

# Storing OS MasterMap with Cadcorp SIS

*Cadcorp SIS – Spatial Information System – version 6 offers comprehensive support for OS MasterMap, and this support extends throughout the product range at no extra cost. Because other articles have covered the characteristics of OS MasterMap at length, this one explores more deeply a key aspect of managing the data: namely, the storage methods that users of OS MasterMap will need to take into consideration, whether they are using Cadcorp SIS or any other GIS software. Simon Doyle and Chris Holcroft explain.*

Loading and storing Ordnance Survey's MasterMap will be a sea change for most users who now work with OS Land-Line data. While smaller 'chunks' of the data can still be loaded into memory, it is wrong to think that this is the only route or even the best one. The importance of database storage will certainly come to the fore for *all* users looking to manage medium-to-large holdings of OS MasterMap. This will have implications for many users over and above the well-established practices for managing OS Land-Line.

## Storing OS MasterMap 'flat files'

Storing 'flat files' will be the simplest approach if you just wish to view your data in its native format. For Cadcorp SIS users, files can be read from the CD or DVD. They can also be copied to a directory on a PC or server in the same way as with OS Land-Line. Because Cadcorp SIS can read the compressed Gzip files, you do not need to uncompress the data. This will have considerable hard disk savings.

You are strongly advised to use high-performance SCSI hard disks if possible. Slower

speed networks such as 10BaseT may also struggle to cope with excessive data transfer with OS MasterMap in comparison to OS Land-Line. Your IT manager or systems administrator may be able to offer some advice here. Cadcorp recommends that you store any data on a dedicated server. Hardware specifications should favour RAM (1 Gb is not uncommon now) and also leave scope for expansion, for example empty slots for additional fast SCSI hard disks. RAID disk arrays also provide significant performance advantages when dealing with large amounts of data.

Before you take delivery of the data, it is worth finding out from Ordnance Survey the size of any initial supply files. Some sample statistics on compressed file sizes (taken from Ordnance Survey guidance notes) are shown in table 1.

Although Cadcorp SIS can read the initial supply and change-only data, it does not provide the data-management opportunities found within databases. The critical issue is the management of 'change-only update' files because

TOIDs (topographic identifiers, or the features) will be modified, created or will depart from the seamless data through time.

Houses may be built or extended, roads may be modified and so on. This means that all of the chunks in a data holding will need to be checked for such updates. *Performing this on flat files is not recommended.*

First, there is no guarantee that all the chunks in a data holding reside in the same directory or even in the same computer.

Second, there is no guarantee that such chunks came from the same order or consist of the same OS MasterMap themes. If the integrity and currency of the data cannot be guaranteed, the benefits of a change-only update mechanism are compromised, if not negated.

## Storing OS MasterMap in databases

For many organisations, storing OS MasterMap as a series of flat files will not be an appropriate means of storing or distributing the data. Several central government agencies have already taken delivery of national coverage of OS MasterMap. This dataset consists of approximately 415 million TOIDs or features. To manage this volume of data, these agencies have implemented database solutions. Many local authorities are looking to use relational databases such as Oracle to maintain and manage their base mapping in a similar manner.

Cadcorp has undertaken extensive research and development into the use of databases as a storage mechanism for OS MasterMap. To complement Cadcorp SIS's ability to read native GML (Geography Markup Language) files, Cadcorp has extended the software's capability to load initial supply data and manage any related change-only updates on that data holding. The Cadcorp SIS OS MasterMap Loader will allow an initial supply of OS MasterMap to be loaded into the following supported databases:

- Microsoft Access 2000
- Microsoft SQL Server 2000
- Oracle 9i.

Other databases exist, but these three are the most suitable for

**Table 1: Sample compressed file sizes for OS MasterMap**

Location	Area km <sup>2</sup>	Mb
Birmingham (urban)	267	940
Carmarthenshire (mainly rural)	2,454	400
Edinburgh (urban)	272	260
Horsham (suburban)	530	170
Rutland (mainly rural)	392	70



Figure 1: Cadcorp SIS OS MasterMap Loader Wizard.

OS MasterMap users because they can deal with the management and storage issues.

Microsoft Access 2000 does have a 2 Gb file limit, but this may still be a suitable medium for smaller data holdings. Microsoft Access 2000 is inexpensive, and many users are already familiar with it. Furthermore, it costs nothing to use this database format, so you need not have the Microsoft Access application installed on your PC.

Microsoft SQL Server 2000 and Oracle 9i are aimed at enterprise-wide operations. They normally require dedicated hardware and usually need a dedicated DBA or Database Administrator to monitor and manage the data. They are both used in a variety of enterprise-scale IT environments and are readily available to GIS users.

Getting spatial data into a database can be a complex procedure, and Cadcorp has

designed a software utility to make this as easy as possible to perform. The Cadcorp SIS OS MasterMap Loader is a cost-free application for Cadcorp SIS Map Modeller users. You are presented with a wizard-based interface in which you can create (and populate) Microsoft Access 2000, Microsoft SQL Server 2000 or Oracle 9i databases with OS

you filter your data. You can upload individual or multiple files and specify which *themes* to upload, and which *attributes* and *geometry* you wish to include in your database table. Once you have created the database, based on the above criteria, you can apply change-only updates later. Any update will follow the same filter pattern as the initial database –

---

***The critical issue is the management of ‘change-only update’ files because TOIDs (topographic identifiers, or the features) will be modified, created or will depart from the seamless data through time.***

---

MasterMap data. You can subsequently manage the change-only update process for any databases you create (with the Cadcorp SIS OS MasterMap Loader), with ease.

The Cadcorp SIS OS MasterMap Loader also lets

for example, the same themes, attribute columns and geometries will be uploaded. This is essential if the database integrity is to be maintained. During this process, the utility will check for TOID revision or duplication (or both) and

update the database table accordingly. Throughout the process, you are presented with progress dialogues, stating the number of features and files processed. You can save this information as a text file for future reference.

You may wish to use the utility regularly (weekly, monthly or less frequently), depending on when you acquire your change-only files from the Ordnance Survey online service. It is important to note, however, that the results of queries regarding the change date of a TOID are based on the Ordnance Survey database and not on the change in your own data store. For instance, a TOID may be modified on 1 January 2002, but it may not be uploaded into your database until some time later.

For example, you receive an initial supply on 21 February. You upload the data into a database on 21 March. You

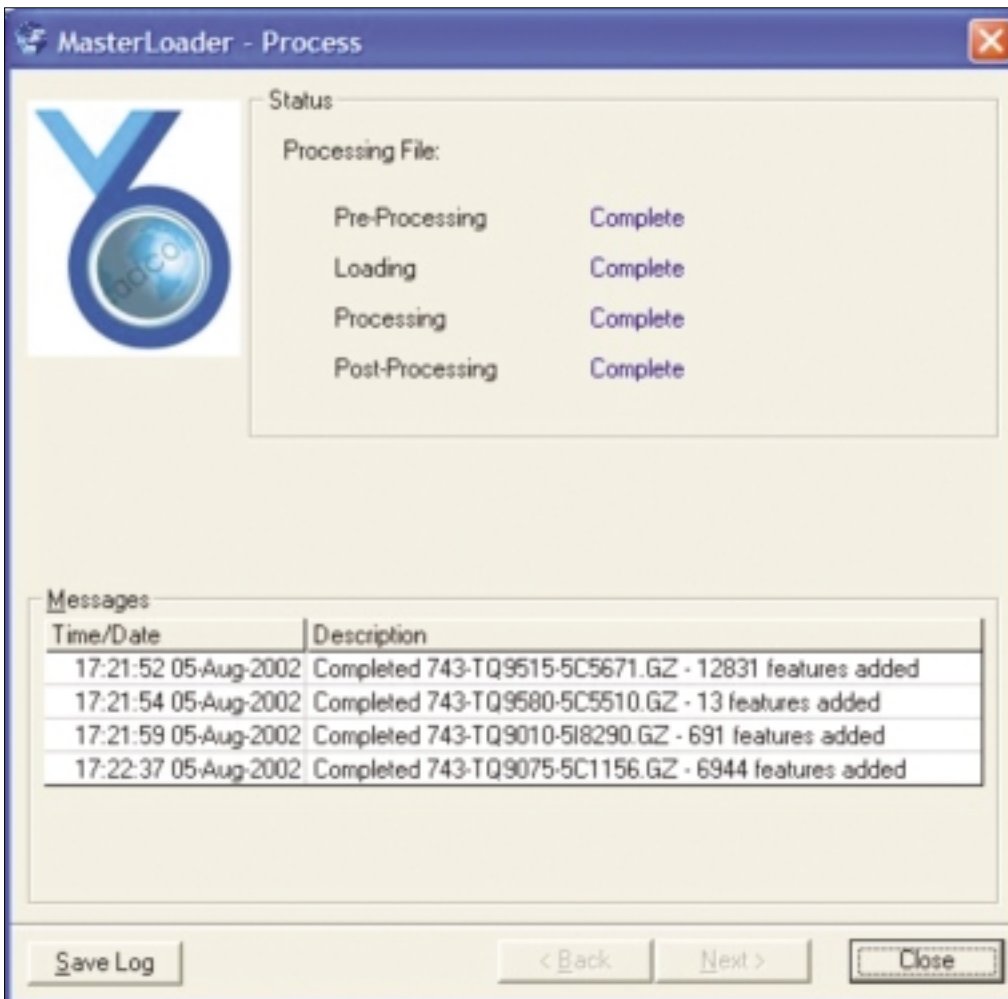


Figure 2: Cadcorp SIS OS MasterMap Loader 'Progress' dialogue.

then request change-only data on 21 April but do not upload it until 21 May. The database will be modified on 21 May but it will reflect the state of the Ordnance Survey database as of 21 April (the date it was requested). The database can only ever reflect the state of the Ordnance Survey MasterMap database on the date of the initial supply and the date(s) of any change-only updates. It is important to note this because, although Cadcorp SIS can perform historical queries for a given date and time, these

queries will only show the state of the local (user) database (relating to that date and time). This will often be exactly the desired effect, as you may need to examine the data you had in use on a certain date, but this will usually differ from the data Ordnance Survey held in its database for the same area on that day. *You are urged to consider the consequences and frequency of your acquisition of change-only data.*

Databases containing OS MasterMap data will be large. This is because they have to

store the OS MasterMap geometry within database tables. They will also store metadata and spatial and feature indices in order to make file access workable. To illustrate, we've chosen an area centred around Brighton and Hove (see table 2). The holding consists of 12 independent polygon chunks. Using the Cadcorp SIS OS MasterMap Loader, we have processed and uploaded the files into a Microsoft Access 2000 database and an Oracle 9i database. (These figures are based on our own preliminary tests and serve for illustration only.) All themes and all attributes have been uploaded.

The relationship of features to file size is generally in the range of 15,000–20,000 TOIDs per megabyte. This size may vary, but the densest OS MasterMap five-kilometre chunk has 863,144 features within it. Compressed, it is just under 36 Mb, which gives a ratio of around 24,000 features

per megabyte. As a rough guide, the ratio of file size to database size is in the region of 1 to 7 or 1 to 8. Uploading time for data of this magnitude takes hours. Larger datasets will take longer to process.

**In review**

OS MasterMap can be stored and read as compressed (Gzip) files, saving considerable disk space.

The Cadcorp SIS OS MasterMap Loader is a cost-free utility available to Cadcorp SIS Map Modeller users, which can handle the entire data uploading, processing and update of OS MasterMap chunks in relational databases. No cumbersome log files are needed to analyse the data or the upload process, as the database can be queried to provide all the metadata a user will realistically require (and use).

If you use OS MasterMap, you should consider purchasing dedicated hardware for your storage requirements. You should also review your current strategy and consider deploying copies of your derived databases to end users in their entirety if your networks are not capable of dealing with the new data volumes.

Uploading the data may take hours or days. But even if data uploading took a week, you would still have access to the databases for 51 weeks of the year.

You can hold data as files or within a database, as Cadcorp SIS can deal with both formats. We would recommend, however, that the database approach be adopted within any operational enterprise.

It is crucial that you assess and plan for your adoption and use of OS MasterMap data, especially with regard to data storage and deployment.

For more information, contact Cadcorp at sales@cadcorp.com.

Simon Doyle is a product engineer for Cadcorp and a member of the AGI Direct Action Committee. Chris Holcroft is Cadcorp's marketing manager.

**Table 2: Brighton and Hove – sample database sizes.**

Number of features (TOIDs)	1,053,569
Compressed OS MasterMap chunks	51 Mb
MS Access 2000 Database	418 Mb
Oracle 9i:	
- Data Table	376 Mb
- TOID Index	44 Mb
- Spatial Index	139 Mb
Total (not including other dB files)	559 Mb