

Publishing Local Open Data

Important Lessons from the
Open Election Data project

Overview

During the May 2010 local elections, Socitm joined with the Local Government Association to support the **Open Election Data Project** (<http://openelectiondata.org/>) set up by Chris Taggart, developer of OpenlyLocal.com and a member of the Department of Communities' Local Public Data Panel, on which Socitm also sits.

Local authorities were encouraged to publish election results on their websites as '**linked open data**' – data that is published under an open licence that allows unrestricted reuse, and that is marked up to identify the structure and meaning, making possible its automated collection for re-publishing and mashing up with other data.

The purpose of the project was threefold:

- to start building an open database of local election results (none currently exists)
- to help local authorities develop skills and knowledge in publishing open data
- to identify issues, pain-points and blockers to publishing open data in general on a relatively contained body of information.

Given the Coalition Government's commitment to publishing open data, and stated aim of having Local Authorities publish their information as machine-readable open data, the project has been useful as an unofficial pilot for many other data sets that authorities may wish to publish in the future.

This document details the key lessons learned from the project, most of which are applicable to other local government datasets, and indeed other projects undertaken by a wide range of bodies interested in publishing linked open data to the same standard. We believe this is probably the first attempt to have a group of public bodies publish linked data to the same consistent standard, and as such there are important implications for similar future exercises.

The project and its outcomes

As indicated above, no freely available database of local election results exists. This is because the data required to compile it is located in different sections of hundreds of different council websites. The information is presented in many different formats and many different ways, and the only way the information can be compiled for re-use is manually - a laborious affair involving finding, cutting and pasting lots of individual pieces of information published by 433 different councils.

The Open Election Data project asked that instead of publishing their results using arbitrary and often inaccessible formats, councils should publish the results as HTML (the language used to write web pages) enriched with '**RDFa**'. Though invisible to normal users, publishing the information in this way gives it structure and meaning, and means it's possible for machines to read the information as data.

A dedicated website (<http://openelectiondata.org/>) was built by OpenlyLocal.com and the markup was agreed with leading linked data practioners. A special Open Election Data community of practice was established so that webmasters and others participating in the project could to report problems and seek advice as they set about implementation. In addition, presentations were made by Chris Taggart at localgovcamp/govcamp un conferences and Socitm canvassed its members and members of its web improvement community to take part

Two things happened after the launch of the project to help things along. Firstly, Lichfield DC webmaster Stuart Harrison wrote some code for the Jadu content management system, and this has been taken up and released by Jadu to its other local government customers to short cut the process for them.

Secondly, Modern.gov, providers of the main democratic/committee software applications used by local government, began to look at ways to add the functionality to their software so that results published through the application could be automatically be marked up as linked open data.

Many other web teams worked on individual solutions appropriate to their own back office systems and web content management systems, or even by wrote the HTML by hand.

The following authorities took part in the project:

- Barnet
- Cheltenham BC
- Coventry City
- Eden DC
- East Riding of Yorkshire
- East Staffordshire BC
- Gosport BC
- Lambeth
- Lichfield DC
- Lincoln City
- Manchester City
- Portsmouth City
- Redbridge
- Richmond-upon-Thames
- Stratford-on-Avon DC
- South Cambridgeshire DC
- Southampton City
- Southwark
- Trafford MBC
- Warwickshire CC
- Wyre BC

Socitm and OpenlyLocal are continuing to encourage councils to re-publish results of the 2010 and previous elections as open data and set things up in their systems and on their websites so that future results will be published in this way.

Lessons from the Open Election Data project

- 1) **There is a lack of 'corporate' awareness/understanding of open data issues, and this will inhibit take up of open, linked data publishing unless it is addressed**
For many of the councils the project came into contact with, response to the initiative was by an individual, often the web manager/master working on their own initiative, rather than as a result of any corporate interest from their councils in open data.
- 2) **There is a lack of even basic web skills at some councils**
Worryingly there were councils where no-one had even basic web-authoring skills (i.e. a good understanding of HTML), being merely relegated to fill in forms in an (outsourced) content management system. In a world where the web is becoming the main method of communication with citizens and between bodies, this is not unlike having a finance department with no-one who understands the core rules of accountancy. Without those basic skills as a foundation there is no way a body can hope to produce linked data.
- 3) **Many councils lack web publishing resources, never mind the resources to implement open, linked data publishing**
Many council webteams are significantly under-resourced, eg consisting of one person, possibly part-time, sometimes with £0 annual budget, meaning any

expenditure has to be bid for and a formal business case made. This stifles innovation (the driving force of the web), leads to websites that become increasingly out-of-date, and makes participation in such a project problematic (although there was no direct cost involved in participation, it did require an investment in time).

The open, linked data issue may not, therefore, come high on web managers' priority list, unless publishing in open data format becomes a requirement on councils.

4) **The understanding of even the basics of linked data and the steps to publishing public data in this way is very, very limited**

Publishing information as machine-readable data is not new, but the standards and methods for publishing it on the web so that anyone can consume it have only recently moved out from academic circles.

Because of this, even those experience with core web technologies have little familiarity with it, and the core principles behind it (using web addresses, or resource URIs, as identifiers for objects and relationships).

In addition, there are few tools for the publishing of linked data, meaning much has to be done by hand and checked manually, which is time-consuming and error-prone. (This is similar to the early days of the web, before the creation of content management systems and web-authoring tools.)

The **Open Election Data Project** has increased knowledge in local government of this by an order of magnitude, as well as providing concrete examples of publishing of open data by councils and establishing an informal network of people who have the knowledge and are willing to pass it on.

5) **The tools for and knowledge of consuming linked data are also limited**

Although linked data exposes data in a richer form, consuming it is not as well-known or straightforward as other formats (e.g. XML, JSON) requiring a greater investment in time and knowledge by data users. While it is appropriate for election data (where it is critical that electoral areas and parties are properly identified), it may not be as essential for all data sets, and should not necessarily be the first format supported, as it may delay publication and consumption.

6) **Publishing RDFa conflicted with some existing setups**

A common query councils had was validation errors caused by a conflict between the RDFa and HTML Doctype (which defines the version of HTML), either because councils were unable to change the Doctype, or because their existing code would break if they did. Although the validation errors were insignificant (it would not affect either browsers or screenreaders), some councils used their strict policies about validation as reasons for not taking part (even though there were invariably validation errors on other pages).

Another problem (for data users, rather than publishers) was where the HTML included the base url, and there were also problems when consuming data that had line breaks within the RDFa.

7) **Getting councils to publish data in open, linked formats will require dedication of a range of resources and application of a mix of skills**

The success of the Open Election Data Project, through which some 20 councils successfully published results as open, linked data at minimal cost to them required the following resources:

- A project leader, in this case, OpenlyLocal (which also funded the project), and endorsement and in kind support from 'recognised' authorities, in this case, LGA and Socitm.

- Access to 'official' information resources (e.g. The Electoral Commission, lined data experts at the Cabinet Office) to ensure the subject was properly understood and represented in the data.
- communications skills to explain the issues and benefits and identify appropriate channels to persuade the necessary people that their organisation should participate in the project
- communications resources including a 'campaign banner' and an interactive communications 'hub' for the project (*see below*)
- research and publication of resources required for implementation by webmasters, eg resource URIs for political parties
- technical expertise around web coding and mark-up and standards and resource URIs and the resource to deploy this as advice to councils

8) **Councils need a dedicated, interactive, information and advice source**

Web managers/developers in councils motivated to participate in the project were able to go to the **Open Election Data** website (<http://openelectiondata.org/>) and access a page called 'How to mark up your election data'. This contained step-by-step instructions (*see below*) with links to sample code and real examples of what the end result should look like on their website.

How to mark up your election data as linked, open data

1. Mark up the Election Results landing page using the code on the [sample landing page](#) as an example. You need to link to each election you are going to mark up as data, and annotate those links as shown
2. Mark up those Individual Election Result pages you want to expose as linked data using the [sample Individual Election Result page](#) as an example.....
3. etc etc

Also provided through the website was a list of all the official parties together with their resource URIs, and a list of poll and ward URIs for every council:

Resource ids for Birmingham City Council election

- Local Authority Birmingham City Council
- Election date May 6 2010
- Position Council member

Poll Resource URIs for this election

These are the wards (electoral areas) for this Local Authority and also the resource URIs to use in on the Ward Result pages (code example)

- Acocks Green (ONS SNAC id: 00CNGS)
Poll resource URI: <http://openelectiondata.org/id/polls/00CNGS/2010-05-06/member>
- Ward resource URI: <http://statistics.data.gov.uk/id/local-authority-ward/00CNGS> ...etc

Using these resource URIs enables results to be published as linked data and therefore be machine read. This is essential for automating the process of collecting the election results from different authorities' websites to compile a national database.

<http://openelectiondata.org/> lists more than 400 different names for parties involved in the local elections, and ward details for the UK's 433 councils – in total of over 12,000 pieces of information. All this information was provided through an easy look-up facility, and was driven by OpenlyLocal's existing database, making what could have been a tricky and time consuming task trivial.

<http://openelectiondata.org/> also provided a service to councils in checking their code before publication, using the w3c Distiller and Validator. However, even with these tools, which check the validity of markup and shows it in an easier-to-check format, there were still many mistakes and errors.

The exercise demonstrated that, even with detailed instructions, and significant motivation on the part of webmasters, marking up web pages was not always straightforward, and external help was needed to get things right.

The Open Election Data Community in the IDeA CoP was also a source of technical help to webmasters, enabling them to compare notes on issues and ask for, and receive help from others who'd overcome similar problems with coding or working supplier products like CM systems or election management systems.

9) Publishing each type of data will require appropriate resource URIs if data is to be published as open, linked data, and furthermore require councils to be able to use those URIs

Every data set to be published as linked data – be it expenditure data or information about recycling facilities or numbers of free school meals – will need to have a appropriate resource URIs agreed, published and communicated across local authorities. It is worth noting that one of the barriers the Open Election Data website solved was that councils in general had no idea of the Office of National Statistics SNAC ids used to identify the councils and the wards, and therefore needed to be told them by the website. It is to be expected that similar knowledge gaps will exist in other areas.

10) Many content management systems are inflexible and made the adding of RDFa to HTML at best problematic at best, and sometimes completely impossible

One of the biggest problems local authorities faced with publishing their election data as linked data was that their content management systems did not support this, and at times made this impossible. One council which had recently bought an expensive new proprietary content management system spent many hours trying to add RDFa markup to the HTML, and eventually gave up.

On the other hand, other content management systems used this as an opportunity to show the flexibility of their systems, as the 'plug-in' created for the Jadu CMS by one of its customers (Lichfield District Council) in connection with the Open Election Data project illustrates.

Local authorities that have or acquire content management products that do not allow use of RDFa and XHTML will find it very difficult to meet the Government's goal of publishing data in open, linked formats wherever possible.

11) Upgrading council software systems to generate 'open linked data' automatically may be possible but will not always be straightforward

During the Open Elections Data Project, the supplier of the Modern.gov committee/democratic data publishing system agreed to modify the system so that it would publish linked data (it published this as a web service using the XML version of RDF rather than RDFa). As a result, data published on the website directly from this system would not need further modification for it to be published as open, linked data.

However, even this brought its own issues, as the existing data in the Modern.gov systems didn't include standard ward identifiers and some councils were unable to expose the web service due to the setup of their firewalls.

12) Outsourced web publishing contracts often make it difficult and expensive to publish open, linked data

A common comment from councils wanting to participate was "We'd like to do this, but we've outsourced our web systems and it will cost us £XXX just to pick up the

phone to our supplier.” Even aside from the specific issues of the Open Election Data project, with the web developing so quickly this sort of outsourcing agreement doesn't seem capable of keep pace with those changes.

In addition, it also became clear that many outsourcing suppliers had no more knowledge about linked data than the average council. This is clearly a problem that goes beyond the requirement for open data and is about the wider fitness for purpose of websites commissioned from external suppliers, and the need for appropriate website commissioning skills within councils.

Add stuff re validation rules

Chris Taggart, OpenlyLocal.com, countculture@gmail.com
Vicky Sargent, Socitm, vicky.sargent@socitm.net