GE Energy

Smallworld Core Spatial Technology™ 4

Integrating data from many sources using a consistent and extensible approach

GER-4237 (10/04)

Integrating data from many sources using a consistent and extensible approach

Abstract

Enterprise GIS systems are most effective when they are able to leverage data from a wide variety of disparate data sources. Many legacy systems provide such access using bespoke technologies with inconsistent application programming interfaces (APIs). This frequently requires lengthy translation processes and specialized code to be written to access and manage each of these data sources. This can extend the deployment phase of an enterprise GIS system, increase maintenance costs and make it more difficult to sustain quality.

The Smallworld GIS architecture, from GE Energy, contains several unique technologies that greatly simplify this kind of integration by providing a consistent, extensible mechanism to access and manage data irrespective its original source.

The challenges of data in the enterprise GIS environment

Many GIS products today offer at least some limited support for accessing external data: open database connectivity or ODBC is a common example. However, in the enterprise GIS environment, the data that a business seeks to tap into is often found in a wide range of disparate systems. Some of these systems adhere to common standards but more often than not they are legacy systems with proprietary data formats or specialized systems that have optimized (or custom built) application programming interfaces (APIs).

This environment presents many challenges and hurdles for the application architect and systems integrator.

Faced with the task of integrating data from a single data source, most application architects will hope to leverage an existing open standards technology such as ODBC. If it is then required to integrate a second data source, say for example from a legacy system, then most architects will probably resign themselves to writing a one-off translator to read the data from the legacy system into the new system. If a business need arises that requires access to a third data source, say some high performance satellite imaging system, then most application architects will begin to think there must be a better way.

The scenarios outlined above might sound contrived but in the reality of the enterprise GIS world they are all too common.

This raises several practical issues that need careful consideration.

Writing code to translate from one system to another is a time consuming process and also a frustrating one, if one of the systems does not publish the format of its data. This code will also require testing resources and quality assurance. As in many cases, businesses do not see the value of data that can be shared across platforms until it is truly needed.

Less obviously, the reality of the enterprise GIS environment means that legacy systems are not simply turned off the day the new system is turned on. The practicalities of today's business operations often mean that migration from the old system to the new system is a lengthy process that requires careful management to ensure that the data the business needs day-to-day keeps flowing. From the integration point of view this means that it is quite common for both the legacy system and the new system to operate simultaneously for relatively long periods of time. Inconveniently, it is also often the case that new data will continue to be added to the legacy system as it is still satisfying an operational business need.

Clearly, in this context, translating data from one system to another begins to lose its appeal. Not only does the translator need to be written and tested, but in this scenario the translation process will need to be

repeated periodically to keep the data in the new system in step with that in the old system. However, the data in the new system will be out of date almost the day after it was updated. For some business applications, this might not be acceptable.

Other suppliers might provide APIs that support live access to their data, negating the need for this time consuming translation. But here as well, hurdles present themselves to the application architect. For example, one system might have a different transaction model while another system might have a unique way of accessing data, while yet another a specialized way of managing database connections. The subtle differences in the characteristics of spatial data raise even more perplexing issues.

More often than not, these APIs have been optimized for the data they provide. This means that, at best, the code developed to access data from one system using its API will have to be modified to allow it to work with another system's API. At worst, it means that little or even no code can be reused and another bespoke development is required. This custom code requires development and needs to be tested and maintained, all of which consume resources and time and is compounded by the magnitude and variety of the data demanded by business' using today's modern GIS applications.

A consistent, extensible approach

GE Energy's Smallworld architecture has been designed from the outset to facilitate the easy integration of data from a wide range of disparate sources using a consistent and extensible approach. This advance addresses not only the technological demands of GIS in the enterprise, but also the practical day to day issues regarding its deployment and maintenance.

Instead of having to provide costly and time consuming translation or additional code to support bespoke APIs, the Smallworld platform supports an advanced on-the-fly translation technology called Spatial Object Managers (SOMs). SOMs allow data in an external system to remain there as long as it is still needed, while at the same time, manifesting itself inside Smallworld as if it were native data. This important development avoids the need for repetitive, time consuming translation procedures and reduces the need for additional bespoke code. Importantly, this ensures that applications implemented on Smallworld have access to the most up to date information. It also means having a consistent API to access the data, perform queries and to manage transactions that might involve ensuring integrity across several data sources. Another attraction is that application code written for one data source can easily be reused for another. The underlying power of the Smallworld architecture deals with the special implementation details in a way that is transparent to the application developer. This benefit extends not only to conventional attribute data, but also to the special characteristics of spatial data. For example data sources in different projections and co-ordinate systems can be transformed on-the-fly to a common projection and co-ordinate system.

The Smallworld platform provides a number of SOMs that support access to a wide range of popular data formats and help expedte solutions to data providers such as:

- Spatial data held in Oracle®
- Highly compressed image formats from LizardTech (MrSID™) and Earth Resource Mapping (ECW)
- The Open Geospatial Consortium's WMS standard
- Engineering designs published in DWG, DXF or DGN
- Legacy data stored in formats such as MID/MIF, IBM® GFIS IFF and ESRI® Shape files
- Numerous image formats such as JPEG, PNG, TIFF, GIF and so on

Finally, the Smallworld architecture provides an extensible SOM API allowing third parties to develop their own SOMs. This allows them to access data from a particular legacy system or a specialized data source.

Conclusion

Today's enterprise GIS applications require access to a multitude of different data sources to satisfy the needs of an increasingly more demanding business environment.

Some GIS architectures over-extended a piecemeal approach to data integration to the enterprise. These architectures were developed to support small scale GIS applications and were not well suited to the enterprise environment and led to unwieldy integration strategies, increased cost and deployment delays.

The underlying power of GE Energy's Smallworld architecture is that it offers an elegant, consistent and extensible solution that not only addresses the technical issues of integrating a wide range of data sources, but also acknowledges practical realities of the enterprise environment as well.