W06.3 Moving geography

Graham Clough, Development Manager, TerraQuest Group

Moving geography

The aim of this paper is to demonstrate the value of having spatial data which adjusts to match base mapping by use of a stable process. The probable initial state of data will be considered, as will obstacles to be overcome (such as Positional Accuracy Improvement (PAI) and other rule base change) and finally possible methods of easy maintenance.

Historically organisations with land information have maintained and used their own spatial data. Business processes have been set-up to allow for the collection, modification and supply of this data to internal users. This is changing. Pushed by government initiatives and enabled by emerging technology, the requirement of organisations which maintain land information is changing from internal supply only to supply to a broader customer base.

This change in supply has effects even within organisations. Using local government as an example, it is often the case that different departments have different quality standards and ways of using and interpreting spatial data. For example, someone in a land charges department may interpret a line which crosses into a road differently from someone in a planning department. When data is shared the rules for interpretation cannot be accurately shared as well.

What does this mean? From a quality standpoint it would be unusual to find two organisations with identical quality standards with respect to maintaining spatial data; indeed these quality standards are generally the minimum required to allow other business processes dependant upon spatial data to complete within acceptable quality levels. As and when organisations move to supply data to other organisations, the luxury of local quality standards becomes untenable.

So what replaces them? If the sharing of information is between only a few organisations, it is possible to agree quality standards and manage these. If you do not know who you are going to be supplying data to this becomes impossible; an absolute quality standard becomes very useful.

There is another way around this, which is the route being taken by hub providers such as NLIS. In this method a spatial query is supplied via the hub to a data provider (such as a local authority). The data provider then interprets the query and supplies the result back to the initiator via the hub. The responsibility for data quality becomes a responsibility for correct responses to queries under this system.

So why is data quality important at all? Primarily the main reason is the better the quality of data being queried, the less interpretation is required to provide an answer to a query (also the less margin for error in interpretation). This translates into faster turnaround times for queries and more queries being handled automatically which equates to less resource per query. Also IdeA is promoting data quality standards to enable hub providers to provide a more efficient service, so the requirements of e-government are pushing in this direction.

In order for spatial data to be usable for real world spatial searches (such as "Who owns this land?", "Are there any statutes in effect on this area of land?", "What's the address of this property?", "Why can't I build an extension here?") it needs to refer to a modelled view of reality – more commonly referred to as mapping. Good old Landline is one form of mapping, MasterMap is a more modern version.

The important thing to remember about mapping is it is always based on a rule base. What is a rule base? For mapping this will include things such as:

- Surveying techniques
- Precision of eventual mapping dataset
- Data storage specification
- Features to be mapped

In theory if two surveyors mapped the same area using the same rule base, the maps that they produced would be very similar (the only differences would be error in measurement). Rule bases are subject to change. This means that without change in the real world a map produced of an area may appear to be significantly different from a map produced using a different rule base. This was not noticeable on early paper maps because the data and the mapping were merged into a single entity – if the map needed redrawing the data would be adjusted to fit as and when appropriate (ironically, this is pretty much what we are aiming to do here with certain types of data).

The advent of GIS allowed for the divorce of data and mapping, it is now possible to change the mapping you are viewing without altering the position of your data. It is very rare to find any kind of provenance information linking user data to the mapping it was captured from. This is one disadvantage that GIS systems have that paper maps do not have!

Changes to the rule base used to generate mapping have been going on since mapping began. This is not a new issue. So why is this a big deal?

Here's why. As spatial data is shared between organisations (and departments) for the purposes of spatial searches it is inevitable that data captured to different versions of mapping will be compared. Two polygons representing the same real world features may produce different search results. The differences may be due to:

- Data overlapping data it should not
- Data not overlapping data it should

In the real world this can mean that a search may uncover information it shouldn't have or (more worryingly) not uncover information it should have. This cannot be tolerated for a system which is being used to make decisions which have financial effect, as it may increase the legal risk to the data supplier (as the data they supplied could be proven to be unfit for purpose).

A couple of recent events have caused this ongoing problem to become more immediate. These are PAI (which is essentially a large change to rule base (surveying techniques being part of the rule base used to capture the map) causing greater than average shift to mapping) and MasterMap (which holds our salvation?)

Positional Accuracy Improvement (PAI).
Originally this was viewed as being a solution to the general shift due to the pragmatic gluing together of County Series maps by the OS to produce 1:2500 scale National Grid map sheets. It is now clear that local positional shifts due to improved surveying techniques are also part of this. The upshot of PAI is that data captured to mapping which has undergone positional accuracy improvement will need snapping back onto the new mapping – this process is logically no different to the process needed for correcting any other shift due to a changing rule base. Technically the process is a little different in that the OS are the in the process

of providing some link files which, in conjunction with some kind of rubbersheeting software, can move data roughly back to where it should be. Once back to roughly where it should be a further process to snap data back to mapping can be carried out.

For general snapping back some kind of software is required which can snap data back to the OS within a given tolerance and alert the user when a situation is reached that cannot be automatically resolved.

The idea of snapping data to mapping is not untested. Two years ago TerraQuest undertook to snap data back to mapping based on the Northern Ireland grid. This was successful.

MasterMap

For the first time we now have a set of digital mapping which has identifiers and versions on the items which it is composed of. This new mapping product is called MasterMap. OK, so there's still a few issues about who should maintain version history etc, but overall it's a step in the right direction.

So, once the processes outlined above (rubbersheeting and snapping where PAI exists, and snapping in all other instances as required) are carried out spatial data and the mapping it is based on will once more be in synchronisation. The problem is then maintaining this. Once spatial data is spatially identical with the latest version of MasterMap, it will be possible to transfer the identifiers from the MasterMap data onto the spatial data. It is envisaged that three types of spatial relationship will be recorded:

- Spatial data wholly contains MasterMap feature
- Spatial data partially contains MasterMap feature
- Spatial data touches MasterMap feature

Using these three relationships it should be clear when a feature needs reviewing due to change of mapping.

Why do this? Spatial data with identifiers linking them to mapping have the potential to be automatically updated as and when the mapping is updated. If the business logic surrounding automatic updates is too complex then a user can be informed and can make a judgement call as to the degree of update required. This means that a process can be put in place which ensures that the spatial data is always cospatial with the mapping where this should be the case. Major systems suppliers are aware of the need for this type of process and the average time advertised for systems to be supplied is a year to eighteen months.

To conclude: The PAI program has highlighted the fact that some kind of data cleansing is required prior to the move into MasterMap – this would have been required regardless of whether the PAI program had started or not. The reason to move to MasterMap is that spatial datasets which update in a stable process controlled manner with changes in mapping are very valuable, as they remove a major data quality headache. In order to enable this to happen, data will need to be initially investigated to get it to the specification of the update process.