

Help! Where next for sustainable statistics and geographies for Wales?

Bill Oates

Knowledge and Analytical Services, Welsh Government

bill.oates@wales.gsi.gov.uk

+44 (0)2920 823820

09th January 2014

Problem Statement

Creating a road-map for statistical and geographical publishing, that:

- A. Is evolution rather than revolution for the in-house publishing team's skill-base;
- B. Can be integrated into mature workflows of the data producers;
- C. Leverages existing investments during their natural lifecycle;
- D. Capitalises on the increasing opportunities to move up the data publishing value chain;
- E. Aims for a combined platform for both statistical and geographical datasets.

Background – the team

The Geography and Technology (G&T) division is one of the cornerstones of Welsh Government's Knowledge and Analytical Services. Our responsibilities include back-office data management, analytical support and online publication of the statistical and geographical datasets of Welsh Government. The team has 5 developers / DBAs plus assorted project managers and support staff. Significant constraints are in place on headcount although cash budgets retain flexibility. The core strengths of the team lie in the Microsoft stack: SQL Server and .NET, with growing Javascript front-end skills.

This team provides their services to the 150 internal statisticians and social researchers as well as geographic information specialists right around the organisation and in external data bodies in Wales. The general model is to create technical publishing infrastructure which can then be used on a self-service basis by the wider professional community.

Current publishing activities

There are three main strands to the data publishing activities:

1. Statswales – statistical data cube publishing, also exposed as XML;
2. Map services and front-end publishing – online spatial data repository, exposed as WxS, with multiple topic-specific 'apps' built for different government needs;
3. Public Service Performance Portals – a growing portfolio of data driven sites that communicate government performance data in key sectors (health, education) to a lay audience;

These publishing resources are well patronised, with a user-base of between 5,000 – 10,000 views requests per month for the websites and online presentations. There is however, no current demand for the underlying data services.

Business Need

A “data-publishing as a service” capability that allows statisticians and other data production colleagues to make their data available on a publish-once, use many approach.

Candidate technologies

Over the last 2 years, the team have transitioned many of the publishing resources onto a cloud-based delivery mechanism, utilising both IaaS and PaaS offerings. Approaches currently in use include:

- A. 100% Javascript front-end driven by cached nuggets of JSON and tiled maps. These apps can be served from a low-cost, highly-performant CDNs. JSON data generated by queries off back-end databases with manual upload and refresh;
- B. Open-source WxS software stack hosted on Amazon or Azure IaaS. Scripted back-end to facilitate data loading and conversion;
- C. MS SQL Server Analysis Services for statistical cube publishing, with a bespoke back-end for cube preparation and transfer to the cloud;

In 2013 there has been a rise of 3rd party data publishing as a service opportunities that may also hold promise. Two particular candidates include Google Maps Engine (with associated public maps programme) and the ESRI ArcGIS Online hosted data layers solution, which are available at no cost to Welsh Government. Both of these offer the opportunity to have our business data hosted – under our control – on an architecture that promises significant scalability and availability to us, controlled communities and the general public through a number of accessible and well documented APIs and services. These include OGC compliant services.

Similarly, Google offer a public data hosting programme for statistical datasets that again offer zero-cost, scalable and performant open publishing for the cubes of data published each year.

The Challenge of Linked Data

In 2012, Welsh Government experimented with Linked Data through a contract to publish the Welsh Index of Multiple Deprivation. This project demonstrated an attractive and usable front-end for the dataset that integrated data and maps. It also however raised some significant challenges for the team in terms of the paradigm shift required for the technology platform and the data structures. At the time, the toolset and the required staff skills were not available in sufficient strength to pursue this within the current resource base.

Furthermore, a key learning was that to properly realise the potential of Linked Data meant not just applying it as a thin veneer on the top of existing (largely RDBMS-

based) working practices but to integrate it completely into operational delivery. This meant not just re-training of the G&T team but also exploring how Linked Data can be integrated into the practices of the wider data creation teams and the 'data publishing as a service' functions.

The roadmap

It is appreciated that there is no magic bullet approach, but advice and input would be welcome on how this small-medium sized publishing entity can further develop its data publishing activities to maximise the benefits to itself and to its customers.

Current online resources

<http://statswales.wales.gov.uk>

<http://mylocalschool.wales.gov.uk>

<http://mylocalhealthservice.wales.gov.uk>

<http://lle.wales.gov.uk>

- See the 'apps' section on this site for further front-end resources