Enabling Information Integrity within Spatial Data Infrastructures - The Digital National Framework Concept

Keith MURRAY, Bern MUNDAY and Ian BUSH, United Kingdom

Key words: Spatial Data Infrastructures (SDI), Land Administration, Standards, Spatial Information Management (SIM), DNF & Case Studies.

SUMMARY

As the economic benefits of using "location" and "geography" within information systems gather pace around the world, especially at government level over the last five years, the need for interoperable information increases. To some the term interoperable means the ability to link two computers together, to others it can mean the ability to overlay one geographic dataset over another.

If the power of geography is to fulfill its potential we have to emulate the banking, electronic point of sale and other information industries and develop our data models to operate in automated processes. With this we can move faster towards realising our investments in geographic information and improving services across all levels of society from determining national policy down to local decision making.

We therefore must look to the information we are using and its fitness for purpose in the application and integrate GI into business information systems. Today we are some way from widespread adoption of truly automated decision making applications. Information may have been digitised by different agencies at different times, for different purposes and in different ways. We should not necessarily expect a cadastral parcel by one organisation and mine working records by another to interoperate at the level that the computer can analyse their relationship and determine a conclusion without manual intervention, especially if we intend this as definitive and for which we might be liable.

The Digital National Framework was introduced in Great Britain in 2000. Since then the principles and methods have started to develop and evolve with the primary aim of improving the collective integrity of spatial datasets that are maintained in a distributed manner, by those organisations whose role it is to maintain that data, but which need to be brought together electronically in the decision making process. The paper describes the principles and outlines the methods. Two case studies are described (index to land ownership parcels and a new database of land now open to public access) and finally progress towards the DNF principles are assessed in the paper.

Enabling Information Integrity within Spatial Data Infrastructures - The Digital National Framework Concept

Keith MURRAY, Bern MUNDAY and Ian BUSH, United Kingdom

1. SPATIAL DATA INFRASTRUCTURES AND MAINSTREAM INFORMATION APPLICATIONS

The use of geographic information is becoming more pervasive and applications are growing in all sectors. Indeed it is within government, stimulated by new policies and targets to modernise services where the greatest impact is being felt.

Not so long ago the use of paper for publication was widespread. Digital mapping disturbed the equilibrium of a century or more, this required investment and commitment and hence each organisation set off on a journey at their own pace in their own time. Now that the data is being recognised as an immensely valuable national resource, as it often is within government, so too are the expectations of the data being raised higher all the time.

Our existing Spatial Data Infrastructures (SDI's) are formed of what data and information we have today. This will often have been derived from previous generations of operation within an organisation. This also varies from country to country where the political, economic and commercial environment also varies. For example at a national level of the Bureau of Land Management in the US, the Land Registry of England and Wales and the Kadastre in the Netherlands all hold important land records to land but do it in very different ways.

The pressure is to move beyond digital mapping to meet the expectations users have today and to be able to share and exchange data in applications ideally from definitive sources. While this is achieved to some extent today, it is very much a "GIS" operation, requiring specialist skills, systems and knowledge. Further, the integration of two or more datasets from different agencies requires processing and invariably some form of human intervention to analyse the results. Where these are "overlays" of for example land use, land ownership and land contamination the digital mapping practices employed in creating these datasets will typically vary. Automated and assured analysis is the expectation but rarely achievable due to limitations of the source data.

If we were to employ an "overlay" approach in transferring funds into and out of our personal bank account, which would rely on "implied connections" how many of us would be comfortable with such an approach? Mainstream Information & Communication Technologies [ICT] relies on explicit connectivity to enable automation and assurance. Our challenge now is to transform the rather laborious and map-centric view of digital mapping and GI processes into ICT application ready information. In this capacity "*data integrity is critical in mainstream information applications.*"

2. WHAT IS THE INFORMATION INTEROPERABILITY PROBLEM?

Information is collected, used and stored by a large number of organisations. Unless they are the national mapping agency or some similar body, geographic information (GI) is often a means to an end for them. An organisation such as a single local authority may manage land occupation records for taxation, hold responsibilities to protect the environment with the support of legislation and provide a variety of other services. There can be hundreds of local authorities in a country. Hence the information will inevitably be distributed.

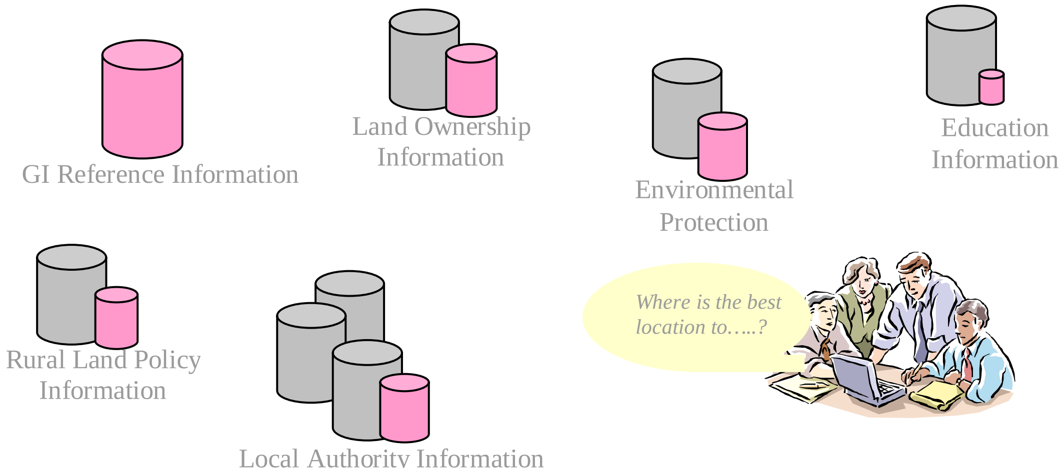


Figure1: Distributed information: Information is distributed where the primary business focus lies and is also best maintained here – but the lack of connectivity **across datasets** can be a barrier to problem solving and business applications.

Not only will the information be distributed, it will often have been collected to different specifications. This may occur even though all the parties may have employed data from the mapping agency. A recent study into the development and maintenance of address lists in Great Britain has revealed significant duplication (VOA, 2004). For area objects, the overlay approach often used in digital mapping readily enables rapid local adoption of digital methods but the lack of a common approach to georeferencing will often be revealed in several ways:

- inconsistency across datasets (addresses, land parcels, assets etc) where consistency should exist
- inability to use the information in automated processes without prohibitive up-front data clean-up costs
- two datasets of the same entity from different sources e.g. a property extent, cannot be easily combined
- massive data duplication – little data reuse
- laborious maintenance systems to ensure the overlay remain faithful to the current state on the ground.

In addition to the lack of a common approach, some argue that the adoption of digital methods is further hampered by:

- the perception that geographic information remains a specialist medium and is not easily incorporated into mainstream ICT applications
- disillusionment by organisations where significant investment in data has not realised the expected business benefits
- information liability issues where the provenance of combined datasets is unclear
- and
- the reluctance of new markets to take-up, integrate and use geographic information.

In summary the lack of consistency and interoperability across organisational (geographic) datasets has become a major issue in realising the full potential of geographic information underpinning the needs across government, commerce, society and for the individual.

3. JOINED UP GEOGRAPHY

3.1 Promoting reuse and connectivity

The need to further develop existing SDI's is well known and well documented. Models vary and evolve according to the political and business climate. So does interoperability; this can mean different things to different people and organisations. In the United Kingdom the push for electronic services had offered an opportunity as well as dangers. In the rush to digitise information it is so easy to take a copy and customise it to meet a deadline for a specific application, the need for wider use and subsequent maintenance are often afterthoughts. The opportunity is to use the thrust for e-services as a way of promoting better connectivity. Hence the concept of the Digital National Framework was born in 1999 to meet the new millennium, and an update on progress has recently been published [Ordnance Survey, 2004].

3.2 Aim of the Digital National Framework

The aim of DNF is as follows:

"The Digital National Framework is a model for the integration of geographic information of all kinds - from national reference datasets to application information at the local level".

and in its implementation:

- *"The Digital National Framework provides a permanent, maintained and definitive geographic base to which information with a geospatial content can be referenced"*
- *"DNF is supported by a set of enabling principles and operational rules that underpin and facilitate the integration of geo-referenced information from multiple sources*"

The primary principles of DNF are:

1. The concept and methods shall be driven by the strategic needs of the wider GI community and the needs of the information industry.

2. Data should be collected only once and then re-used.
3. Reference information/data should be captured at the highest resolution whenever economically possible.
4. Such information may then, where appropriate, subsequently be used to meet analysis and multi-resolution publishing requirements.
5. DNF will incorporate and adopt existing de facto and de jure standards, wherever they are proven and robust.

Hence DNF will foster an environment where users should not need to capture information that already exists. In future information can be reused and added together to form new datasets building on existing proven components. Equally there are many who already hold data and wish to migrate this to improve connectivity and consistency, and there are several ways of achieving this [Rönsdorf, 2005]. The DNF web site will hold several case studies and evolving models to support sharing of experience and a common approach [DNF. 2005].

3. 3 Evolving the framework

The concept accepts that we start from where we are today and if we converge on a common point in the future, we can take steps towards that common point as our business environment permits. We will gain advantages at each stage. There is no intention to define a grand architecture since the models and methods need to work in the short term and have to provide business benefit to the user immediately. However, the issues are often alike across different user areas from land and property to utilities to environment, statistics and so on.

The model involves ”mapping” the ”*user’s application view of the world*” to the physical geography as represented by the map i.e. land ownership, land occupation, land use, land cover – all of which are ”different geographic views”. At the lowest level these are represented by point, line and area features (physical and real world objects), hence the common nature of the issues across the user areas. Figure 2 illustrates the basic model of referencing different data types.

Such models are being developed by user groups of different kinds through case studies, data models and methods of data creation or data migration and these will soon be published on the DNF website (DNF, 2005 & ICE, 2005).

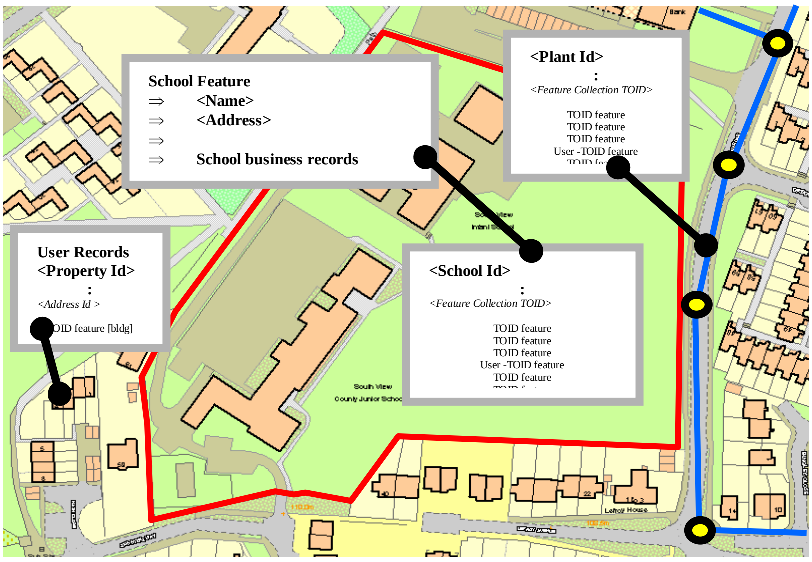


Figure 2. Examples of point, line and area geographic referencing

The case studies that now follow are a proof of concept data migration study (which is not yet complete) and an operational system based on data capture.

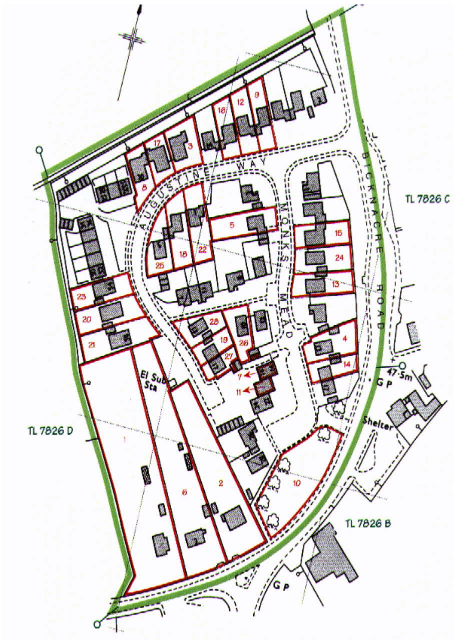
4. CASE STUDY 1: LAND REGISTRY INDEX MAP

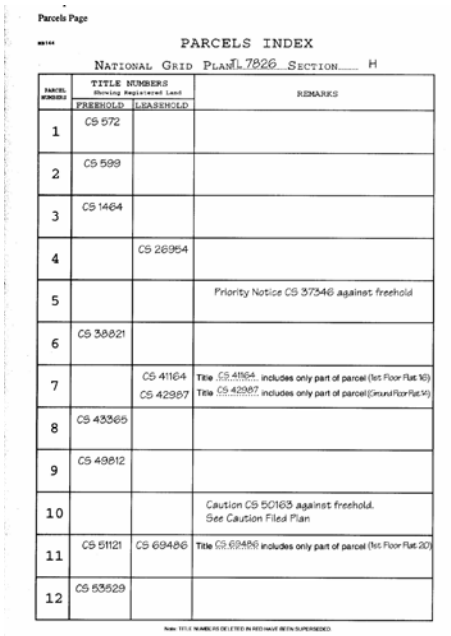
4.1 Index Map - Background

The Land Registry is responsible for the register of titles to land for England and Wales. It was established in 1862 and is a government agency with an annual turnover in the order of €580 million (2003-04) [Land Registry, 2004].

Guaranteed land registration forms the cornerstone of a stable, modern economy, providing everyone with the confidence they need when dealing with property. The Land Registry provides that guarantee and holds one of world's largest property databases, guaranteeing ownership of €1,885 billion worth of property in England and Wales. Land registration in the United Kingdom is based on the the general boundaries model. The Register for each property is formed of three parts: A) The Property register (i.e. the property), B) The Proprietorship register (i.e. ownership) and C) the Charges Register (a record of restrictions, easements affecting the propery etc). These text records are supported by the "title plan" of the property. This is either a vector plan or scanned copy of an old paper plan, using the Ordnance Survey map with the extent of the property marked in red.

Current registration coverage is about 50% of the land area of England and Wales, but this constitutes about 80% of the total number of land owners. It therefore follows that the remaining 50% of the land is owned by only 20% of the population. Land Registry has a ten-year plan to get all remaining land onto the Register by 2012 and will be taking proactive steps to meet this target. The Index map acts as an entry point to the database. Although an address or the property title number may be known and used, the Index map also indicates the spatial relationship of adjacent properties and cross references information as shown in Figure 3.




Parcels Page

PARCELS INDEX

NATIONAL GRID PLAN TL 7826 SECTION H

PARCEL NUMBERS	TITLE NUMBERS Showing Registered Land		REMARKS
	FREEHOLD	LEASEHOLD	
1	CS 572		
2	CS 599		
3	CS 1464		
4		CS 26954	
5			Priority Notice CS 37346 against freehold
6	CS 38821		
7		CS 41164 CS 42987	Title CS 41164 includes only part of parcel (1st Floor Flat 16) Title CS 42987 includes only part of parcel (Ground Floor Flat 14)
8	CS 43365		
9	CS 49812		
10			Caution CS 50163 against freehold. See Caution Filed Plan
11	CS 51121	CS 69486	Title CS 69486 includes only part of parcel (1st Floor Flat 20)
12	CS 53529		

Figure 3: An example of a paper Index Map, and an associated table of cross references of title numbers with the map. *Note: not all properties are registered.*

Between August 2001 and February 2004 Land Registry ran a project to convert all of its paper Index Maps into vector form. They captured 17.2 million titles and created 22 million polygons. This represents the majority of the urban and peri-urban land in England and Wales, but only a relatively small proportion of the rural land.

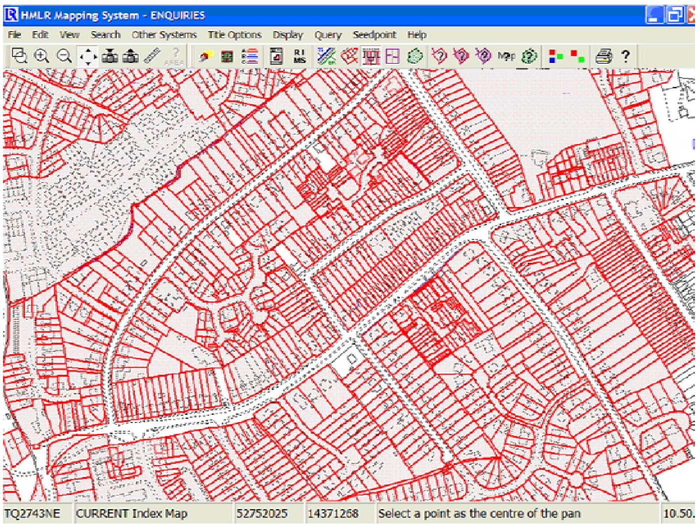


Figure 4: An extract of the Electronic Index Map [London area]

The rural properties have been more difficult to get onto the Register as they comprise mainly farms, manorial estates and moorland and these change hands infrequently; the main triggers for registration being activity due to a sale or creation of a charge (mortgage) on the property.

4.2 Electronic Index Map – DNF Proof of Concept Testbed

Even without the remaining rural land, the LR polygon dataset is a very valuable resource as it includes all major domestic and commercial properties. If its true potential is to be realised, its data structure must be such that it is compatible with other organisations' data. Different organisations, be they government departments, local authorities or private landowners, will realise considerable benefits and economies if all spatial and address data is in the same form. This will permit accurate comparisons and positive identification between data sets.

The Index Map was digistised using an advanced application of the overlay method and met the budget and timescales set for production. Research is now taking place to investigate the benefits of re-engineering the Index Map parcels as references to the topographic features. A first pass included the development of a set of agreed Business Rules and implementation in a Mapinfo system showed that this was possible. Approximately 100,000 parcels were then processed to develop a table of cross references for properties in the Bournemouth area.

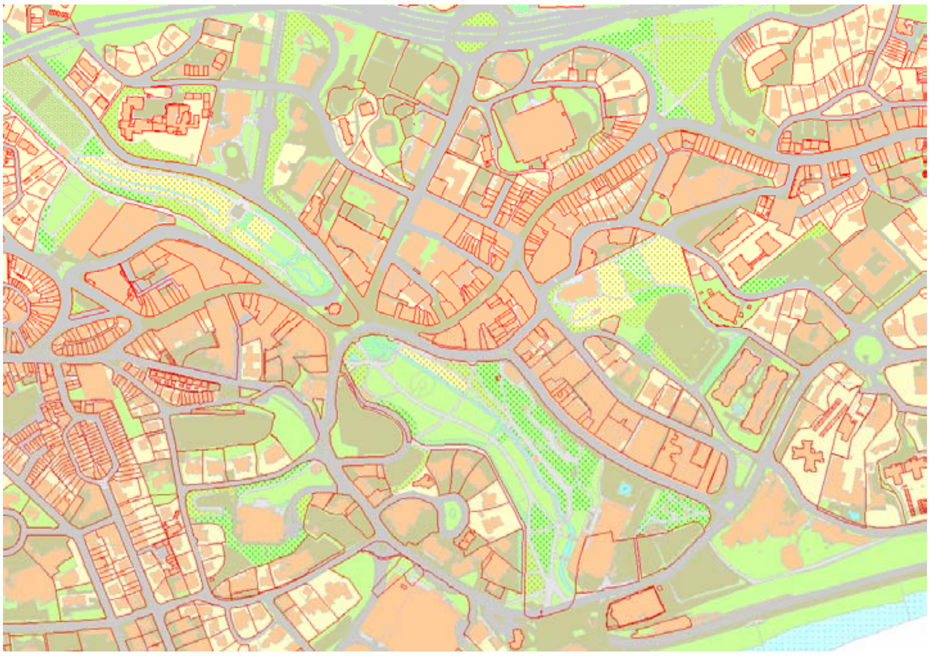


Figure 5: An illustration of the Index map referenced against geographic features. [Bournemouth Area]

A second phase is now under way to replicate the work using different software and skills to demonstrate the approach is transportable. Toward this end, Land Registry are working with Ordnance Survey and ESRI to develop software to further investigate the relationships between LR polygons and OS TOID geometry. The perceived benefits are:

- Improved data integrity (parcel to topographic feature) ie no overlaps or underlaps
- A managed maintenance regime based on feature level metadata with topographic object updates
- Potential to support future database developments
- Reduced risk to liability from errors
- Improved consistency with data from other organisations in data sharing.

These will be assessed when both phases of the research are complete. This work is still in its infancy as far as production is concerned but holds great promise for the future.

5 CASE STUDY 2: DATABASE OF OPEN COUNTRY AND REGISTERED COMMON LAND

5.1 Background to the access land programme

The Countryside and Rights of Way Act 2000, sometimes called ‘Right to Roam’ placed a duty on the Countryside Agency to prepare maps of all open country (mountain, moor, heath or down) and registered common land in England. Mapping is conducted in three stages, first a draft map is published and subject to public consultation. Once all comments have been assessed, a provisional form is issued and is open to appeal by those with a legal interest in the land. Once all appeals have been heard, the map is issued in conclusive form; a commencement order will start the new right of access. This new area of access amounts to nearly 1.3 million hectares of land or approximately 10% of England.

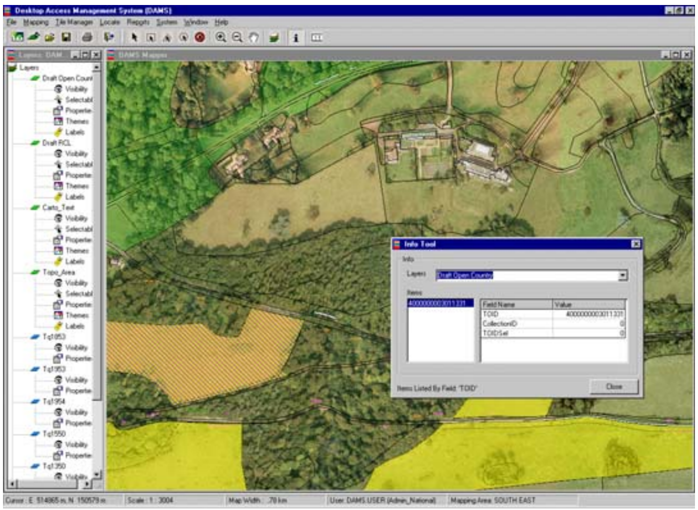


Figure 6: The programme required the collation of 350 reference datasets

All mapping and consultation work relating to this duty is being carried out by Black & Veatch via a four year rolling programme that divides England into eight mapping areas. This is a huge undertaking involving the use of nearly 350 reference datasets, a team of 80 full-time staff and three subcontractors. The solution places GI at the heart of the project, and in doing so has achieved a number of noteworthy milestones.

To date Black & Veatch have completed the issue of all draft and provisional maps and conclusive maps for four of the eight mapping areas. In addition they have identified approximately 10% of the total area of England as open country or registered common land. The last of the conclusive maps will be completed by Autumn 2005, with access likely to be implemented across the country on a region-by-region basis. A Commencement Order for the first two completed mapping areas was given on the 19th September 2004 by the Secretary of State.

5.2 Data management and georeferencing the land information

The project involved managing a complex process of analysis, mapping and public consultation, staggered across eight regions covering the whole of England. The solution required a geographic information system (GIS) at the heart of the project, providing benefits which include;

- comprehensive information on public access rights to the countryside
- engagement with stakeholders, including a statutory public consultation (e-enabled)
- digital audit trail of all decisions making the process traceable and transparent and
- resultant mapping that can be generated dynamically and viewed over the Internet.

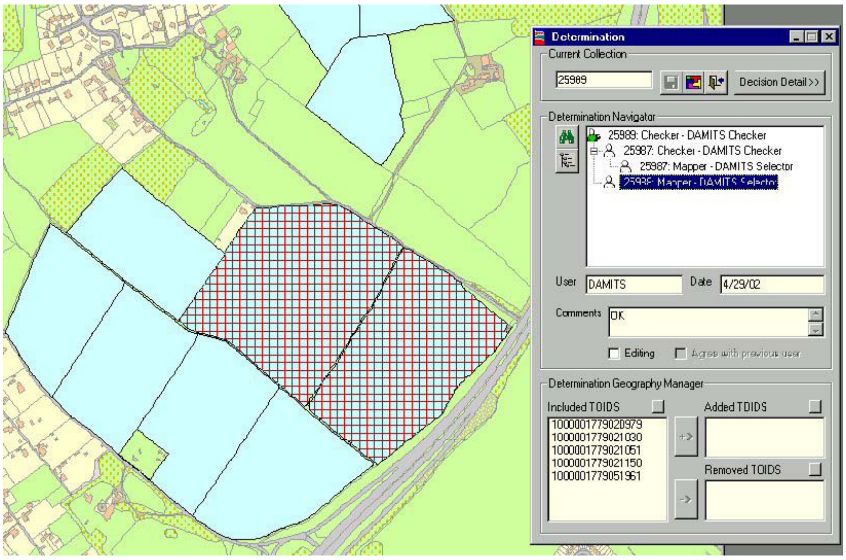


Figure 7: Example of access land referenced against physical land parcel objects

The approach to the mapping of open country was based on a desk-study exercise, analysing and interpreting existing datasets (including aerial photography for the whole of England) and exploiting the inherent 'intelligence' and accuracy of the Ordnance Survey of Great Britain's OS MasterMap data. The project team did not need to draw (digitise) new lines and boundaries but could select those areas (polygons) that made up a parcel of open country. The land parcel TOIDs were used to reference the access land and these parcels were aggregated and referenced under a collection identifier where the individual parcels were contiguous.

This approach turned the project from a data capture exercise to a data selection and categorisation exercise. This meant that the mappers were free to focus on the value-adding business of identifying open country through analysing and interpreting nearly 350 different datasets (amounting to almost 650Gb of both vector and raster data).

5.3 Public Consultation

Whilst the mapping itself was a huge technical task, the consultation stage was possibly more challenging as it needed to consider the aspirations of multiple stakeholders. For this reason, a mapping process that was transparent and consistent was agreed [Bush, 2004]. An integrated GIS to support this process from end-to-end keeping a detailed audit trail of all decisions made on every parcel of land was then developed. The system currently holds around 3.5 million records.

To facilitate the public consultation process, a website was developed that dynamically created maps from user-provided location information such as postcodes (www.ca-mapping.co.uk). The website allows the public to query the maps, submit comments on-line or print off their map and post it to us along with a comment form. The website was created in eight different languages and in compliance with the International Web Accessibility Initiative (WAI) guidelines from the World Wide Web Consortium (W3C). Special attention was given when developing the web site to ensure people that maybe colour-blind are not excluded.

5.4 Project statistics

By the end of the project, Black and Veatch will have issued over 100,000 individual paper maps. These have been deposited with 414 local authorities, 163 individual libraries, and almost 11,000 statutory consultees, including 10,000 parish councils. They organised 118 individual roadshow events, which attracted over 12,000 visitors to come and talk about the maps. The public consultation process resulted in almost 30,000 comments on the Draft Map and approximately one third of these responses have resulted in a change to the map. Since its launch in November 2001, the mapping website has received over 31 million hits from 531,000 visits, and has generated nearly 2.6 million unique user-defined maps.

The Mapping Access Database is based on Microsoft SQL Server and currently holds 3.5 million records. It holds records on the 1.8 million hectares that we have mapped and made decisions on, involving 89,000 parcels of land comprised of 1 million TOIDs. At periods of peak demand, the website has delivered up to 100 maps per minute.

6 BUSINESS BENEFIT THROUGH BETTER DATA INTEGRITY

Several other testbeds are under way and operational systems are also in place such as the Dudley Metropolitan Borough Council implementation of cross referencing addresses land parcels and the underlying physical real world features represented in the topographic data. The Institution of Civil Engineers involving several key utilities is leading a significant initiative just starting. The industry has a poor record on being able to share data owing to records being held in several different forms and epochs of mapping (ICE, 2005).

While DNF does not seek to set a standard, it is aiming to bring best practice to the surface and promote this through models and methods so that others can share, adapt and develop these to meet there own needs. Formal and industry standards will be incorporated where they are proven, but many of the DNF application areas often commence where ISO and similar standards finish.

The benefits, real and to be tested, outlined in the case studies offer real business advance over the methods of the past. There is no gain without pain however since the best practice does require more rigorous data management practices by all players, greater use of feature level metadata and versioning of datasets. Again we would not expect anything less in equivalent terms from the financial institutions that manage our bank accounts.

7 CONCLUSIONS

It is clear that distributed information systems will not only continue but these can be expected to expand both in number and volume as more data comes on line. Data-sharing needs will also grow, not just at the national level, but internationally.

The growing number of applications to improve existing services, reduce costs via automation or develop new information services will require better connected data and information and hence the need for high levels of data integrity.

Reuse of proven data based on cross-referencing will underpin the information and the knowledge economy of the future. However this does require vision, investment and adoption of best practice. There are frequent opportunities for all these as organisations modernise their information services at frequent intervals.

A small number of key government datasets connected up, especially in the field of land and property will have the potential to enable faster growth and significant benefits for all stakeholders, from government to the commercial sector and beyond to the individual.

REFERENCES

Bush, I et al. 2004. Public Participation – Drawing the Boundaries. CIWEM Annual Conference 2004.

Countryside Agency. 2005. http://www.countryside.gov.uk/WiderWelcome/open_access/index.asp [last accessed 2nd February 2005]

Digital National Framework [DNF]: 2005. http://www.dnf.org [under construction]

Institution of Civil Engineers [ICE]: 2005. Geospatial Engineering Board. Buried Services Working Group. Use of a common framework for positional referencing of buried assets. http://www.ice.org.uk [when published]

Ordnance Survey. 2004. The Digital National Framework - evolving a framework for interoperability across all kinds of information. A White Paper.

http://www.ordnancesurvey.co.uk/aboutus/reports/dnf_white_paper.pdf [last accessed 2nd February 2005]
Land Registry. 2004. Annual Report and Accounts. http://www.landreg.gov.uk [last accessed 2nd February 2005]
Rönsdorf, Carsten. 2005. Achieving and Maintaining Interoperability of Spatial Data. Proceedings of the FIG WW and GSDI-8 Conference.
Valuation Office Agency [VOA]. 2004. Reports of the Acacia Programme: http://www.voa.gov.uk/publications/acacia/final-report-june04/ [last accessed 2nd February 2005].

BIOGRAPHICAL NOTES

Keith Murray is Head of Geographic Information Strategy at the Ordnance Survey based in Southampton. He has worked as a surveyor and photogrammetrist. He has extensive experience in the use of imagery both in production and in research following a Masters Degree from University College London. More recently he has developed the concept of DNF, the geographic information strategy for the Ordnance Survey and is involved in developing the European Spatial Data Infrastructure. He has published numerous articles on the use of geographic information, new developments in GI strategy and research. He is a member of the Royal Institution of Chartered Surveyors (RICS) and the Remote Sensing and Photogrammetry Society (RSPSoc).

Bern Munday joined the Land Registry at their Stevenage district office in 1965, where he worked as a surveyor and plans technician. Whilst at Stevenage he acquired an extensive knowledge of Land Registration mapping practice which he took to Land Registry Head Office in London in 1991. In 1996 he joined Survey and Mapping Services at their Head Office, where he was heavily involved in the development of their computer mapping system. In 2001 he was appointed to the post of Project Manager and given the task of converting all of their paper Index Maps to vector form. This project was successfully concluded in February 2004 and he now occupies the post of Geographic Information Manager.

Ian Bush is the Technical Director and Business Development Lead for the Information Solutions Group at Black & Veatch. Since 1982 Ian has been using and developing computer technology for geospatial analysis related to hydro-environmental engineering applications with prominence given to Geographic and Management Information Systems (GIS/MIS), both in the UK and overseas. Specialising in the field of geomatics, Ian originally trained as a cartographer and engineering land and hydrographic surveyor, He has a Masters Degree in Construction IT and is a Fellow of the Institution of Civil Engineering Surveyors. Ian was the Technical Co-ordinator and leading architect responsible for developing the methodologies and IT/GIS infrastructure created to support and implement the Mapping of Access Land in England Project for the Countryside Agency. He is currently developing integrated project databases and virtual environments with particular emphasis given to collaborative working and Internet based mapping applications.

CONTACTS

Keith Murray
Ordnance Survey, Great Britain
Romsey Road,
SOUTHAMPTON
SO16 4GU
Tel: +44 23-8079-2729
Fax: +44 23-8079-2660
Email: Keith.Murray@ordnancesurvey.co.uk
http://ordnancesurvey.co.uk

Bern Munday
Land Registry
32 Lincolns Inn Fields
LONDON
WC2A 3PH
United Kingdom
Tel. +44 20-7166-4455
Fax + 44 20-7955-0110
Email: Bern.Munday@landregistry.gsi.gov.uk
http://www.landregistry.gov.uk

Ian Bush
Black & Veatch
Grosvenor House
69 London Road
Redhill
Surrey
RH1 1LQ
Tel: +44-1737 856268
Fax: +44-1737 772767
Email: bushi@bv.com
http://www.bvl.bv.com/isg