What the hell is GML?

In an industry where there are many vendors, each with their own product version and proprietary data format, there comes a time when two systems need to be able to talk to one another. In the past this was accomplished in a number of differing ways: write an import/export routine for your application, or hope the vendor will do it for you; use a third party translation program; or write a piece of middleware to translate data sent to it automatically. All of these solutions have pros and cons, especially if you have a number of different systems on your network.

One of the fundamental problems is deciding which format to base the data on. This becomes especially problematic when you are a data provider, where having to produce your data in a number of formats increases overheads. Also, if you choose to provide data in one vendor's format what about the users of another product? Will they buy your data? What if the data format you choose cannot handle the complexity of the spatial model you are trying to provide? It's a simple model for line and points, but what if you are trying to model cable networks or areas of forest, each with their own specific relationships between data types? How can you share your data between disparate Internet systems over a wide area network? All of these problems have not been unique to the GIS industry. Such problems of data interoperability have long been an issue throughout computing, an issue that the eXtensible Mark-up Language (XML) was designed to help solve.

The adoption of GML by Ordnance Survey shows that even though GML may still be new to many people, over the next few years it may well become one of the most important data formats in the country.

GIS is not limited to single corporate networks. The provision of on-line spatial services and vendor-independent spatial clients means that, increasingly, systems are interoperating with systems not just on the data level but also at the application level. This means that we not only need to share the data between systems but that we also need to structure that data in a way that has meaning in the appropriate paradigm. We need a method for storing and representing data in an extendable and flexible way, but also in a way that is vendor-independent. While individual systems, especially major vendors' packages, will continue to do things in the way they perceive is best, a widely adopted neutral format could revolutionise data sharing and interoperability throughout the GIS industry, eliminating the need for expensive third-party translators.

Of all of the proposals and standards that have come out of the Open GIS Consortium one of the most important recent developments has been the Geography Mark-up Language (GML) specification. GML is a language for representing spatial data and its meaning in an XML-based form. It has its own structure for encoding data and their associations, but as with all XML these can be extended to model an application's specific foibles.

At its most basic level GML provides a method for representing features. A simple polygon can be represented as text within an XML document as below:

```
<Polygon grid="_1000">
    <outerBoundaryIs>
        <LinearRing>
            <coordinates>
                100000,200000
                103000,201000
                104000,205000
                101000,201000
            </coordinates>
        </LinearRing>
    </outerBoundaryIs>
</Polygon>
```

Such documents are easy to decipher and can be extended to show the relationships between not only other features but also non-spatial attributes and concepts such as a city (with its associated building polygons, road lines and points of interest) as well as attributes such as name and population. While we can only scratch the surface of GML here, it does throw up a number of possibilities beyond the simple transference of data.

The overriding goal behind the implementation of GML was providing a means for transporting and storing data in an Internet environment. To do this it had to be able to cater for a wide variety of spatial feature representations, remove the presentation mechanism from the data and provide the ability to relate features to other features' attributes, to facilitate the modelling of spatial relationships.

One of the most important qualities of GML is its ability to provide a framework for modelling application relationships in a way that is implementation-neutral. This allows for an application to be designed independently but still be able to inter-operate as it is has been designed using the same base object model. In large systems with many interrelated components this can save weeks of time spent trying to relate disparate models. GML also lends itself to modelling using such languages as the Unified Modelling Language (UML). UML is a graphical object modelling language that can be used to help visualise the relationships between features and attributes in an application before any system implementation is done. If all systems are designed based on GML then the length of time it takes to design and deliver could be reduced, although it is very rarely so unless the project is tightly specified.

While GML is not yet widely adopted, there are some key areas where it is popping up. One of the most important for many spatial data users in the UK is the adoption of GML by Ordnance Survey for distributing its forthcoming Digital National Framework dataset. The document 'DNF data in GML' from Ordnance Survey shows exactly how GML can be used to represent real world data features and topology. The adoption of GML by the Ordnance Survey shows that even though GML may still be new to many people, over the next few years it may well become one of the most important data formats in the country.

Matt Toon is Internet Team Leader, ESRI (UK)