

Linked geospatial data, and the BBC

A position paper for the March 2014 “*Linking Geospatial Data*” Workshop of the W3C.

Chris Henden, Paul Rissen, and Sofia Angeletou

Future Media News and Knowledge, BBC.

This paper aims to communicate the current state of play for geographic data and content for BBC Online. We want to show why place matters for the BBC audience, and provide a simple overview of the approach taken to date. Our hope for the conference is to provide any insights possible, find if we are in sympathy with others, and ideally to improve our approach.

Why place matters to the BBC

Part of the BBC’s explicit purpose touches on geography: “Representing the UK, its nations, regions and communities”. In order to do that, we make content for all of the UK, and want people to be able to find it.

Sometimes place can seem secondary, but for some BBC activities it is crucial. The recent storms showed how critical it is to know what’s happening where you are. The BBC provides timely information on Weather, Travel, and Local News, all of which are rooted in space.

However, while place matters, it is often one of the many properties of content, so it can sometimes be de-scoped during product development.

How do we use geography in the BBC?

BBC has the mandate to inform the license-fee paying public of the United Kingdom of news relevant to their lives. We aim here to provide an overview of how we use geography within specific BBC products, and to touch on some of the challenges being worked through.

It is also worth noting that the BBC has a particular fit for the geographic areas and levels of required granularity.

Traditional Localised Broadcasting and Online Services

Traditionally, the BBC has been broadcasting locally relevant news via broadcasting outlets (television and radio) by dividing the country into regions, based on transmission areas and editorial coverage, and until recently, the same approach has been taken online. Examples of this are the Local Radio¹ and Regional Variations of TV services² and the Local News Regions³.

Local Live News

The BBC wants to ensure that its news content is made available to the audience in the most relevant ways possible. By using the Locator service, which holds canonical geographical data for the organisation as a whole, along with a Linked Data tagging approach, we are beginning to tag our content with geographical areas. At present, we are running a series of trials in local news regions. Journalists are tagging their articles as they write, with the relevant locations - either at a high level (e.g. the tag 'Birmingham and Black Country'), or at a lower level, such as a particular village within the area. In terms of RDF triples, news articles are regarded as 'creative works', the predicate used is 'about', and the object is the place in question.

The use of predicates poses an interesting challenge for the organisation. It may not always be strictly semantically correct that a news article is 'about' a place. In an analysis of 100 articles taken from the current BBC News Online index for Derby, only 20 could reasonably be described as being 'about' Derby. Instead, we are often trying to convey a sense of relevance to the population of the place.

¹ <http://www.bbc.co.uk/radio/stations>

² <http://www.bbc.co.uk/tv>

³ <http://www.bbc.co.uk/local/>

The analysis piece mentioned suggested three other predicates that could be used - 'mentions', for when the article in question merely includes a reference to the place; 'impacts', for when the article is about something that will potentially affect the population in the place - for instance a debate in Parliament about the proposed high-speed rail network, HS2, may take place in London, but may be relevant to people in Derby, as the network proposed will pass nearby the city; and 'event', for when the article is about an event which took place in the location in question. The analysis revealed that there were 21 articles out of 100 that merely mentioned Derby, 14 that can be said to have described things that might have an impact on the population, and 43 which were describing events that took place in Derby. This suggests that, from the point of view of the journalist, there may be many ways to describe the relevance of an article to a place, and that one approach could be to instead associate the article with an event, where relevant, and then the event with the place where it happened. We must bear in mind, however, the usability of any form of tagging tool which has this more detailed flexibility built in - too complex a user interface, and journalists simply won't use it.

From the point of view of the audience, it is also clear that they can have many different relationships with location. Someone may live in one place, work in another, commute through a third, and have family on the other side of the world. When considering how we make our news output as relevant as possible to our audience, it will be interesting to develop news propositions that a user can customise to describe the different relationships they have with locations, so that we can bring news to them in an appropriate way. Probably the most relevant example of this at the moment is the BBC's online Travel News service.

Travel News

Similar to the main news output, Travel News has traditionally been organised in terms of the same BBC regions. The use of BBC regions makes sense in terms of television and radio output, in that depending on where you are, travel updates are essentially scoped to that geographical area. However, users may be interested in other places, and the regional divisions can be unfamiliar concepts. There is a new online proposition, currently in beta, which is no longer tied to the regional areas. Users can search for their local town or postcode district, and see the results displayed for the exact location they are interested in. The new functionality gives users much more freedom and control over the level of detail displayed, as you can now see incidents for specific roads or areas. The static map, showing only results for the predefined regions, will be replaced with an interactive map, allowing panning and zoom, dynamically refreshing the information displayed.

Elections

The BBC has a responsibility in its Government licence agreement to ensure audience awareness and understanding of the United Kingdom's democratic process. Thus, its election service is especially important, generates high audience figures and indeed, was one of the founding stones of the BBC's presence online. Informing users that they are able to vote in a particular area is an important part of this service.

The two main types of election that are currently of interest are those in which the UK population votes to elect representatives for a particular area of the country; the other is where they will elect members to serve on a council which represents and governs the area in question. Examples of the former are General Elections to the UK Parliament, or the European Parliament Elections. Examples of the latter are local council elections.

For the first type of election, it is crucial for the BBC to be able to accurately inform the user of the constituency that maps to the area of their interest. The user may provide this information in the form of a town, or a postcode. The geographical co-ordinates and shapes for all of the UK Parliamentary Constituencies, and EU Parliamentary regions, are held within a boundary service ("Locator"), enabling the return of relevant constituencies. There is a considerable user experience challenge, however, with overlapping constituencies (especially in terms of the difference between a UK Parliamentary constituency and a EU Parliamentary constituency). The European Election case also presents an interesting challenge in terms of its different mode of election - whereby candidates and parties, familiar to a UK audience, are allied to pan-European political groupings. Explaining both the choices available, and the results of the European Parliament, are challenges which the BBC has to engage with and resolve.

For the second type of election, that of the local council elections, the issues are slightly different. There is a subtle, but important difference, between the region itself, and the council which governs it. In a similar case to the Local Live example outlined above, we are encouraging journalists to tag their articles and election updates with the relevant council. Therefore, we need tags for the councils, as well as a mapping between the council and the geographic location, which Locator holds, and identifies uniquely through the use of GSS IDs, which are also used by the Office of National Statistics. For Councils, we are considering using the identifiers provided by Open Data Communities, which hold URIs for every Council in the UK, and also hold a relationship between the Council and the relevant area's GSS ID.



Weather

The local weather forecast provided by the BBC is an example of integrating distributed data sources. Weather is provided for postcode districts and towns. The URLs of the pages incorporate the ids from the Geonames database, or postcode districts. Each location page brings together the actual forecast information, together with video clips for the region, maps of rasterised cloud data, with Weather Warnings to be added soon. This brings in the requirement to relate the locations to the various data sets, mostly from external providers. Similarly to Travel News, one of the main shifts of recent years was from inflexible region maps to ones centred on specific locations. One of the challenges in Weather has been to correspond the level of granularity given by gazetteers and postcode districts (and then selected by the user) - with that of the forecast information. For example, some of the postcode districts in Scotland can be very large.

Coverage

The BBC has an unusual scope. Again from the Royal Charter, one of the public purposes is “bringing the UK to the world and the world to the UK”. The “UK” here includes all licence fee payers, which adds the Channel Islands and Isle of Man. Along with Northern Ireland, this makes uniform use of identifiers problematic. The Ordnance Survey, and the Government Statistical Service ids/URIs, provide the de facto models for Great Britain, but there are challenges in establishing full coverage for our licence-fee paying audience. Yet there are also public service and commercial requirements outside the UK, with the World Service and World Wide catering for international users.

Granularity

The BBC purposefully deals with geography at a particular level. For privacy and trust reasons, it does not want to be storing identifiable places of residence. Much of the content applies at the level of “populated place”, or region. There are, then, potentially fewer issues with detail and scale, and more capacity to work on semantic meaning – for example, when “Birmingham” is referred to, is it the concept, or a political area?

Language, and BBC guidelines

The BBC provides services for some indigenous languages: Welsh, together with Scots and Irish Gaelic. This support extends to place names, which need to be searchable, and displayed correctly. This creates the need for places to be semantically identifiable, so the language string becomes one of the properties of the concept.

There are also instances where the BBC has a preferred spelling for a place name, or country. To this end, we must override some place names, and ensure the changes are not overwritten. This means our assembly of location data can be different from that in the public domain.

Geography and technologies

A prime requirement is to relate a location – often that of the user - to various datasets for that location. This is currently fulfilled by using “Locator”, which provides geospatial lookups.

Representing reference geography

The current approach is to express and store locations to a reasonable level of granularity, using a subset of the Geonames database (populated places), together with postcodes at a district level. Full postcodes are used where needed for search precision, but not to store location preference.

Relating geographies

Often details are required about a location, such as its broadcast area, nearest railway stations, or Westminster constituency area. Locator incorporates a PostgreSQL database with PostGIS, holding Geonames, postcode districts, and various domain specific shapes such as TV and Radio regions. A location can be mapped to the correct spatial data using operations such as “within” and “nearest”. Solr is used to search locations, disambiguating entered strings, or for simple reverse geolocation.

While “on-the-fly” location mapping is provided by the Locator, the ambition is that more stable geographical relations will be persisted in the Linked Data Platform. This will enable reasoning, and answer more complex and interesting queries involving locations such as “I want all BBC’s video clips about people born in Wales”. Such reasoning requires a combination of relationships to provide results, given that people’s birth places are usually more granular than country level. One challenge here is to maintain integrity of the data in different parts of the business. Another is in providing the correct flexibility in retrieving the data held within a triple store, with sufficient performance. What is the best way to find all news articles within 20 miles of a location?

Syndicating terms

The locations stored in Locator are available to other parts of the organisation through an atom feed, with any changes immediately published and made available. These are ingested to the Linked Data Platform, where they can be used for tagging and incorporated in other models. The requirement is for there to be a “single version of the truth”, so any place referred to is consistently available in different parts of BBC Online.

Linking data for the BBC

Many domains

As a media organisation, the BBC is fairly unusual in its wide range of activities. Doctor Who sits in the schedules along with world news, wildlife, and University Challenge.

Place could provide pathways through these different domains. China, South Sudan, 221b Baker Street, Leinster Gardens, Albert Square, Ambridge, and the Oval Basin in Cardiff, are all places which could mean things to our audience. The Torch Relay, or the current Queen’s Baton Relay, and the upcoming World War One season, are all underpinned by places.

If the geographies of our content could be clearly, meaningfully expressed, then potentially our users could create their own journeys through the wealth of BBC output.

Saving time

This may appear a mundane benefit – but it reflects the complexity of geography. Much more time is spent than necessary attempting to find the joining points between the datasets of different organisations – regardless of technical difficulties. Often conversations between organisations are held via proxies, in the form of account managers, far from the detail. Machine-readable, unambiguous identifiers, understood beyond individual organisations, promise to make everything a easier and more straightforward.

Hooking datasets together within the BBC

Elections are a good example of the benefit of different parts of the BBC making use of the same currency. Data comes from external providers, and we want it to match up with the geographies we understand. This often needs to be done quickly, with development teams allocated to projects just in time. The place name and postcode searches, the maps, the result data and the stories, all need to be using the same currency in order to snap together quickly. It is partly the geographic aspect of the elections that creates this need to bind together the different incarnations of the data.

Opening our datasets to the world

BBC has recently made a strategic statement on opening up its data⁴, signing four memorandums of understanding with the Open Data Institute, the Open Knowledge Foundations, the Mozilla Foundation and the Europeana Foundation:

We have worked closely with many other organisations to promote open data, open standards and the open internet, and we plan to build on these MOUs to find new ways to engage audiences and deliver the BBC’s aspirations online.”

James Purnell, BBC Director, Strategy and Digital

Linking our internal representations of geography to open geographies, such as the Geonames database, and other datasets, can only enable the consumers of our data to understand and reuse it to its best.

⁴ <http://www.bbc.co.uk/mediacentre/latestnews/2013/memorandums-of-understanding.html>