Interoperability and data handling

The last five years have seen a marked increase in the amount of time and energy spent by GIS software vendors on the subject of "interoperability". This article explores the reasons for this interest and considers the various meanings that the term has in the context of GIS systems.

According to ISO/IEC 2382-01, Information Technology Vocabulary, Fundamental Terms, interoperability is defined - rather dryly - as "The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units".

In a software engineering context, and more particularly for GIS software, interoperability is used to describe the capability of different programs to exchange data via the ability to read and write a common set of exchange formats, and to use the same protocols for data exchange and interaction.

Mainstream IT has used a standards-based approach to resolve interoperability issues for a long time. We see the manifestation of these standards day-to-day in areas such as mobile phone technology, e-mail and the web, and we have grown familiar with standards such as HTTP, HTML, GSM and Bluetooth.

We encounter interoperability issues every day. All the standards mentioned above come into play when, for example, using a mobile phone to provide an internet connection for a laptop.

Most people who work in the GIS arena will be aware that the industry has perhaps been slower to adopt standards than mainstream IT. However, end user demand driven by the requirement to integrate GIS with related technologies such as GPS, relational databases and the internet, has forced GIS to come out of the pseudo-academic niche it occupied in the 1990's and into the broader IT arena. GI systems and vendors have been forced to address interoperability issues as a result.

For many years the major hurdle in making GI systems interoperable was the lack of geospatial standards among different software systems.

This hurdle has been overcome through the activities of organizations such as the International Organization for Standardization (ISO), and more particularly in the GIS arena, of the Open Geospatial Consortium (OGC).

When it was formed in 1994 the goal of the OGC was to define interface standards to allow seamless, vendor independent access and interchange of GIS data. This interface definition approach was the single major differentiator between the OGC and other attempts to introduce standards into GIS, most of which had been focused on data translation between systems. Today, most GI vendors are members of the OGC at some level, together with major IT companies that have some spatial products, such as Microsoft, Oracle and Google.

The interface standards defined by the OGC can be viewed as the building blocks of GIS. Some of the more well known standards are:

Simple Features (SF) – covering vendor-independent data storage, this standard specifies the digital storage mechanism for geographical data e.g. points, linestrings and polygons. Two well-known examples of this are the OGC-specified GML encoding and Oracle's SDO_GEOMETRY data type. The SF standard also covers the definition of nine spatial tests such as 'Intersect' or 'Contain'. The standard ensures that consistent results will be derived from all implementing systems.

Web Map, Web Feature and Web Coverage Service standards (WMS, WFS and WCS) – covering data distribution over the internet, these

Figure 1. Example of user interface showing application of OGC spatial tests.

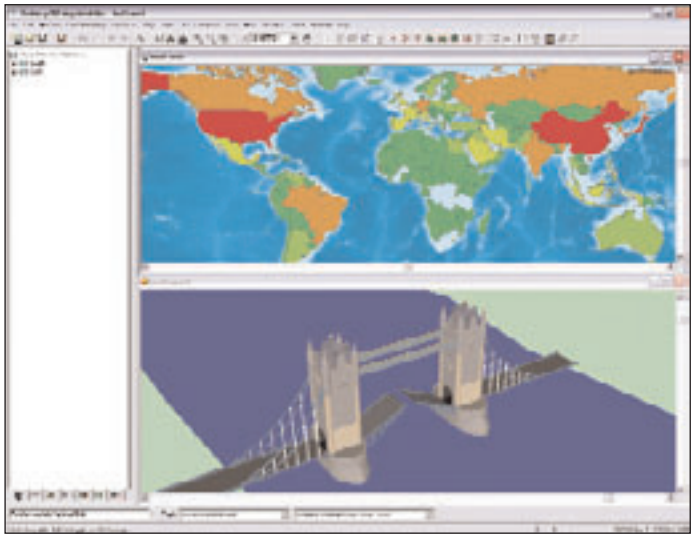

Figure 2. KML – the latest OGC standard.

standards provide opportunities for achieving high levels of interoperability between systems. They are the most commonly implemented standards in the web mapping arena and provide a method of viewing and downloading data in a standard form.

Catalogue Services specification (CS/W) – which deals with management and location of online data stores of GIS information.

Geography Markup language (GML) - An XML encoding for the transport and storage of geographic data, including both the geometry and properties of geographic features. GML has achieved widespread recognition and endorsement within the UK because it has become the data distribution choice of the Ordnance Survey (OS) MasterMap® product set. The advent of GML has allowed the OS to distribute their geographic data in a single, vendor-independent standardised format accessible to all vendors. National mapping agencies around the world are looking to follow suit in the future with GML-based supply mechanisms.

The latest standard is KML, formerly Keyhole Markup Language. Originally developed for use with Google Earth, KML is an XML-based encoding for expressing geographic annotation and visualization on two-dimensional maps and three-dimensional Earth browsers. KML 2.2 became an official OGC standard in April 2008.

Of course interoperability means different things to different people in the GI business.

To some it means translation of data from one system to another. In this activity certain proprietary data formats have taken on the role of de facto standards - for instance DXF, DWG, MIF/MID and SHP. Although valuable for graphic data interchange, these formats have limitations when it comes to transferring underlying intelligence between systems. Differences between data structures can lead to unsatisfactory results and data being lost. Additionally data is transferred at a point in time

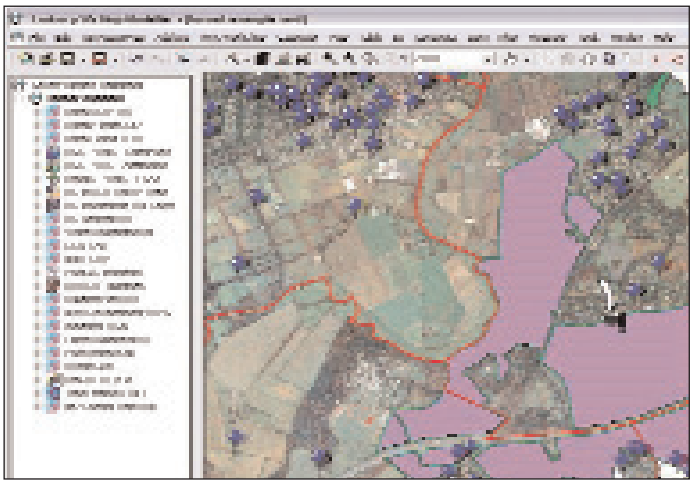

Figure 3. Interoperability example showing integration of data from multiple sources onto the GI desktop – raster data, Oracle data, SHP, DXF, DWG, TAB, MIF, DGN (and others) displayed with live links to data source.

providing a snapshot rather than any kind of live data update.

To other people, interoperability means integration of data from multiple sources onto the GI desktop. Several systems on the market today permit the direct read of data from system disk onto the target platform. The link is often live, thus changes in the source data are visible in the target at the redraw stage. While this technique can have some of the drawbacks of translation (for instance interpreting styling), the live nature of the link means that the user of the target system can maintain an up to date view of data even though it may be sourced from multiple systems.

To other organizations, interoperability means managing the interface between CAD and GIS data. At this level, interoperability issues are often clouded by establishing where the boundary between CAD and GIS data really lies.

Perhaps the major difference appears when coordinate reference systems have to be used to correctly calculate the position or location of particular CAD items, or where CAD data needs to be geo-referenced, e.g. to be overlaid with smaller scale GIS data. It can be argued that engineering plans are large scale maps contributing to an overall picture of a larger built environment. Recent moves within the OGC to incorporate Building Information Modelling (BIM) standards show a significant move in this direction.

As data volumes inevitably increase over time 'spatial data warehouses' have become increasingly important, as has interoperability between spatially-enabled databases. Usually based on industrial strength relational database technology such as Oracle, Microsoft SQL Server, and PostgreSQL, these databases offer the prospect of a single corporate data store for all GIS and CAD data managed within an organisation. However, much work still needs to be done by vendors in this area. While most CAD and GIS vendors can connect to a spatially-enabled data store, and while standards such as OGC Simple Features for SQL exist, the way in which data is actually held within the database still differs across systems. This can lead to 'middleware' software to cater for these differences, somewhat hampering interoperability.

In whatever field of GIS you practice, there is no doubt that interoperability is here to stay. Standards will have an increasing influence on the GIS market as they have done in other areas of IT.

International endeavours to foster 'joined up government' (real or imagined) will only accelerate moves in this area. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) was published in the official Journal on the 25th April 2007. The INSPIRE Directive entered into force on the 15th May 2007. Look out!

Martin McGarry and Paul Bigg-Wither (Cadcorp)