

# Spatial ETL: Making Data Accessible

---

Spatial data is all around us. Businesses are looking to their spatial information now more than ever for assistance in analyzing situations and making decisions. But before spatial data's full value can be realized, it must be accessible to the people who need it. The ability to freely share and use spatial data is too often plagued with interoperability challenges. Proprietary, evolving formats, unique data models and a lack of standards all contribute barriers to the flow of spatial data.

More than 15 years ago, spatial ETL tools emerged as a way to help GIS professionals make spatial data accessible to end users in the applications of their choice. Regardless of the data's originating format or data model, spatial ETL tools make it possible for data users including surveyors, planners, developers, executives, the general public, and even GIS technicians themselves to work with spatial data in the GIS, CAD and other applications they use for analysis, visualization, and planning.

To demonstrate how important spatial data accessibility is, imagine for a moment that you've come across someone needing medical assistance. Here, calling for an ambulance is only valuable if the person receiving your phone call makes your request (the data) available to an ambulance driver who can respond. But not only does the ambulance driver need to know that someone needs his help, he also needs the spatial data – *where* to find the person – before he can successfully respond.

Data is worth nothing if it just sits there. Across all industries, spatially enabled data must be accessible to the people who need it before its full potential can be achieved.

## Extract, Transform and Load

If the acronym ETL sounds familiar, then you won't be surprised to hear that when Safe Software coined the term "spatial ETL" in 1993, they simply meant "extract, transform, and load spatial data." Traditional business intelligence ETL products were designed for the same purpose, but with one distinct difference: they neglected to include the spatial world of an organization, which industry pundits estimate makes up 80 percent of all corporate data.

The goal of spatial ETL tools is to leverage this existing spatial data – often the most untapped, yet key, corporate asset – to enhance business analyses and ultimately influence business decisions.

So how do spatial ETL tools make spatially enabled data accessible? They provide an economical mechanism for quickly converting spatial data into the required format, projection, and data model so that the information can be used by the people who need it.

To achieve this end, spatial ETL is comprised of three capabilities:

- *Extract* the data from its source
- *Transform* the data as required to make it usable
- *Load* the data into the destination view or dataset

## EXTRACTING DATA FROM ITS SOURCE

The first step in spatial ETL is extraction. In the extraction process, the spatial ETL tool gathers the source data for processing, copying it from its original location. This ensures that when transformation is performed in the next step, it can be done without impacting existing infrastructure.

## TRANSFORMING DATA TO MAKE IT USABLE

Transformation is the core functionality of spatial ETL, as it is the process that converts the data into the format, structure and coordinate system that end users require. Transformation occurs in three forms, enabling users to:

- *Translate* spatial data into the format that the end consumer requires;
- *Restructure* the data into the required data model and coordinate system;
- *Integrate* disparate data into a single, central view or database.

### Translate

The process of translation is, in its simplest form, simply converting a set of data from one format to another. This process enables data to be opened and used by a different application than the one which corresponds with its original format.

### Restructure

Restructuring occurs when the data's schema must be modified to suit the destination format, destination database or the users' needs. In many cases, restructuring can involve changing the coordinate system in which the data is stored and filtering data so that extraneous information doesn't clutter the destination dataset. For databases, restructuring also routinely means rearranging the names of tables and columns, and even restructuring how the data is organized within these tables and columns. File formats often require similar restructuring of the data model to correspond with the application they are designed for.

### Integrate

Integration focuses on combining data from multiple sources. While simple integration processes may gather data from separate datasets which are already stored in the same format, complex integrations can combine data from disparate sources and data types. For example, data created in a GIS system can be integrated with data authored in a CAD system and other data authored in a BIM system.

Effective spatial ETL tools allow users to mix and match these three processes – translation, restructuring, and integration – into a single spatial ETL task. This allows the users to create destination datasets that are immediately usable to end users according to their specific needs.

## LOADING DATA INTO DESTINATION DATASETS OR VIEWS

Once data has been extracted from its source, converted into the required format, data model, and coordinate system, and even integrated with data from other sources, it is time to load the data. The data can be loaded into either a destination dataset or a data view.



Today datasets are the most common destination, but as technology advances and people who wish to only view spatial data realize that this is possible, data views are becoming a popular way to load data.

However, the advantages of a destination dataset – including the ability to save and open the data within applications for analyses, manipulation, and authoring – will certainly maintain the popularity of datasets as destinations, especially for data authors such as GIS professionals and planners involved in CAD drafting.

Once the data is loaded either into a destination dataset or view, it can be accessed and used by the people who need it because it has been converted into the format, data model, and coordinate system they require. End users can now open the data in the application they use to perform analysis, visualizations, and more.

Additionally, when the data has been combined from multiple sources, the end user can access everything they need from one central source. This concept is termed “create once, deliver to many” as authors can continue using the systems of their choice while end users can access all of the data they need from a central location – with spatial ETL tasks running quietly in the background to keep the destination dataset up-to-date.

## Key Requirements for True Spatial ETL

Now that the basics of extracting, transforming and loading have been explained, it is important to understand three key design features which effective spatial ETL tools use to accomplish these tasks: centralized processing, semantic translation, and model-driven transformation.

### CENTRALIZED PROCESSING

The most effective spatial ETL tools extract data through a centralized processing engine that uses a set of readers (extractors) and writers (loaders) which act as plugins. When spatial data is extracted, it is pulled into the processing engine by the reader which matches the data’s originating format, and then is transformed by the engine and loaded into its destination using a writer plugin. In cases where data is being integrated from multiple disparate sources, the tool simply makes use of multiple readers.

This centralized processing technique ensures that the spatial ETL tool can convert data to and from the widest range for formats, fluidly growing to meet market changes by adding support for new formats as they emerge. When a new format is identified, developers introduce support by simply creating a reader and writer plugin for the format rather than writing individual tools to translate data between one specific format and another.

In today’s market which already consists of over 200 formats, centralized processing is crucial. Without such a capability, end users would be faced with searching out a specific spatial ETL tool for each different combination of formats. Centralized processing ensures that a single spatial ETL can grow to support whatever format-to-format combination is required. Future support can be achieved in a timely manner, and end users can count on one tool to accomplish all of their spatial ETL needs rather than selecting unique tools for each need they have.

### SEMANTIC TRANSLATION

While many applications claim to import and export data, successful translation only occurs when data models are taken into consideration. Semantic translation offers the complete approach to translating data from one format to another because it considers the data’s original data model and matches it to a structure that the destination format can recognize.



Before understanding how spatial ETL tools accomplish this, it is important to understand a bit more about data models in relation to formats. Many formats carry their own unique data models – in fact, the entire basis of some formats such as standards-based sharing formats is their data model – yet others, often proprietary databases, enable organizations to choose their own data model within some limitations such as what geometries are allowed and the quantity of data which can be maintained.

In both cases, data models are tied very tightly with formats. Achieving semantic translation requires a spatial ETL tool to be aware of the rules and regulations of each format and data type, and to be capable of translating data between them with minimum involvement from a user. This capability involves rich handling of the geometry types available across the many formats and data types in order for data quality to be maintained.

Proficient spatial ETL tools overcome this barrier to translation in two ways. They can:

1. Automatically account for data model transformation during translation between formats which have set structures
2. Enable users to restructure data models as needed for a translation – but without requiring the users to handle development code.

### USER DRIVEN DATA MODEL TRANSFORMATION

Although format-to-format translations are common, they are not the only situation in which data models must be considered. Safe Software's recent customer survey demonstrated that more than 50% of spatial ETL tasks do not translate data from one format to another, but instead restructure the data's schema so that it can be more effectively used.

Spatial ETL tools enable their users to perform data model restructuring without having to develop a single line of code by offering intuitive user interfaces which enable models to be restructured visually.

These visual workflows can then be saved, reused and even modified for future transformation tasks. This is often necessary to maintain up-to-date datasets by pushing new data that has been created using various applications into a centralized destination dataset.

### Applying Spatial ETL

The above three design features ensure that organizations choosing to use spatial data as a core part of their business strategy have an efficient way to make their spatial data accessible to the people who need it. Gone is the idea that individual departments must work in their own data silos, with IT structures limiting the company's potential to truly work as one. Instead, spatial ETL tools find their strength in tying data from disparate systems together, simplifying data transformation projects, and ultimately making data accessible across the organization and beyond.

The best way to truly understand how spatial ETL can make your data accessible in a more productive way is to try it out for yourself. Download Safe Software's FME and try it free for 14 days by visiting [www.safe.com/eval](http://www.safe.com/eval).

