) Semantic Interoperability between Topographic Data and a Flood Defence Ontology**,** http://www.ordnancesurvey.co.uk/partnerships/research/publications/docs/2005/Semantic%20Interoperability_CDolbear_sem.pdf.

Frank, A. U. (2001) Tiers of ontology and consistency constraints in geographic information systems *International Journal of Geographical Information Science,* **15,** 7 667-678.

Galton, A. and Warboys, M. (2005) *Processes and Events in Dynamic Geo-Networks: GeoSpatial Semantics: Proceedings of First International Conference, GeoS 2005.* Mexico City, Mexico, November 29-30, 2005. Springer.

Greenhalgh, A. (forthcoming) *The Value of Standards-Based Metadata in Local Authority GI Management.* Unpublished MPhil thesis. Newcastle University.

Grenon, P. and Smith, B. (2004) SNAP and SPAN: Towards dynamic spatial ontology *Journal of Spatial Cognition and Computation***,** 4 69-104.

Hart, G. and Greenwood, J. (2003) *A Component Based Approach to Geo-Ontologies and Geodata Modelling to Enable Data Sharing: 6th AGILE Conference on Geographic Information Science.* Lyon, France, April 2003. Presses Polytechniques et Universitaires Romandes.

Higgs, B. and Malcomson, E. (2005) *Building a single-geometry land and property database at Dudley Metropolitan Borough Council using the Digital National Framework.*

Kuhn, W. (2003) Semantic reference systems *International Journal of Geographical Information Science,* **17,** 5 405-409.

Kuhn, W. and Raubal, M. (2003) *Implementing Semantic Reference Systems: 6th AGILE Conference on Geographic Information Science.* Lyon, France, Presses Polytechniques et Universitaires Romandes.

Laborda, C. P. d. and Conrad, S. (2005) *RelationalOWL - A Data and Schema Representation Format Based on OWL: Conceptual Modelling 2005, Second Asia-Pacific Conference on Conceptual Modelling (APCCM2005).* Newcastle, Australia, January 30 - February 3, 2005. Australian Computer Society.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. and Schneider, L. (2003) *The WonderWeb Library of Foundational Ontologies Preliminary Report.* National Research Council, Institute of Cognitive Sciences and Technology: Padova, Italy.

Mizen, H. (2005) *OS Methodology for Building Conceptual Domain Ontologies.* Southampton: Ordnance Survey

Mizen, H., Hart, G. and Dolbear, C. (2005) *Ontology Ontogeny: Understanding how an Ontology is created and developed: GeoSpatial Semantics: First International Conference, GeoS 2005.* Mexico City, Mexico, November 29-30, 2005. Springer.

Niles, I. and Pease, A. (2001) *Towards a Standard Upper Ontology: 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001).* Ogunquit, Maine, October 17-19, 2001.

Ordnance Survey. (2000) *Digital National Framework, Consultation Paper 1/2000.* Southampton: Ordnance Survey

Ordnance Survey (2004) *The Digital National Framework.* Ordnance Survey: Southampton.

Ordnance Survey (2005a) *Implementing OS MasterMap / Information sheet 1: Data models and management v1.2.* Ordnance Survey: Southampton.

Ordnance Survey (2005b) *OS MasterMap User Guide - Product Specification.* Ordnance Survey: Southampton.

Ordnance Survey (2005c) *Research & Innovation: Semantic reference systems*. Available at: http://www.ordnancesurvey.co.uk/oswebsite/partnerships/research/research/semantic.html (Accessed: 10th December 2005).

Racer Systems (2005) *RacerPro Reference Manual*. Available at: http://www.racer-systems.com/products/racerpro/reference-manual-1-9.pdf (Accessed: 10th December).

Simmons, D. (2005) *Migrating to a National Street Gazetteer based on the Digital National Framework: Streetworks 2005.* Cardiff, Highway Authorities and Utilities Committee.

Spong, C. (2005) 'Beginner's Guide to UKPMS', in UKPMS User Forum(Ed), *The UKPMS User Manual.* Chris Britton Consultancy: Guildford.

Tsarkov, D. and Horrocks, I. (2005) *FaCT++*. Available at: http://owl.man.ac.uk/factplusplus/ (Accessed: 10th December 2005).