Geospatial Linked-Data - In need of Common Practice

Stuart Williams
skw@epimorphics.com
13/1/2014

This topic is offered as potential theme for panel discussion.

With the exception of WGS84 point positions, there is a lack of "dominant common practice" for the storage and expression of geospatial information as linked data. This creates problems for linked data publishers, consumers, application, tool and intrastructure developers/vendors alike.

Lack of guidance for linked data publishers on how to prepare and publish spatial-data creates impedance to publication in the first place. The diversity of data publishing practice amongst 'early-adopters' induces uncertainty in data-users, particularly application developers, about available formats and access mechanisms. This creates impedance to the use of such geospatial linked-data as has been published. The lack of critical mass around any given approach leads to a lack of shared application libraries and storage infrastructure such as geo-spatial capable linked-data stores. Platform providers, both open-source and proprietary, offer different capabilities. Standards like GeoSPARQL have yet to have their impact in terms of creating a solid technology base.

Contrast this with the rich open-source and properitary ecosystem associated with the mainstream GIS and web mapping specifications from ISO TC211/OGC and lead industry players.

An example of diversity

The purpose of this example is not to critique either publishing approaches below but simply to note their difference.

Looking at Ordnance Survey(OS) and Office of National Statistics(ONS) linked data for South Gloucester Unitary Authority we see diversity in the way that each have represented spatial of the linked-data they have published.

The OS publication makes a separation between the adminsitrative region and its geometry. Amongst the properties of the geometry is a GML serialisation as a 'big-literal'. This is very close to the approach advocated by GeoSPARQL.

```
# Feature
<http://data.ordnancesurvey.co.uk/id/7000000000025559>
    a            <http://data.ordnancesurvey.co.uk/ontology/admingeo/UnitaryAuthority> ;
    rdfs:label  "South Gloucestershire" ;
    <http://data.ordnancesurvey.co.uk/ontology/geometry/extent>
            <http://data.ordnancesurvey.co.uk/id/geometry/25559-20> ;
    owl:sameAs
            <http://statistics.data.gov.uk/id/statistical-geography/E06000025> ;
    .

# Geometry
<http://data.ordnancesurvey.co.uk/id/geometry/25559-20>
    a        <http://data.ordnancesurvey.co.uk/ontology/geometry/AbstractGeometry> ;
    <http://data.ordnancesurvey.co.uk/ontology/geometry/asGML>
            "<gml:Polygon xmlns:gml=\"http://www.opengis.net/gml\"
                          srsName=\"http://www.opengis.net/def/crs/EPSG/0/27700\">
               <gml:exterior>
                 <gml:LinearRing>
                   <gml:posList srsDimension=\"2\">
                     356578.2 193831.7
                     356616.9 193903.4
                     ....
                     356578.2 193831.7
                   </gml:posList>
                 </gml:LinearRing>
               </gml:exterior>
             </gml:Polygon>"^^rdf:XMLLiteral ;
             .
```

Note the embedding of spatial reference system name (a URI is given for EPSG:27700) within the GML literal and the dimensionality of the polygon. There are open-source libraries capable of parsing and handling GML encoded geometries.

The ONS linked data for the same administrative region provides a different encoding of the geometry. Again, a 'big-literal' approach is adopted:

```
#Combined Feature and Geometry
<http://statistics.data.gov.uk/id/statistical-geography/E06000025>
    rdf:type stats:statistical-geography ;
    rdfs:label "E06000025" ;
    skos:notation "E06000025"^^<http://statistics.data.gov.uk/def/statistical-entity> ;
    statistical-geography:officialname "South Gloucestershire" ;

    statistical-geography:hasExteriorEastNorthPolygon
        "364835.100971767 196693.800540854
         364892.79927582 196480.599644655
         ....
         364835.100971767 196693.800540854" ;

    statistical-geography:hasExteriorLatLongPolygon
        "51.6679115 -2.5098613
         51.6659982 -2.5090055
         ....
         51.6679115 -2.5098613" ;
         .
```

Indeed ONS provided two renderings of the polygon, one using OSGB eastings and northings, the other using WGS84 lat and long. Whilst non-standard, the internal syntax of the literal values closely matches that of the gml:posList of the OS GML serialisation. The two spatial reference systems (WGS84 and OSGB) are implied from the property names used.

Whilst these small differences in presentation are not huge, they are unfortunate and create impedance because the variety so easily created by the data publisher is a barrier to the data consumer and particularly application developers.

The OS data is further complicated by some entries, eg. Scotland http://data.ordnancesurvey.co.uk/id/7000000000041429, having multiple extents and by some extents that use more complex GML constructs such as `MultiSurfaces`. This latter emphasises a need to be aware of the full syntax that may be used in a 'big-literal' expression.

Variation is further compounded when the RDF data is made available as or is transformed into other common formats, say XML or JSON. Rarely do the resulting renderings follow common JSON or XML idioms for the publication of spatial data e.g. GeoJSON or GeoRSS - not withstanding an corresponding 'big-literals' are preserved the renderings.

Further variation arises through the possibility of coordinate transformation; geometries rendered at different resolutions; and service mechanisms to trigger such transformations.

Whilst only two examples are shown here, there are several other diverse practices to be found.

Conclusion

The lack of a "dominant common practice" for the linked data publication of spatial-data hampers both its initial publication and the use of such spatial linked-data as has been published. The current diversity of practice stands in the way of the development of: common platforms for publishing and querying geospatial linked-data; and application libraries and frameworks for its consumption and presentation to end-users e.g. open-source libraries that are useful in the production of multiple data publications or in multiple data consuming applications.

GeoSPARQL is one specifications that may help to address the gap. However, widespread adoption and implementation remain a challenge.

The 'big-literal' approach meets with mixed reception. It hides the internal structure of a geometry from an ordinary (non-spatial) query engine making simple bounding-box queries hard; and it has a more complicated expression of simple point position data than current RDF norms. It effectively makes geometry a datatype, albeit with a complex syntax, that standards based engines can handle - and has the potential to bring most of the ISO/OGC Simple Feature Access (ISO 19125) capabilities into an RDF/SPARQL linked data world.