A practical approach to OS MasterMap

OS MasterMap is of keen interest to all users of Ordnance Survey large-scale mapping data. Early adopters' workshops, industry-level marketing communications and media coverage have all raised the profile of the Digital National Framework and its rebranding as OS MasterMap. Cadcorp's Chris Holcroft *examines issues and questions that potential users need to consider and offers practical advice on how to take advantage of this new data's many benefits.*

OS MasterMap brings new and different challenges to its users, especially when compared to the relatively straightforward practices of implementing products based on NTF, the National Transfer Format. The gains that Ordnance Survey promises with OS MasterMap cannot be achieved unless users confront and take a range of important decisions. But users also need to learn what OS MasterMap offers and how their own organisation can employ it.

Let's answer some basic questions before turning to implementation issues.

What is OS MasterMap?

First released in late 2001, OS MasterMap was developed by Ordnance Survey, Great Britain's national mapping agency. OS MasterMap encompasses new ways of managing and providing large-scale digital data to customers, and Ordnance Survey envisages that this new database will eventually replace its flagship Land-Line digital data product.

OS MasterMap introduces significant changes that affect the way large-scale base mapping data is provided, stored, used, updated and priced. In particular, OS MasterMap is provided in Geography Markup Language v2.0 format (GML2), an Open GIS Consortium standard for data distribution. For those unfamiliar with it:

- OS MasterMap is made up of uniquely identified real-world features created as either points, lines or polygons. Each feature has a unique topographic identifier, or TOID.
- OS MasterMap can be supplied topologically structured. Data comes in seamless, irregular shapes or 'chunks'.
- Related data can be selected for supply as themes based on real-world association: for example, roads, buildings and railways.
- Users will be able to get hold of data updates via the web.
- Data will be available compressed. Indications are that the compressed chunks delivered may be very large, at least in early transactions, and this will determine the recommended delivery media (web, CD or DVD).

Why OS MasterMap?

OS MasterMap has been developed in response to an increased demand for information referenced to location. The uptake in the use of GIS, the Internet and the increased awareness of the benefits of applied spatial information have also served to stimulate this development. OS MasterMap is expected to offer advantages both to its users and to Ordnance Survey.

First, OS MasterMap has a better structure for data than Land-Line. The justified gripes about unstructured Land-Line vectors with errors (slivers, undershoots, overshoots, orphan features and so on) should be a thing of the past. Second, better detail about features in the form of metadata (descriptive data) is there to answer those calls for richer descriptive detail referenced to map features (although sources indicate that errors are cropping up in the metadata delivered so far). Third, Ordnance Survey sees OS MasterMap as the database

OS MasterMap has a better structure for data than Land-Line. The justified gripes about unstructured Land-Line vectors with errors (slivers, undershoots, overshoots, orphan features and so on) should be a thing of the past. Second, better detail about features in the form of metadata (descriptive data) is there to answer those calls for richer descriptive detail referenced to map features (although sources indicate that errors are cropping up in the metadata delivered so far). Third, Ordnance Survey sees OS MasterMap as the database for many new map products and not just as a Land-Line replacement. Fourth, going down the web data delivery route helps fulfil OS objectives in e-government.

for many new map products and not just as a Land-Line replacement. Fourth, going down the web data delivery route helps fulfil OS objectives in e-government.

The increase in the number of spatial data providers has also prompted OS to re-evaluate its product line. The aim of OS MasterMap is to provide a seamless database of features across the UK at 1:1,250 and 1:2,500 scales. Ordnance Survey certainly claims that OS MasterMap brings many benefits – and there is no doubt that the new data is far more detailed than Land-Line, which it replaces. However, while the agency's points have validity, most benefits will come at a cost to the user in technical, resource and financial terms.

Why GML2 for OS MasterMap?

GML2 was not the automatic choice; in fact, the structure of OS MasterMap was apparently established before any delivery format was selected. Despite original OS hints that it would supply OS MasterMap in common GIS formats, this option was discarded once it was recognised that the database would be compromised if delivered to customers in this way. Why? Quite simply, any format that was to deliver OS MasterMap successfully would have to be able to handle continuous, topologically structured datasets (containing a potentially huge number of vertices), manage change-only updates to features within that data and entail users managing potentially massive datasets. No mainstream GIS format could offer this capability.

Some people also expressed the opinion that OS MasterMap data should be provided in a neutral format because OS should not be tied to any particular GIS vendor. While not perfect, GML represented an available OpenGIS standard that could do the job. It had relevance to the web. It also removed any competitive advantage afforded to individual GIS vendors and avoided the 'dumbing down' of the richly structured OS MasterMap data for distribution at the point of origin.

What is Open GIS?

The Open GIS Consortium (OGC) is a global initiative that creates open interfaces allowing transparent access to disparate geospatial data and its processing systems in a networked environment. The OGC meets every two months at venues throughout the world. With around 300 institutional members drawn from government, academia and the major GIS vendors, OGC has already created five standard interfaces for GIS, as well as the global geospatial data distribution standard GML2.

Implementation issues

Although many people have heard of OS MasterMap, most still want to know how it will affect their working practices. The main issues we'll discuss here, using Land-Line as a baseline for comparison, are supply, loading, data storage and change-only update. These four are not the only ones, however; others include the manipulation and editing of OS MasterMap data and the issue of licensing and pricing.

Supply

Let's pause to remember: Land-Line data in NTF format has been provided in tiles, either 500 metre or 1,000 metre squares named against the National Grid. The Land-Line naming convention has allowed users and software to locate any tile within the National Grid. Land-Line has been the core dataset of choice for government, utilities, data capture agencies and emergency services for many years, and the ways of working with it are ingrained.

In contrast, OS MasterMap data is not available in tiles. Initial supply of OS MasterMap is either as a topological dataset or as independent polygons in chunks. Choosing either option yields different results from

Figure 1: Tiled NTF Land-Line dataset.

those experienced with NTF. Additionally, each option produces different OS MasterMap datasets with distinct implications for the user.

Figure 1 shows how the traditional Land-Line approach (on the left) allows a regular grid structure of geocoded map tiles.

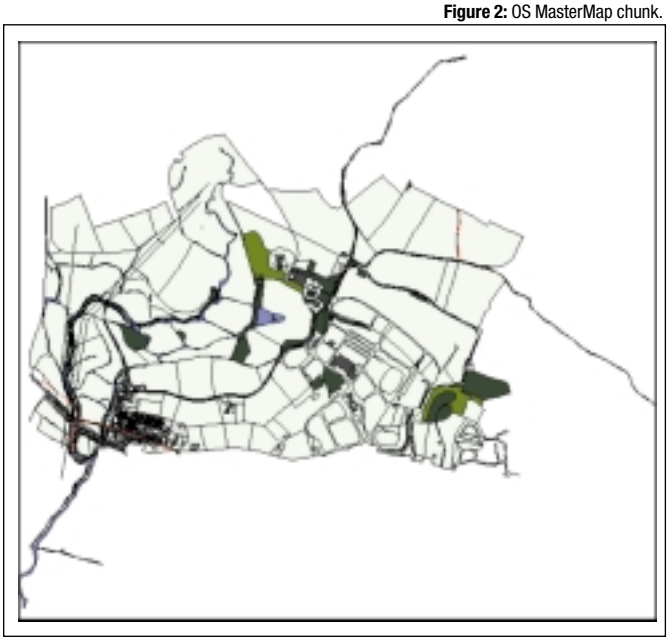
Conversely, the OS MasterMap chunk is an extraction of structured data with an irregular, user-requested, 'geographically' shaped edge (figure 2).

When we consider supply, several pertinent questions become apparent. For example:

Can your existing software that handles Land-Line data load OS MasterMap? If it can load it, how does it do it? As a direct read without translation, or by loading the data translated into a third-party data format?

How do you want your data – topologically structured (therefore interrelated and more attuned for complex analysis) or independent polygons that are better structured than Land-Line but lacking in the richness of the topological flavour?

Figure 2: OS MasterMap chunk.



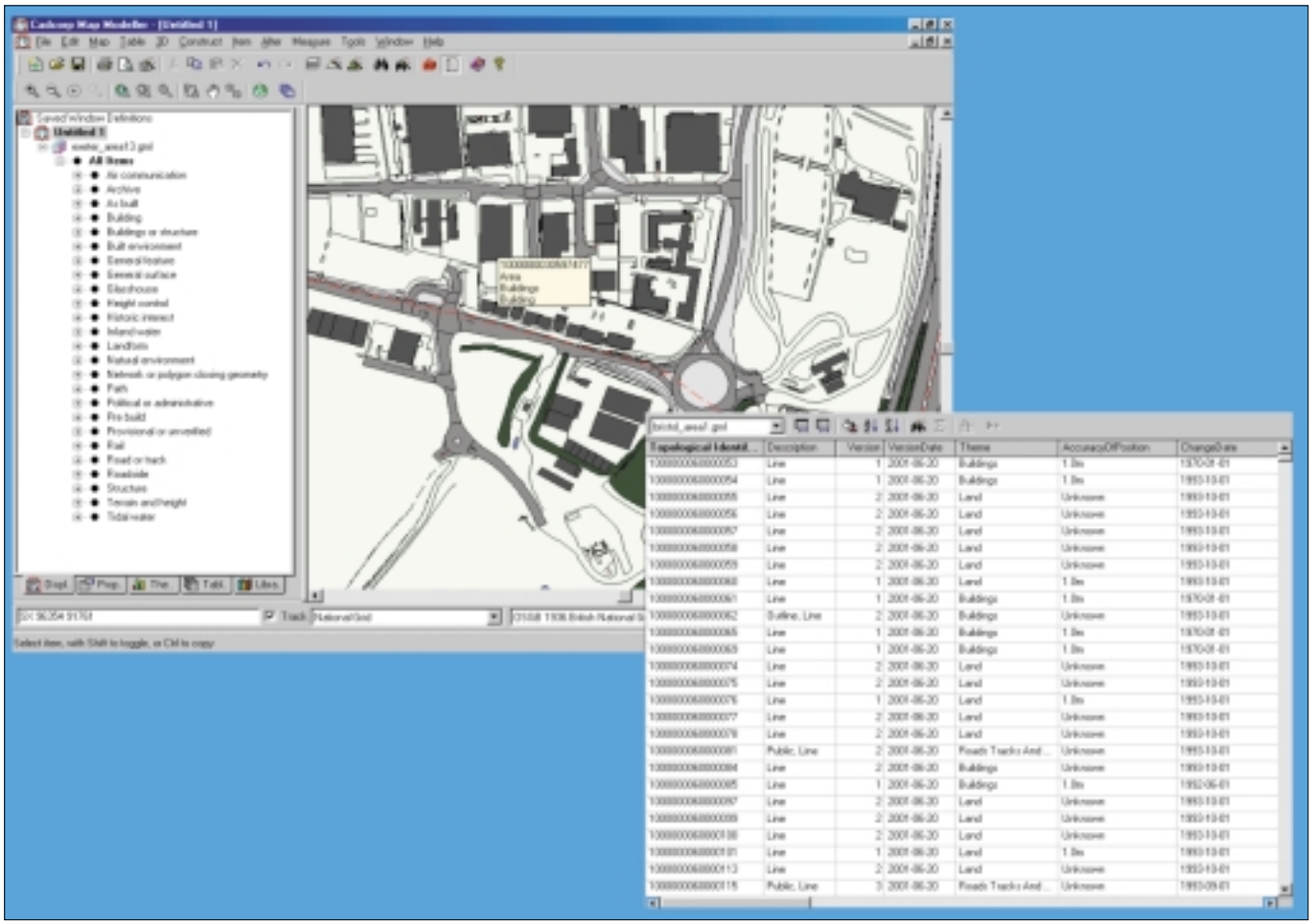


Figure 3: OS MasterMap data with full integrity loaded into GIS software.

Either way (and to be fair, this will become more apparent when we look at storage issues), how will the options you choose affect your RAM, storage and network capacities, which may already be creaking under the jobs you are doing now?

Which OS MasterMap themes – for example, Administrative Boundaries, Buildings, Land, and Roads, Tracks and Paths – are important to you? What are the storage and pricing issues related to your choice? Who manages these decisions?

It is clear that supply is no longer as simple as just ordering the predefined tiles that cover your area of interest.

Loading

In the market so far, both high-level and low-level approaches to getting OS MasterMap chunks into GIS software have emerged. The high-level approach is to read OS MasterMap GML2 native, thus ensuring the integrity of the original data (see figure 3). The low-level approach is to translate from GML2 to an established format such as MIF/MID or SHP. Because of inherent weaknesses in established formats when handling OS MasterMap data, this low-level approach is effectively a 'dumbing down' exercise and in practice appears to be slower to achieve.

In considering loading issues, it really boils down to how native GML2 support versus translation to another format affects you. If the more time-consuming 'dumbing down' of your OS MasterMap holding to, say, SHP or MIF/MID using the various translation products available is adequate for your needs, then fine. However, if you are used to quick and easy loading of NTF without translation, then native GML2 support is likely to be important to you, and you may be looking for it to be supported in your chosen GIS software.

Perhaps the greatest challenge posed by OS MasterMap is the management of data that is updated over time. We will see, in future, the demise of tile-based OS data in favour of selected chunks of data carved out for any area requested by the client. With such a sweeping change, it is naïve to think that developing the data-management processes will not actively involve the user and require the user to take serious responsibilities. Change-only updates must introduce new practices at the customer end that only good management structures can tackle.

The methods already outlined are only sufficient for managing modest quantities of OS MasterMap, because they gobble up memory resources when loading the data directly into GIS software. For larger organisations – those that must manage large territorial extents or need national coverage – what is a suitable method for managing medium to huge quantities of OS MasterMap data? The answer is loading to databases (figure 4). Here, OS

MasterMap can be placed in powerful data repositories. The resource overhead on the GIS software client can be kept down, because only the extent of the data being queried is taken into memory. Furthermore, database storage methods open up the capability to manage change-only updates of OS MasterMap.

If you go down the database storage route, you'll have to consider middleware. In particular, you'll have to find out whether you'll need to buy additional middleware to link your GIS software to the database of your choice, or whether this can be achieved by a direct link.

Storage

Unlike Land-Line (NTF), OS MasterMap is a verbose format. This means that a given area takes up more disk space than its NTF equivalent; tests indicate the increase in data volume is in the region of 10 times the size. The growth in data volume is only partly due to the move towards GML as a delivery standard. It is also due to an increase in the quality and detail of the topographic features and their associated metadata within OS MasterMap. In all likelihood, existing hardware based on NTF requirements will not be suitable for storing OS MasterMap data. With this increase in data volume, you will need to review your network requirements if the data is to be delivered in real time within a client-server environment as is possible with Land-Line. Even if used as a stand-alone, OS MasterMap will have significant hardware performance issues when compared with Land-Line.

For many organisations, storing OS MasterMap as a series of 'flat files' will not be an efficient way to distribute the data. The use of databases as a storage mechanism therefore becomes an important consideration. Using Editable BLOBs (Binary Large OBjects) will allow users to store topological information. Conversely, Open GIS SQL92 will only allow 'Simple Features' (points, lines, areas) to be stored. These methods suit themselves to topological and independent polygon OS MasterMap data respectively. Both approaches are theoretically possible in any database.

Figure 4: A database loading and change-only update management tool.

When we consider the storage of OS MasterMap, then, certain questions spring to mind:

- Are your storage resources sufficient?
- How about your network bandwidth?
- How will you store OS MasterMap – as flat files or in a database?
- If in a database, which is suitable or indeed available and affordable?
- Depending on whether you want topology in your database, do you opt for BLOBs or Open GIS SQL92?

Change-only update

Perhaps the greatest challenge posed by OS MasterMap is the management of data that is updated over time. We will see, in future, the demise of tile-based OS data in favour of selected chunks of data carved out for any area requested by the client. With such a sweeping change, it is naïve to think that developing the data-management processes will not actively involve the user and require the user to take serious responsibilities. Change-only updates must introduce new practices at the customer end that only good management structures can tackle.

Ordnance Survey has said that it is not responsible for dealing with the management of change-only updates. Its stated role is to produce OS MasterMap and then supply it to customers. OS has also said that managing change-only updates is the responsibility of the GIS software vendors – which we feel is only half the story. The role of Ordnance Survey is analogous to an electric power generation and distribution company's role in manufacturing. It's a very important role, but nonetheless a base component in the broader industrial process. We can argue that the management of change-only updates in OS MasterMap is really going to be a joint effort, where the vendor and customer collaborate to manage an exercise that will be very different from that experienced with Land-Line.

Change-only update solutions are possible and already exist based on the use of relational databases such as Oracle 9i.

Points to remember

- OS MasterMap is new and very different.
- OS MasterMap presents new costs and challenges.
- The user faces individual choices that are more complex than before.
- Users and GIS vendors need to cooperate to deploy OS MasterMap effectively.
- It is naïve to believe that managing OS MasterMap will be simple and low impact.
- OS can be congratulated for its many achievements, but there are still many challenges to be surmounted.

Note: Cadcorp has published a guide to managing OS MasterMap data for all Cadcorp SIS v6 users.

Chris Holcroft is the marketing manager of Cadcorp.
Tel: 01438 747996
Email: chrish@cadcorp.com
Web: www.cadcorp.com