

Linking Geospatial Data Workshop: 5th - 6th March 2014

Proposed discussion item:

Modelling and managing temporal changes to geographic datasets

Authors:

Steve Peters, Department for Communities and Local Government

Bill Roberts, Swirrl IT Ltd

Introduction

This paper outlines the common problem of handling temporal change to geographic LinkedData datasets, with thoughts on how it might be resolved.

We would welcome feedback, and would like to work with others to agree a best practice or standard that could be applied consistently by data owners with similar data.

The problem

Geographic and spatial information does not remain static over time. Changes commonly arise through developments in the physical environment, for example extending transport networks or adapting buildings, or legislative/policy amendments such as adjustments to Local Authority or Parliamentary Constituency boundaries.

This matters to the UK Department for Communities and Local Government (DCLG) because we maintain and publish significant volumes of data (principally statistical outputs) about Local Authorities and other geographic entities over long time periods. For instance, our house-building and house price series include annual totals for individual Local Authorities, starting in 1996. Other series report data back to 1968, and – in one case – 1930.

We are now publishing various statistical series via our OpenDataCommunities triple-store (<http://opendatacommunities.org>), and will progressively extend the range and volume of Departmental data here in response to user demand. It is therefore essential that we identify a robust solution for managing changes to geography spanning long time periods.

Examples of changes that have occurred are:

- On 1 April 2012, amendments were made to the boundary between St Albans and Welwyn-Hatfield authorities.
- On 1 April 2009, Cheshire County Council and its 6 associated districts were disbanded and replaced by two new councils: 'Cheshire West and Chester' and 'Cheshire East'.

On a practical level, these changes will typically trigger:

- Amendments to information about the organisation/s or legal entities governing geographic areas – e.g. a change to description of an authority, with no change to the associated map polygons; and/or
- An amendment to the underlying geometry – including merging or splitting multiple areas.

Although the problem has manifested itself for Local Authority datasets, it clearly is not confined to this geography. It would occur where, for example, we have changes to postcode areas, census geographies - e.g. adjustments to Lower Super Output Areas used in the 2011 Census - and policy geographies: for DCLG, these include Enterprise Zones, Local Economic Partnerships and Neighbourhood Planning Communities.

Possible solutions

We have begun to tackle various aspects of the problem.

First, we have created a “Local authorities in England” URI set¹, comprising unique identifiers for current local councils and other local public bodies – such as Police, Fire, and Waste Management authorities. Its purpose is to define local public bodies as organisations, as distinct from the geographic areas they govern and serve. Relevant URIs are already defined and maintained by Ordnance Survey and the Office for National Statistics, so our set simply cross-references to them.

We are currently working on extending this “LAs in England” set to incorporate additional information about each public body. For Local Authorities, this will include address of headquarters, contact telephone numbers, and services provided to the public.

Second, we are exploring options for managing associated information about change over time. Alongside the “LAs in England” URIs we have a "Historic local authority information" dataset describing how LAs were at some point in the past.

We are considering using the Versioning Ontology² (first developed by Epimorphics to describe UK Environment Agency data) to describe the relationships between entities in this and the “LAs in England” set. That means that we need a separate 'base' URI for each authority, perhaps:

<http://opendatacommunities.org/id/district-council/st-albans/base>

¹ The “Local authorities in England” URI set is available at <http://opendatacommunities.org/data/local-authorities>

² <http://purl.org/linked-data/version>

This base URI would have a series of versions. By employing the Versioning Ontology, we could define an interval over which each version is valid, and relates it to its predecessor and successor.

The base URI is in some ways like a Wikipedia disambiguation page. 'You asked about St Albans. Did you mean St Albans as it is now? Or St Albans as it was before 1 April 2012'.

```
<http://opendatacommunities.org/id/district-council/st-albans/base>
  dct:hasVersion <http://opendatacommunities.org/id/district-council/st-
albans> ;
  dct:hasVersion <http://opendatacommunities.org/id/district-council/st-
albans/19960401-20120331> ;
  version:currentVersion <http://opendatacommunities.org/id/district-
council/st-albans> .

<http://opendatacommunities.org/id/district-council/st-albans>
  dct:isVersionOf <http://opendatacommunities.org/id/district-
council/st-albans/base> ;
  dct:replaces <http://opendatacommunities.org/id/district-council/st-
albans/19960401-20120331> ;
  version:interval { time interval during which this version is valid 1
Apr 2012 to present} .
```

Events that would trigger a new version:

- new authority created
- existing authority disbanded
- existing authority changes GSS code or boundary
- existing authority changes its type
- other legislation-driven significant changes

We could also include provenance information to describe the event that led to the creation of a new version.

If an authority disappears entirely and is replaced by new authorities, then this isn't a case of a different version, but we can still relate the old authorities to the new authorities that replace it/them.

In Open Data Communities, all data is organised into datasets. To keep the data simple for the most common use cases, we anticipate having a 'current' dataset that holds only the most recent version of all authorities, uncluttered by information on previous versions. Therefore finding information about current authorities would be same as now.

Previous versions would be held in a 'historic' dataset, containing only those authorities that used to be different to how they are now. This could be queried to find data that was valid on a particular date, by examining the interval of validity of each resource.

A SPARQL query to find an old authority would be reasonably complex, but possible. To make the process of historical querying easier, we are considering adding a 'convenience API', that would return a full set of authorities at a particular date in the past, or the version of a single authority at a particular date in the past.