Abstract
This paper describes the challenges and benefits surrounding a web-based data distribution system.

Implementing a web-based spatial data distribution can be a challenge, but many organizations are doing it successfully and reaping the benefits. Before enterprises undertake such projects however, they must first allocate time to understanding the quality of their data, and the target audience.

Traditionally, it has been very difficult for an enterprise to consolidate its disparate map data into a single, seamless database, and integrate this significant asset into the decision making process. For the past 30 years, organizations around the world have been capturing spatial data digitally in a wide variety of data formats. With thousands of data formats, sharing mapping data is a complete process. Industry sectors, governments, and even departments often work in data formats that most appropriately address their needs. Recent developments in spatial database technology from a variety of vendors are making it possible for organizations to realize the dream of integrated spatial and attribute corporate databases. Although the transition from mapsheet files into a spatial database can be difficult, many organizations are now successfully doing this in order to leverage their significant spatial data investment. This article provides guidance in preparing for such a migration, independent of spatial database type.

Overview
Spatial data is being used by an ever-increasing number of organizations - from city to national governments, and from small companies to large corporations - who all view spatial data as a strategic asset. As spatial data increases in importance, both businesses and governments need to disseminate and have access to the latest data as cost-effectively and as fast as possible.

As the need for spatial data grows, there is also an increasing number of web-based mapping systems that enable users to view data, and perform simple analysis and other basic GIS operations. The focus of these mapping systems was on providing GIS-based functionality over the Internet/intranet; however, the products have a limited native ability to distribute data.

Historically, spatial data has been distributed using physical media; and, since spatial data is voluminous, data providers were often forced to provide the data in a single format and a single coordinate system. As a result, data consumers who wanted the data in a different format or coordinate system had to convert the data either by writing customized software or by using a commercial data translator such as Safe Software’s Feature Manipulation Engine (FME) or Blue Marble Geographics’ Geographic Translator.

The growth of Internet and web-based technologies provides new distribution possibilities for spatial data users and providers. Web-based data distribution products are now hitting the market, such as Safe Software’s SpatialDirect.

When you are choosing a web-based data distribution product, you will need to consider some key points to ensure that the system satisfies both your immediate and future needs.

Spatial Database-Based
Spatial database-based systems provide superior performance in addition to the benefits of a Relational Database Management System (RDBMS). While it is possible to distribute data via the web from data holdings that are not stored in spatial databases, spatial databases such as Oracle Spatial and ESRI’s ArcSDE, provide much better functionality.
Web-based data distribution systems that are built on spatial databases are also in no way limited or complicated by file boundaries or other tiling issues, that is the complete data holding can be represented as one contiguous dataset. The first generation of web-based data distribution systems often require all the data to be periodically dumped into a proprietary file format. This restricts first generation systems to smaller datasets and to systems that do not require live updates. It also means extra work is necessary since the data has to be copied from the data source to the web-based format on a regular basis.

Open and Standards-Based Architecture
The web-based data distribution system must have an open architecture that enables integration with third-party products and be based on OpenGIS® Consortium (OGC) standards so that it can easily work with other standards-based products.

In particular, the web-based mapping server should be able to act as both an OGC Web Mapping Server (WMS) and an OGC Web Feature Server (WFS). The Web Mapping Server specification is the simplest of the OGC Web Services. It defines three operations:

- **GetCapabilities (required):** Obtain service-level metadata, which is a machine-readable (and human-readable) description of the WMS’s information content and acceptable request parameters.
- **GetMap (required):** Obtain a map image whose geospatial and dimensional parameters are well defined.
- **GetFeatureInfo (optional):** Request information about particular features shown on a map.

The OGC WFS specification defines the following operations for a Web Feature Server:

- **GetCapabilities (required):** A web feature server must be able to describe its capabilities.

Specifically, it must indicate which feature types it can service and which operations are supported on each feature type.

- **DescribeFeatureType (required):** A web feature server must be able, upon request, to describe the structure of any feature type it can service.
- **GetFeature (required):** A web feature server must be able to service a request to retrieve feature instances. In addition, the client should be able to specify which feature properties to fetch and should be able to constrain the query spatially and non-spatially.
- **Transaction (optional):** A web feature server may be able to service transaction requests. A transaction request is composed of operations that modify features (that is create, update, and delete operations on geographic features).
- **LockFeature (optional):** A web feature server may be able to process a lock request on one or more instances of a feature type for the duration of a transaction. This ensures that serializable transactions are supported.

The OGC divides WFS implementations into two categories:

- Basic WFS -implements the required operations (a read-only implementation), and;
- Transaction WFS -adds the Transaction capability, and possibly the LockFeature capability, to the Basic implementation.

When selecting a web-based data distribution system, it is definitely worth knowing what level of compliance the system has with the OGC and what the future direction with the product is.

User-Directed
The data export operation must allow the user to select the data to be exported. This should, at a minimum, allow the user to do the following:
Web-based spatial data distribution systems

- view the data,
- perform simple mapping operations such as pan and zoom,
- select the themes that are to be delivered,
- select the format for the delivered data,
- select the projection for the delivered data,
- specify the geographic extent of the delivered data.

Once the data is selected, the system should only send the data that the user selects. This requires the system to be able to do on-the-fly data extraction, and data clipping. It’s a more efficient use of communication bandwidth if as little data as possible can be packaged for shipping to the user.

Scalable
The system must be able to satisfy clients using the smallest single CPU machine data distribution systems to large clients using multi-machine systems, and must be able to easily grow from one extreme to the other without causing organizations to lose their investment. The architecture must thus be flexible enabling software components to be easily moved from one machine to another with minimal change to configuration files.

Secure
The system must be secure in two ways:

- it must not allow users to see any restricted data, and
- it must guard against requests for too much data that, if processed blindly, would result in loss of or degradation in service.

Since the Internet can be a very hostile environment, there must be a layer of software between the underlying database and users, ensuring that users cannot find ways to sensitive data. If the system detects any attempts to thwart the security, then it should log this information with as much user and/or IP information as possible, and notify system operators.

The system must also be capable of handling requests for too much data. For example if there is a theme named “Roads” that contains all the roads in the continental United States, the system should guard against a misinformed or hostile user that requests all the roads for a particular state or for the whole country. This is too much data for a real-time request and processing such a request would greatly degrade the system performance.

Ideally, systems administrators should be able to define the size of data that is to be distributed on a layer-by-layer basis and provides for different levels of service based on the amount of data that is requested. An example of one possible set of different levels is described below:

- Real-time Service: This is for small requests. This value is dependent on a number of factors: server bandwidth, client bandwidth, number of expected simultaneous clients, and throughput of server. For these requests, the system processes the request immediately with a turnaround time that would be acceptable for a user waiting at a browser.
- E-mail Service: These requests are the next level in size. The server still processes the requests immediately, but it is recognized that the delay is beyond the threshold of a user waiting at a browser. The user is sent an e-mail message with an FTP link that points to the extracted data.
- Physical Media Service: This level of service is for data requests that are performed off-line and then put on physical media. These results are deemed to be simply too big to be sent via the communication infrastructure.
- Prohibited Service: This is for requests that are deemed too large to process. The request is logged and the client is simply notified that the data request is too big for the data distribution system.

For more information about Safe Software, call us at 604-501-9985 or visit our website at www.safe.com
Web-based spatial data distribution systems

Reliable
The system must be reliable, and at the same time it must have an administrative capability that catches and reports faults. It must also have a statistics reporting capability so that administrators can see how the system is performing. If any bottlenecks exist; administrators need to know where they are located so that future performance issues can be identified before there is a serious impact on the users.

Cost-Effective
The data distribution system must be cost-effective, providing a cost based on server configuration or number of concurrent users and not on the total number of users.

The data distribution system must also be able to be used without requiring software be installed on the client machine. For Internet-based solutions, it is best if the software can run from a standard browser such as Internet Explorer or Netscape without requiring plug-ins.

Summary
The move to web-based data distribution systems builds on the trends to move spatial data into databases and GIS functionality to the web. When choosing a web-based data distribution system, an organization must ensure that the system meets both their immediate and future needs. The chosen data distribution system must have an open architecture and must adhere to industry standards so that it can easily work with the web mapping solutions from both current vendors and future standards-based products. The product must be scalable, able to grow with the need to distribute data. Last but not least it must be cost-effective - not priced on number of users, but on server configuration, which enables the deploying organization to benefit from the continual decline in computer hardware pricing.