Conflation of multi-source data

How using the 1Integrate rules engine for conflation of multi source data can automate and ease the process resulting in big benefits

Challenge

Many organizations manage or rely on location data. They must deal with the fact that the data can come from a variety of sources in many different formats, scales, data models, granularity, segmentation etc.

Typically data is integrated by converting it into a common format and a common data model, but this is only the start of the process and is often where problems begin.

Very often, organizations are confronted with challenges due to the precision, scale, meaning and structure of the data. For example the same objects may be duplicated among the different sources such as a street being represented as one line feature in one dataset and as multiple line features in another. This problem of merging different representations of the same data is known as conflation and requires some intelligent matching of the disparate data.

In this process, because the features are similar but not exactly the same, it is difficult to establish a measure of "spatial" similarity on which integration can be based, especially for datasets created at different times and with different structures, levels of accuracy and precision.

This is the case in the following example: objects in the red layer must first be matched to objects in the black layer before they can be combined. It is easy to see the complexity of this operation, due to the difference in precision and location of the data and also the difference in how the data has been split up.

Solution

1Integrate applies rules that can be used to match heterogeneous geometries based on the similarity of the objects' shapes as well as their connectivity and attributes. 1Spatial is a global leader in this type of conflation due to successful implementations at many client sites, using 1Integrate software to solve complex conflation problems and a repeatable, automated rules-based approach.

There are several use cases for this kind of matching, in particular:

- To allow different applications to interoperate, each one having access to the same data in the same structure, as opposed to different versions of the same data
- To create an improved dataset by coupling the geometric accuracy of data produced for mapping purposes with the attribute completeness of another data source created for more specific analysis needs
- To create an improved dataset by retrieving the most up-to-date data from the two sources, making the resulting analysis more complete.

Figure 1: Complexity of matching two versions representing the same real-world features

1spatial
YOUR WORLD SMARTER

Advantages of the solution

The 1Integrate solution addresses the following problems, which are essential in order to achieve a high level of conflation and reliable results:

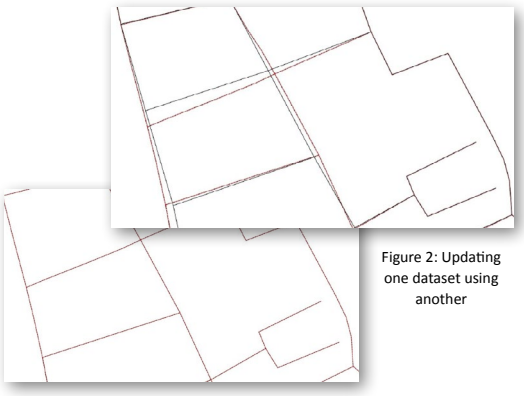

Figure 2: Updating one dataset using another

- Matching objects between datasets even when their geometries are different and there is no common identifier
- Identifying moved objects and differentiating from new or deleted objects even when there is no common identifier
- Matching of datasets that represent the same data but are structurally different (e.g. 1 object represented by multiple objects in another dataset)
- Recording a confidence match to help operators review the result
- Retrieving and incorporating geospatial data available in an external dataset (e.g. open data) into your own dataset
- Using rules to allow the process to be easily and cost effectively repeated every time the data changes.

Implementation

Due to the typically large amounts of data and frequent inter-feature checks involved in these processes, traditional scripting, GIS or workbench-based tools struggle to process the data. 1Integrate however is a rules engine that provides both powerful functionality, high performance and scalability over large datasets.

1Integrate is not a black box: rules are authored and run using a collaborative web interface, making it possible to refine processing at any time, to add or modify rules, or to apply them to new datasets. In this way, the organization builds its catalogue of rules for its processes of quality and integration of company data, connected to its ERP or any other computer system.

Due to the web-based interfaces, 1Integrate can be installed either on-premise or made available as a cloud service.

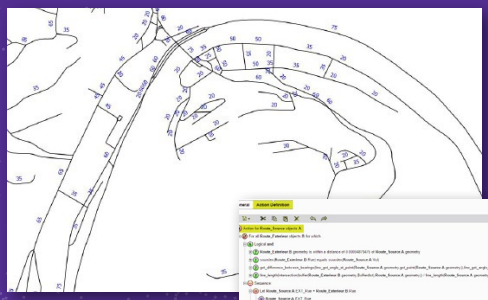

Figure 3: Copying speed limits from one dataset to another to improve client's road network

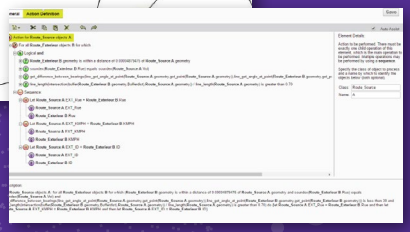

Figure 4: Extract from one of the speed limit conflation rules

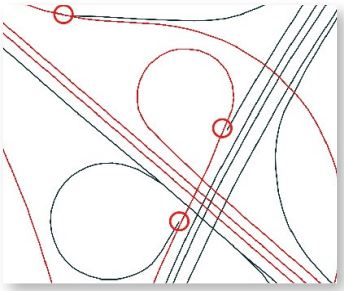

Figure 5: (Above Left): Matching of Topo10NL from the Dutch Land Registry (red) and Open Street Map (black) data using automated rules.

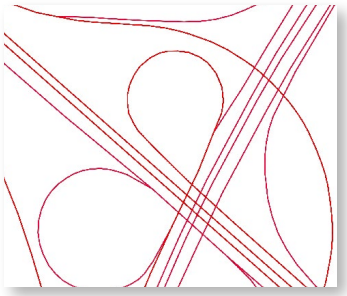

Figure 6: (Above Right): Shows the conflated result with connectivity issues corrected.

Conclusion

Data conflation enriches your spatial data, quickly and reliably. It allows you to combine the most accurate or up-to-date sources of data to get the best overall result or to enrich them with free, open data.

A manual approach to data conflation is costly and time-consuming (often several man-years of work), and its quality is not guaranteed due to tedious and repetitive human operations. 1Integrate allows reliable, repeatable and very fast automation of this task. The automation and minimal manual intervention means that an investment in 1Integrate can produce very large savings for repeated processes. The openness of the interface means that it can also be used for many other automatic validation and data processing tasks within the organization.

For more information about 1Spatial's solutions
email info@1spatial.com
or visit 1spatial.com

1spatial
YOUR WORLD SMARTER