

Spatial Data Quality: Happy Hunting Ground

Laser-Scan has contributed to the OGC interoperability specification process since 1996, playing a significant role in test-bed programmes and OGC contributions to ISO TC211. Within the context of OGC, Laser-Scan is synonymous with topology and its role in GML. Indeed, the company played a key role in the decision to base GML on XML. These past accomplishments are fine, but I am mindful of the comment in the eSpatial column in these pages a year ago: "It is often difficult, expensive and risky to integrate GIS software with general applications." From a geospatial industry viewpoint I believe we, the industry, are making progress, but from a Business Intelligence perspective we are only scratching the surface of placing our capabilities in the decision-making process. Why?

Contract of Rules

Quality is one reason. As a body of professionals working together in OGC we have made a significant contribution to Web services. But as an industry working in a distributed computing environment (as part of the overall interoperability paradigm) how do we deal with the issue of the provenance or quality of data returned? How do we conflate what is returned? In Web terms, how do we deal with the spatial equivalent of "Error 404: The requested URL could not be retrieved?" We believe that the concept of spatial data quality can be simply represented in terms of a contract of rules. Consider, as an illustration, a land management/property registration application used to record ownership rights. In such a system, the business rules concerning the spatial data are well understood:

- every piece of land (parcel) has owners
- land parcels do not overlap
- land parcels do not have gaps between them.

Generic Concepts

Once these rules are adopted it becomes possible to monitor them and to quantify the impact of drift. Most importantly, these rules state the formal set of conditions that should be met before data can be said to be fit for purpose. The rules can be used to assess the current state of spatial data quality and to plan for data quality improvements. By making rules explicit as enterprise metadata rather than embedded in applications, much of the ambiguity associated with spatial data use can be avoided and the care and ownership of data better managed. While the rules themselves are application-specific, each relies on well-established and generic concepts. Finally, the required level of investment in spatial data quality can be gauged and monitored using information networks to plan the necessary rules frameworks and assess the impact of any new data usage.

We firmly believe that the next round of standards activity should be the creation of open specifications for spatial data quality, as data alone is insufficient. I suppose we could call it the happy hunting ground where the Semantic Web joins with the spatial world!