Spatial data on the web

By Linda van den Brink and Ine de Visser (Geonovum)

Background

Geonovum has been working for years towards the realization and success of the public sector spatial data infrastructure (SDI) in the Netherlands, and with success: an infrastructure was built on ISO 19100 Information models (UML) and Open Geospatial Consortium (OGC) standards like WMS and WFS, which makes geographical data readily available. The European INSPIRE approach is also a good working example of this. Most web services available for spatial data are WMS services, which serve pictures, not data. Although CSW and WFS services serve metadata and data, seen from outside the geospatial domain the data behind the OGC services is part of the "Deep Web": the data is published behind specialized web services and not readily available for the majority of web developers [Taylor and Parsons 2015]. This group of users is increasingly making use of, and creating, geospatial data, and is therefore seen by Geonovum as an important new target group to disseminate geospatial data to, in addition to our existing users of the SDI. This brings the following question: 'In which way can, in an evolutionary way, the current SDI be leveraged and the majority of web developers be reached as well?

Geonovum wants geospatial data to be used. The public sector creates a lot of geospatial data, a lot of which is open data and could be useful to others, who are often not experts on spatial data. In addition, we have observed how geospatial data is becoming more and more important for the web and its importance is still growing, among other things because of the rise of the internet of things. And this importance is far beyond serving static 'pixels'. Semantically meaningful information on objects is required. Not just information about buildings, roads or waterways; but also e.g. legislative boundaries, permits and ordinances. In our opinion it's very important to integrate spatial data with other data on the Web. This is one of the reasons why we started the Platform Linked Data Nederland as a pilot a few years ago, and it is also part of the mission of the Spatial Data on the Web working group (SDWWG) that OGC and W3C have formed together, in which Geonovum is participating.

#geo4web testbed

Geonovum sees the web as an important dissemination channel and wants geospatial data to be accessible to web developers with no specific geospatial expertise. To explore the possibilities of making spatial data a useful, integral and common part of the web, Geonovum organised a 'testbed', an experimental project in which several market parties cooperated to make spatial data findable through search engines, and usable for web programmers.

Within the testbed a lot of experimentation was carried out and several spatial datasets were published on the web in different ways: using Linked Data, as HTML pages with embedded schema.org markup, with APIs, or using a proxy on top of existing SDI services. The testbed participants were asked to report all their findings and based on these findings, a set of Lessons Learned was compiled and published. In a second run of the testbed, more experimentation was done with the goal of testing and improving the Lessons Learned. This testbed phase was only recently concluded and the Lessons Learned have not yet been updated.

The Lessons Learned and the latest testbed results contain some interesting points with regard to vocabularies and metadata. For example:

Structure your data in a way search engines understand: use Schema.org markup. In the testbed, a mapping of several datasets to schema.org was created, as well as a mapping of spatial metadata to schema.org, a GeoJSON to schema.org converter, and a proposal for improvements to the spatial part of schema.org.

Serve your data in many different flavours and use content negotiation to allow selection between them. While this is a basic lesson, testbed participants identified some issues with its practical implementation. For one, content negotiation via the accept header is not always practical; in some cases URL based content negotiation is still necessary. Also, the fact that Linked Data vocabularies use WKT for geometry serialisation while (Geo)JSON uses arrays complicates the use of spatial data on the web.

Serve your data for many different communities, not forgetting the geo community itself. Within the testbed a SDI proxy was developed that makes it possible to publish crawlable spatial data dynamically from SDI web services.

In conclusion, we can present the results of the testbed and the lessons learnt up to now.