

# Enabling Distributed GIS – OpenGIS in the Real World

Simon Doyle & Martin Daly, Cadcorp



Geographic Information Systems (GIS) are now commonplace within various commercial, governmental and academic settings. From their inception in the late 1960's and early 1970's, such systems have developed continually; migrating from the mainframe to the desktop and, more recently, to the Internet and the mobile device. The advent of component based GIS toolkits has also supported the broadening of the GIS audience and the increase in its uptake.

Within this process, users have traditionally adopted a proprietary or *homogeneous* approach, *i.e.* using a single GIS product and using that system's proprietary file format as a *de facto* standard. Many factors have driven this approach, such as implementation, maintenance management and communication issues. Many of which are further exasperated through the potentially complex or expensive adoption of multiple systems and data formats. These factors have all previously restricted GIS interoperability *i.e.* the ability for the systems to communicate effectively - regardless of vendor, data format or operating system. The benefits of overcoming these limiting factors are many, not least that of distributing data and systems in an 'open' manner. This increased state of interoperability is being promoted through OpenGIS Consortium (OGC) initiatives and standards. This paper discusses the trend towards interoperability and uses examples of OGC software implementations carried out by Cadcorp Ltd, a UK GIS software house and OGC Technical Member.

## Introduction

The evolution of Geographic Information Systems is well documented (Dangermond 1991, Longley *et al* 2000). They are now commonplace within many sectors but are still not considered to be 'mainstream' in the same way as other software applications *e.g.* spreadsheets or word processing tools. It cannot be denied however, that GIS is heading towards this end, if the history of GIS is any indicator *e.g.* the migration from the mainframe to the desktop personal computer (PC) to the browser and the portable device. This growth is represented fiscally too, with GIS revenues accounting for *circa* US\$1,047 million in the year 2000 (Daratech 2000,p3). There are now hundreds of software houses developing GIS applications based on their own or third party technologies. The advent of GIS toolkits or component based developer applications has, no doubt, helped fuel this trend. End users are also developing their own applications too, *via* macro or scripting languages and through the customisation of existing GIS. The upshot is that there are a large (and growing) number of systems and file formats in circulation.

There are, however, increased requirements for such systems to communicate with each other – both at system and data levels. In the United Kingdom, Central Government has established several 'e-Government' initiatives, such as the e-GIF (Government Interoperability Framework), in which Public Services are expected to interoperate and make better use of Information Technology (I.T.) within, across and beyond constituent organisations (Stationery Office 1999). This trend is not just restricted to the UK, it is evident in other economies too. It may also be seen in the private sector, where the recognised value of spatial information has increased through many channels. The rise in popularity and performance of the Internet Browser and associated services has also contributed to this growth. Such initiatives and demands are placing new pressures on existing GIS infrastructure. As such the importance of distributed or *heterogeneous* GIS will grow. The following paragraphs highlight the limitations of current practices and the potential of broader based GIS.

## Proprietary GIS (Homogeneous model)

The majority of GIS in operation today follow an homogeneous or proprietary approach. In the simplest form the software would be sourced from a single vendor and perhaps use a single operating system. Within this model, an organisation would also, typically, adopt an associated vendor file format as an internal standard – perhaps through necessity. Whilst there is nothing 'wrong' with adopting this approach, several limitations occur. Functionality is restricted to that of the incumbent system, data standards can be ruled by *de facto* standards which are difficult to share with other non-aligned software and dependency is placed upon specific service and software service providers. Many 'common' GIS have their own associated, proprietary formats which have served their parent systems for two decades or more *e.g.* SHP, MIF, DXF, all be it in a myopic way. These formats tend

to require ancillary software and human processing so that data can be shared across systems. This is costly and time consuming as well as being open to error.

Whilst many datasets are available in proprietary formats, such formats are not ratified by any central or organising body. The reality of translating, importing or exporting such data between different systems is more often than not a harsh one. To accompany institutional limitations, there are also technical limitations attached to *de facto* data standards. For instance, some may have to store geometries, such as points, lines and polygons separately. Some may not be capable of storing or manipulating topology. Some may only be capable of storing two-dimensional data, and so on. Users are also dependent upon the authoring vendor to not alter (or drop) the format specification. There are advantages to this model with regard to dealing with a single supplier and system management and maintenance. Users need only learn about one product (range) and data can be moved between related systems without much effort. However, once an organisation is required to interact with another and between non-aligned systems the rigidity of this approach is exposed as a weakness.

### Distributed GIS (Heterogeneous model)

In an attempt to alleviate the restrictions imposed in the proprietary model (above), tools and technologies are being developed to encourage and enable distributed or heterogeneous GIS. The term heterogeneous refers to a computing environment in which a variety of software and hardware co-exist and interact. In this environment, users are not restricted to specific vendor systems or formats, nor are they necessarily aware of the diversity of the system from an end-use perspective. This, is interoperable GIS.

The Internet (World wide Web) is often cited as an heterogeneous environment due to the diverse range of hardware, software and data which construct its form. Within such an environment there are common metrics and protocols which bind these disparate entities together. In the case of the Internet, HTML (Hyper-text Mark-up Language) and TCP (Transmission Control Protocol) are the common metrics and are borne through Internet Browser applications. Within the GIS community the benefits of interoperability are growing in acceptance. The major benefit being the ability to distribute and then combine disparate data to produce new combinatorial datasets, regardless of the original format. Therefore promoting information exchange across organisations. An example of such a model is presented.

### Example

A data supplier hosts a variety of datasets across numerous servers. These databases are accessed directly by client applications across a network or through Internet Browsers. The clients communicate with the data servers through a common, *de jure* metric which allows a uniform supply of data to each client – regardless of their origin but dependent on the implementation of the same metric.

For this to work in practice, the emphasis is on the interface between systems. Since 1999 the OpenGIS Consortium (OGC) has been researching the feasibility of such heterogeneous systems. The aim of the OGC Web Mapping Testbed Phase 1 (WMT1) was to "...increase users' ability to access and overlay maps and earth images available from other vendor's Web map servers." ([www.opengis.org](http://www.opengis.org)). Within this and the subsequent OGC Testbeds, many data providers and curators have successfully combined disparate data and systems via the Internet.

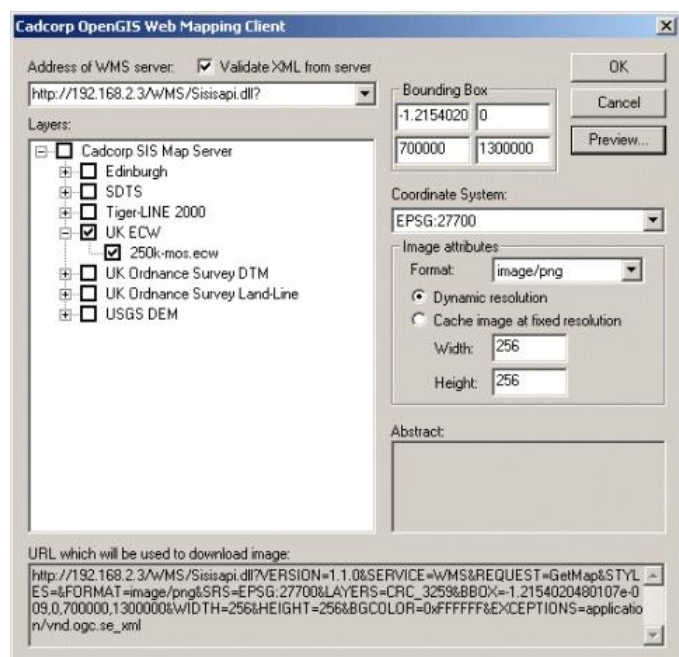


Figure 1a. Cadcorp SIS – Spatial Information System™ user interface to connect to an OpenGIS Web Map Server as a client

To enable the practices described above, several standards have been designed and implemented. The OGC have several ratified interfaces, which 230+ organisations have been party to. These bodies include large GIS vendors and users who recognise the importance of interoperability. These interfaces standardise the semantics and the functionality delivered by any implementing software. Furthermore, software can be submitted for conformance testing by OGC. They also provide an open and stable framework for future application and service delivery. An important branch of this standardisation process has been the distribution of vector (and raster) data over the Internet. This has produced the Geography Mark-up Language (GML) which is used to encode real world features in a modified version of the W3C eXtensible Mark-up Language (XML). GML originated from OGC's WMT1 project.

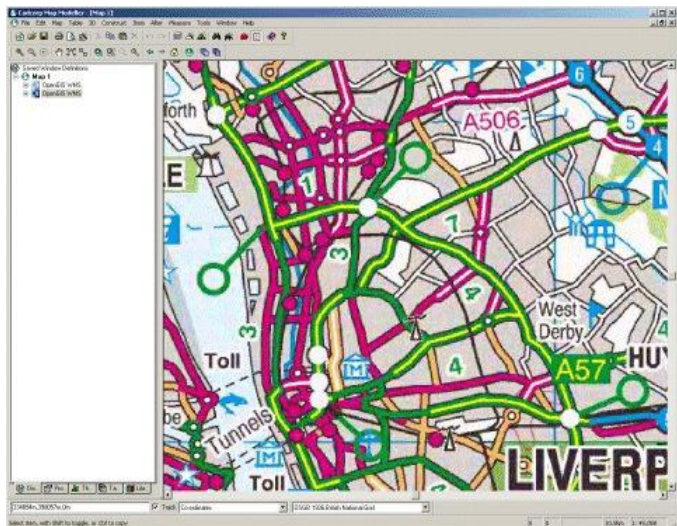


Figure 1b. Cadcorp SIS – Spatial Information System™ displaying an ECW file from a remote OpenGIS Web Map Server, as selected in Figure 1a.

Systems interpret this neutral information through OGC specified interfaces, ensuring standard delivery between systems and organisations. Like HTML and XML before it, GML will no doubt be used for many, as yet, unforeseen purposes.

### The rise of distributed GIS

Technological advances facilitated by OGC standards have permeated through Geographic Information (GI) provision, allowing GIS users to take advantage of new, distributed mechanisms. The amount and type of data now available to end users further aids this process. At a global level many datasets exist, varying in price from free to several thousands of pounds. Their content and quality is also varied. At local levels, where users have access to specific data stores, data use and recognition is also high – this again, has been driven because of the growth in GIS and partly because of the activities of industry bodies such as AGI (in the UK) URISA, and GITA (USA) and so on. In short, there is more data on offer than ever before and the increase in quantity and accessibility of such data places more emphasis on communication between systems, if the potential of this data is to be realised. Users are also keen to examine the data which is held by other organisations, partly through legislation and partly through the 'academic' legacy associated with many GIS. Interoperability is an aid to this end. GML is currently allowing real systems to access real data stores to solve real problems. It is providing the means for tangible results to be created from what are often dry and theoretical bases.

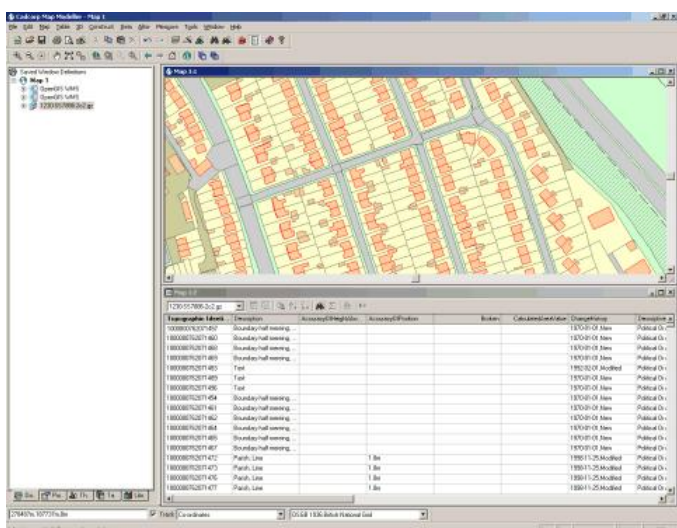


Figure 2. Ordnance Survey MasterMap (GML) viewed in Cadcorp SIS – Spatial Information System™.

### Implications for developers

There are many software houses adopting the tenets of 'open' GIS i.e. adopting and implementing OGC specifications GML2 is probably the most visible sign of this adoption, allowing data to be passed between organisations, systems and processes. GML3 will eventually extend to multi-dimension and complex geometries, further strengthening the case for its use – broadening the scope for data storage and transfer. Whilst not all vendors will adopt OGC specifications or commit to re-engineer their GIS, those who do will be able to serve organisations who wish, and need to share geospatial data, far better. Cadcorp SIS – Spatial Information System™, is conformant in several

OGC specification areas and whilst it is a desktop GIS, it allows users to use it as a client to any OGC Web Map or Web Feature Server. Thereby the user can access raster and vector data, regardless of its original matter – providing that the host server implements the appropriate OGC interface specifications. The restriction here is merely in the number of available OGC servers and the speed at which it can be transferred across a network. Those developers who ignore the importance of GML and interoperable systems do so at their peril.

### **Implications for data suppliers**

There is an obvious requirement for spatial data to be delivered effectively and quickly – the temporal currency of data is an increasingly important aspect of its value. If users can access such data without recourse to explicit data transfer i.e. directly into their GIS – regardless of what that system may be – then the data will penetrate more markets, or the same markets, deeper than any ‘restrictive’ de facto standard. That is not to say that once delivered, GML cannot be converted to another format to satisfy legacy or other systems. With GML for instance, data suppliers can provide data as files or they can provide direct, web based, data leasing. GML also potentially offers a neutral alternative to Binary Large Objects (BLOB) database storage.

Ordnance Survey, the National Mapping Agency for Great Britain, has adopted GML as a basis for its large scale, vector, OS MasterMap™ base map product. It is an extremely detailed national, topographic coverage based on a scale of 1:1250. By doing so, it has changed the face of geospatial data provision in Great Britain and set a template for other major data providers i.e. using GML for data supply and maintenance. By doing so it has increased the pressures on GIS vendors outlined above. It has also allowed OGC specifications to be realised on a very large scale, in an important and operational context. US Census Bureau’s TIGER/GML may be the next whole scale conduit for the realisation of current OGC specifications. In these cases, data providers are forcing the market, although there is a circular process in action; the more GML data provided, the more interoperable GIS will become, the more interoperable GIS there are, the greater the amount of GML data there will be. Web users who may be familiar with ‘show me the nearest’ type applications will be taken on a journey where functionality will increase to a point where they will be ‘doing’ quite complex GIS without realising.

### **Implications for users**

Users will no doubt be as receptive to the benefits of interoperability once these benefits have been fully exposed. In certain areas this is just reaching a critical mass. Initiatives such as the OGC USL and the Ordnance Survey Digital National Framework are at the fore and are lighting the path for increased user uptake of interoperable practices. It is unlikely that rigid homogeneous systems and processes will be able to fully exploit these benefits per se but there will be ‘some’ need for them to interface with the new generation of GIS which have interoperability in their blueprint. The concept of remote data storage, which is accessible through common tools is one which underpins the World Wide Web. This analogy is being applied to the GIS sector but it will be further advanced through increased user uptake.

### **Conclusion**

The distributed model forms the basis of interoperability initiatives espoused by OGC. It is an attempt to move away from the GIS of the mid-late twentieth century and to design systems and services from the ground up; such systems harness open and ‘fit for purpose’ technologies and do not try to fit current operational and enterprise practices into technology which is at its limit. GML is a useful and now, proven delivery mechanism for geospatial data. It is not a direct replacement for traditional formats, but it is a primer for a more ‘open’ and perhaps, ‘mainstream’. Geographic information community.

### **References**

- Dangermond, J. (1991) *The Commercial Setting of GIS* in Maguire, D.J., Goodchild, M. F. & Rhind, D.W. *Geographical Information Systems: Principles and Applications*. Longman.
- Daratech (2000), *GIS Markets and Opportunities: 2000 Forecasts and Opportunities*, Daratech, MA.
- Stationery Office (1999) *Modernising Government (CM4310)* Cabinet Office, Great Britain, ISBN 0101431023.

**Biographies**

Simon Doyle is a GIS Technical Specialist at Cadcorp Ltd. He was previously the GIS Manager at the London Borough of Brent and a researcher at the Centre for Advanced Spatial Analysis, University College London. He is currently working on GIS product development which have OGC specifications at the core. He holds a Master's degree in Geographic and Geodetic Information Systems (UCL).

Martin Daly is the Technical Director of Cadcorp Ltd. Martin has led Cadcorp's development team since 2000, prior to which he was Cadcorp's senior programmer. He is a leading figure in the OGC Coordinate Transformation Services and Simple Features working groups. Martin is a graduate of the University of Glasgow.

**Web resources**

<http://www.cadcorp.com>

<http://www.opengis.org>

[www.ordnancesurvey.co.uk](http://www.ordnancesurvey.co.uk)

<http://www.govtalk.co.uk>