Ordnance Survey's MasterMap

Jonathan W. Lowe

Lessons learned from the development of Great Britain's massive spatio-temporal dataset — the Ordnance Survey MasterMap — will benefit other nations that attempt to undertake similar projects in the future.

Close your eyes and imagine a magical island where virtually all permanent stationary objects — fences, paths, buildings, mailboxes, riversides — are registered in their government's spatial database. On this island, each of these real-world objects has a unique identification number that persists for the life of the object. The island's database delineates not only real-world objects, but also such invisible features as addresses or postal-code boundaries — with topologic relationships maintained between all features, visible or invisible. Sounds like a geographer's fantasy.

If you open your eyes in Great Britain, however, it's not a fantasy at all. Great Britain is home to an astonishingly fine-grained, 440-million-feature spatio-temporal dataset owned and operated by the Ordnance Survey (OS). This column offers a glimpse into the organizational and data-management challenges that all countries may someday face as their spatio-temporal repositories gradually approach the impressive breadth and depth of Great Britain's.

Defensive Mapping

Ordnance Survey is an unusual name. After all, what does the word "ordnance" (which means munitions or military supplies) have to do with mapmaking?

OS traces its inception to 1791, when the British Government realized that the south coast of England was in danger of invasion by the French. Consequently, the government directed its Board of Ordnance (the defense agency of that era) to survey the vulnerable coastline.

During a 20-year period and using a Ramsden Theodolite, OS staff mapped approximately one-third of England and Wales at one-inch scale. From 1801 onward, through nearly two centuries of war and peace, this initial coastline survey expanded to include all of Great Britain on a series of 230,000 paper maps ranging from 1:1,250 to 1:10,000 scale. Then, in 1973, anticipating the growth of computer mapping, OS began to digitize its enormous paper map collection until, by 1995, the entire collection was digital.

Because OS's original mission produced paper-based cartography for humans to interpret, the resulting digital product (called the Land-Line) was essentially a CAD (computer-aided design) dataset or digital map (see Figure 1). Human operators looking at Land-Line data on a computer screen or printed paper map could themselves interpret which lines represented buildings or roads or walls or waterways, but to the computer they were just lines and points scattered across an extent.

To convert their CAD data to a topologic, object-oriented data model, OS first weeded out more than 60 million geometric inconsistencies. OS then developed a complicated rule base and corresponding data model using Laser-Scan's Gothic architecture. Investing one year of automated processing (and minimal manual intervention) in reference to this rule base and model, OS reengineered its former CAD data into GIS data supporting automatic feature classification. The former Land-Line collection is now a fully attributed dataset known as OS MasterMap.

Breadth and Depth

Before exploring the challenges of managing an extremely detailed digital countrywide dataset, consider what sets MasterMap apart from most other spa-

Figure 1. Though rich in detail and meaningful to the human eye, OS Land-Line data lack attribution enabling automated feature classification.

Net Results columnist **Jonathan W. Lowe** covers the role of emerging technologies in the exchange of spatial information. Lowe is the owner of Local Knowledge Consulting (Berkeley, California), where he designs and implements spatial Web sites. He can be contacted at info@giswebsite.com.

Figure 2. OS boasts that MasterMap building footprints even include the bulges of bay windows. Surveyors capture linework with centimeter-accurate GPS-enabled instruments.

Figure 3. OS's ADDRESS-POINT dataset contains more than 26 million addresses, which are geocoded to their building structures and include a quality and accuracy rating.

tial datasets of its kind. Namely, MasterMap's data are fine-grained, comprehensive, and 100 percent complete across all of Great Britain. Comprehensive in this context means that OS captures most permanent, real-world objects in its surveys. For instance, MasterMap includes the following themes: administrative boundaries, buildings, heritage and antiquities, land, rail, roads, tracks and paths, structures, terrain and height, and water (see Figure 2).

These themes may sound as simple and familiar as any data provider's, but they contain considerable and comprehensive detail. The water theme, for example, includes the following physical water features: mean low and high water (springs); canals; lakes and lochs; ponds; bridges and footbridges; moats; reservoirs; rivers; streams; drains and ditches; foreshore features; floating objects; shake holes and swallow holes; sluices; stepping stones; taps; tidal gauges; waterfalls; public-water troughs; weirs; bollards, capstans, and mooring posts; breakwaters and groynes; culverts; perches, pilot beacons, and navigational beacons; pumps, wells, spouts, springs, and fountains; drinking fountains; swimming pools; watercress beds; issues; sinks; and springs. (Whew!) Each feature also has an associated text description and flow arrows — a symbol used to indicate the direction of flow of nontidal moving water. All other themes have similarly exquisite detail. In fact, carefully scrutinizing the OS model, one can't help but wonder who was foolhardy enough to embark on sucha project in the first place, pitted against such a complicated and diverse real world. Thankfully, fools rush in.

If you remain unimpressed, a review of the invisible data in the OS model may win your approval. MasterMap locates not only such physical objects as houses or roads, but also Great Britain's approximately 26 million postal addresses, all pregeocoded. The original address text comes from the Postcode Address File (PAF) of the Royal Mail, a government organization independent of OS, although the two organizations do cooperate. Any errors that OS discovers in the PAF when geocoding the text to a point on MasterMap are returned to the Royal Mail with feedback.

As for post office boxes, OS geocodes them to the delivery office from which they are delivered to the addressee. Saying that OS geocodes the PAF addresses could be misleading to those familiar with street-segment geocoding. The considerably more accurate address points in MasterMap fall within the building footprints of the structures housing their recipients (see Figure 3); they are not interpolated from a range of street numbers.

As an American journalist weaned on TIGER, Navteq, and TeleAtlas national-street-network databases, I must admit to being completely astounded by MasterMap's great detail. Not that the data themselves are unusual — certain individual U.S. municipalities with mature GIS operations boast similarly detailed features, but no national dataset matches OS's granularity across an entire country. If the U.S. does someday amass such a data resource, what challenges might we face in managing it? OS's technical experience provides a glimpse of that future.

Technical Choices

OS has achieved several enterprise-level data and computing goals, including a shift from flat-file to database storage, establishment of topologic data integrity, and adherence to open standards for data distribution. And, as a result of MasterMap's size and quality, a cottage industry of third-party vendors has emerged to assist with loading and updating OS data.

Seamless Database. The conversion of OS Land-Line to OS MasterMap resulted in a shift from tiled, flat-file data storage to seamless database storage. Seamless coverage is an improvement over tiles, which require software to manage computations or output extents involving multiple tiles.

For all the problems they solve, however, large seamless geographic databases also present their own new challenges. For instance, users no longer select data by referencing a tile. Instead, their unique area of interest clips all data it intersects. This ad-hoc nature of user queries over a large dataset threatens the database's performance. Unlike a tiled system, outputs cannot be anticipated or preloaded because each request is different. Consequently, developers need to engineer the database so that requests of any scale, including any set of features, and any anticipated number of simultaneous users do not significantly reduce performance. Though OS does not detail their internal database solutions online, Oracle databases appear to be popular with third-party vendors who help OS customers duplicate MasterMap in their own organizations, combining indexes, partitions, and data-clustering strategies to achieve acceptable performance levels.

Topology. OS describes MasterMap as "an unbroken web of 400 million features stretching from Lands' End to John O'Groats." In other words, it encompasses all of Great Britain. In more technical terms, MasterMap data are both spatially topological and logically object-oriented.

Spatially, OS delineates contiguous objects with the same single line, such as when a building wall abuts a road edge. Logically, OS's data model assigns every feature a membership with one or more object groups. For instance, the shared edge between the house and the road is a member of both the buildings group and the roads, tracks, and paths group (see Figure 4). Maintaining a topologic dataset sidesteps the pitfalls lurking in independent polygon models, such as unintended polygon overlap, slivers, gaps, and a similar data corruption arising during polygon digitization or editing. Enforcing object orientation enables automated feature analysis, such as identifying the total impermeable surface area for a given borough. Though requiring rigorous program logic to maintain, the resulting dataset provides users with a reliable and consistent base for analysis.

Open Distribution. Database storage makes spatial data easier to manage, but many users still rely on static files. And even if all users stored their spatial data in a database, they would inevitably use different database products; there are at least six spatial databases available in today's market. To enable equitable access of its digital data to all users, OS selected a data-exchange format that any spatial database could ingest or which could be converted into any flat-file format — the Open Geospatial Consortium's Geography Markup Language (GML).

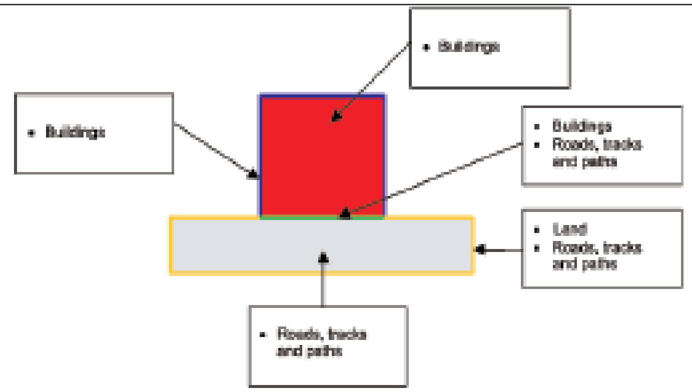

Figure 4. A single line (or edge) may define multiple real-world objects. Areas such as building footprints are implied by topologic relationships.

A spatially augmented form of extensible markup language, GML uses plain text and special brackets (indicating tags) in a hierarchical outline to store spatial data's coordinates, attributes, and metadata. GML earns its "open" distinction by conforming to a published standard and being transparent to all users — it's just ASCII text. MasterMap's complete GML output files fill eight DVD-CD ROMs, and their contents are not spatially indexed. Consequently, reading directly from GML to a map-rendering tool is inadvisable. Several vendors provide tools to translate GML into either a file-based format (such as MapInfo .TAB files) or a spatial database (such as Oracle Spatial).

Vendor Enhancements. By some reports, without the use of clever loading techniques, the first installation of a complete MasterMap dataset can take 40 days from arrival of GML files to a loaded and indexed spatial database. Fortunately, third-party vendors offer tools designed specifically to automate and speed the process. Snowflake Software, for instance, provides the Go Loader, a tool that organizes the complete GML translation into an Oracle database in less than five days. The Go Loader interface simplifies such steps as establishment of the logical and physical models, database schema creation and data partitioning, translation of GML, loading, indexing, and later maintenance (see Figure 5). Cadcorp also provides a MasterMap loader with a Microsoft-style wizard interface for bulk spatial database loads into Access, SQL Server, or Oracle. Cadcorp users can select the themes, attributes, and geometric types relevant to their application prior to initiating a load.

After obtaining the first GML snapshot of MasterMap's 440 million features, subscribers are spared the burden of reloading the entire dataset when updates become available. Instead, OS will deliver only the changes that have been obtained since the snapshot date — a strategy it calls "change-only update." Because each feature has a unique topographic identifier (TOID) that lasts for the life of the object it delineates, change reporting pivots on the TOIDs. While logically straightforward, the volume of changes is a significant data-management challenge, given that OS performs an average of 5,000 maintenance updates to MasterMap each day (applying approximately 2 million updates per year) and the same object may be involved in a series of discrete changes over time. Again, the same third-party loading tools described earlier provide update-management shortcuts, safeguards against duplication, and performance gains.

Semi-Public, Semi-Private

If OS MasterMap provides a glimpse into what's possible for national basemap data, what more can be learned from the organizational structure of OS itself? The director general and chief executive officer of OS is Vanessa Lawrence (formerly with Autodesk, Inc.). Lawrence reports to the British Parliament through the Office of the Deputy Prime Minister. Reporting to Lawrence is a board of directors who oversee six OS business groups. Within these groups are 350 surveyors and data-collection staff working from a nationwide network of field offices. OS contracts with U.K. firms for aerial photography. An additional 1,200 staff members at OS's Southhampton headquarters deliver printing, marketing, distribution, and support services.

At first glance, all these officers, surveyors, and staff might appear to be members of just another U.K. government agency. But, in fact, OS pays its operating costs through product sales, local and international services, and copyright licensing. (Though paper OS maps are popular in the United Kingdom, sales of digital products now fuel the majority of OS's business.) In April 1999, OS was awarded "trading fund" status, giving them more latitude over their finances, planning, and new-initiative development. (For instance, OS employs 30 full-time spatial innovation research staff members.) As OS explains on their Web site,

> Trading Funds are part of government, but have different finance arrangements from other centrally funded departments and agencies. A Trading Fund is an arms-length trading organisation but with a duty to observe specific financial targets set by the Treasury and involving capital returns, borrowing and transparency of reporting. It must also deliver quality standards and fitness for purpose in its products and services within government policy.

The government is OS's sole shareholder and the recipient of any OS dividends in profitable years. Given their estimated annual turnover of approximately £115 million (roughly $200 million), OS expects to deliver a dividend at the end of the 2004–2005 financial year.

Transferable Model? Would such an arrangement result in equally high-quality basemap data if implemented in other countries such as the United States? A recent Geographic Information & Technology Association forum debate suggests that some geospatial users have strong convictions to the contrary.

For example, some fear that if governments operate their national mapping programs by purely financial considerations, certain areas of the country might be neglected. With similar concerns in mind, the British government established a National Interest Mapping Services Agreement (NIMSA) in parallel with the OS Trading Fund arrangement. NIMSA is a not-for-profit contract that funds OS work not justified on purely commercial grounds, such as mapping of remote rural areas.

Some OS MasterMap users complain that the digital data are too expensive, arguing that the government should not be charging citizens for data that, if paid for by taxes and then freely available, would generate even greater economic activity and advance the public good. OS is well-aware of the debate, as evidenced by their corporate message pages on "the benefits of being a trading fund" and reminders of "independent research estimates at [OS] underpinning more than £100 billion of business in Britain."

Clearly, if data quality were their only goal, then OS's trading-fund status has been very successful. If the point of public investments is to create public value and that value is realized when the data are put to use, then there are unanswered questions as to how much more OS data might be used if OS were a fully public government agency rather than a trading fund. OS leaders respond that their data's quality would decline without the current amount of revenue from data sales, and they doubt that British taxpayers would approve the necessary expenditure with tax revenue alone.

At the root of the debate lurks the seemingly insoluble problem of measuring an economy's growth or a society's gain against the cost and quality of its national map data. Lacking such hard numbers, Britain has chosen a semi-private organizational model to create and maintain one of the finest basemap datasets in the world. May we all learn from this leader's experience as our own datasets expand and deepen. ⊕

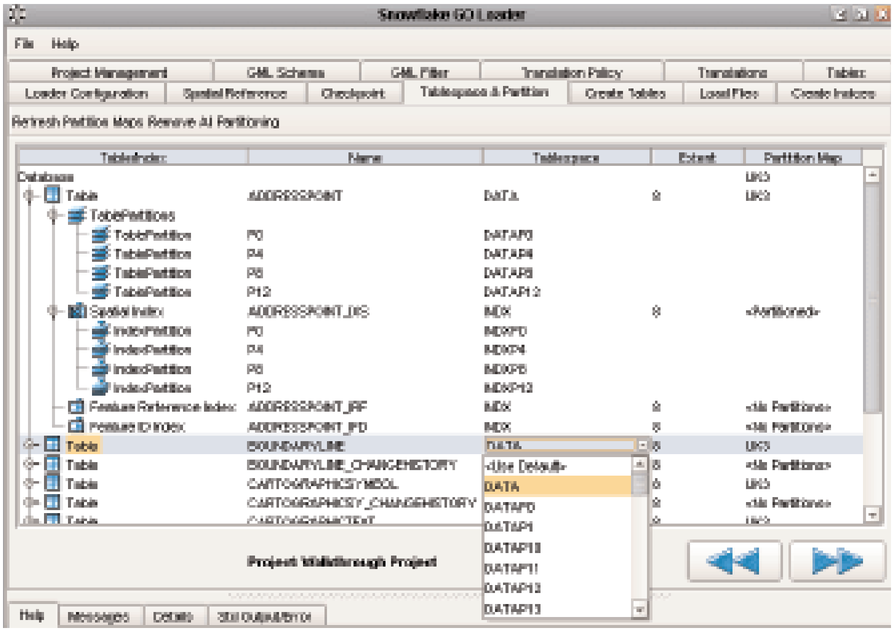

Figure 5. In a well-organized series of steps, Snowflake's Go Loader product assists users with the conversion of MasterMap GML files to an Oracle database. This screen illustrates the tablespace and partitioning step.