

Merging statistics and geospatial information, 2012 projects - the United Kingdom

Statistics Explained

This article forms part of Eurostat's statistical report on *Merging statistics and geospatial information: 2019 edition*

Final report January 2015

Problem

Statistics are measures of things, so every statistic has a geographic value. The Office for National Statistics (ONS) decided to investigate the potential use of 'linked data' as a mechanism for publishing statistical geographies and linking them to statistical data.

Objectives

Formally, this project had five key objectives, namely:

- to establish what difficulties existed with publishing statistical geographical information (for locations or areas) as geo-linked data;
- to make recommendations on best practices for publishing statistical geographies as linked data;
- to set-up a system for disseminating statistical geographies as linked data;
- to provide value added functionality through the development of a human-readable interface to the linked data;
- to investigate the potential issues with publishing statistics that link to statistical geographical data to other types of related linked information.

Method

The provision of technical infrastructure was outsourced using an existing platform called OpenUp, which provides a Resource Description Framework (RDF) — a specification for metadata.

In the United Kingdom there was already a single unique identifier for each geography, known as a GSS code, comprising nine digits. An early step in the project was to use these codes to develop a structure for unique resource identifiers (URIs) that are HTTP encoded. As part of the guidelines of the UK Location Programme it was decided to use data.gov.uk as the domain. For spatial things (real-world phenomena that have a spatial extent or position) the word statistics was added as a theme while for spatial objects (abstractions of spatial things) the word location was added as a theme.

For the development of URIs for statistical geography on spatial things, the following structure was adopted:

http encoding	theme	domain	spatial type	class	identifier
http://	statistics.	data.gov.uk	id/	statistical-geography	{GSS code}

For *spatial objects* the structure was linked to INSPIRE requirements and so included an element for the INSPIRE theme for statistical units. It also included geometry as a class to distinguish these objects from points or lines. This resulted in the following structure:

Having developed a structure for the URIs the main work was to consolidate data from a range of ONS products — using geography as a common linking element — in order to publish data online using a format based on a resource description framework (RDF). The preparatory work involved identifying which datasets would/could be converted alongside creating a data model for each one (using TopBraid Composer), which in turn led to a specific data vocabulary to describe all of the elements of the dataset in question. Scripts (in Python, which is open source) were then developed to use the data vocabularies and models to convert data from a basic format (*.csv files) to the RDF format. The resulting data were then loaded through an open data portal to the OpenUp platform, from which they were made available to end-users. The final step was to put a set of tools in place so that people (as opposed to machines) could find the information they sought.

ONS branding was added to an existing interface that could bridge the gap between the linked data and the query language (SPARQL): effectively the interface provided human readable tools for generating queries in SPARQL (without the user having to understand SPARQL) and then exported data in various formats such as *.xml or *.csv.

A visualisation tool called the Explore tool was also developed. This displayed selected geographies on a base mapping layer. Users could enter a geographic name or postcode (or select names from a list) and then be provided with all of the information available for that geography. A Locate tool was also developed. With this, users could identify an area of a map (a box with a specific boundary) in the form of a rectangle or freehand polygon with information for various geographies within the selected area.

http encoding	theme	domain	spatial type	INSPIRE theme	class	identifier	resolution
http://	statistics.	data.gov.uk	so/	su/	geometry/	{GSS code}	{resolution code}

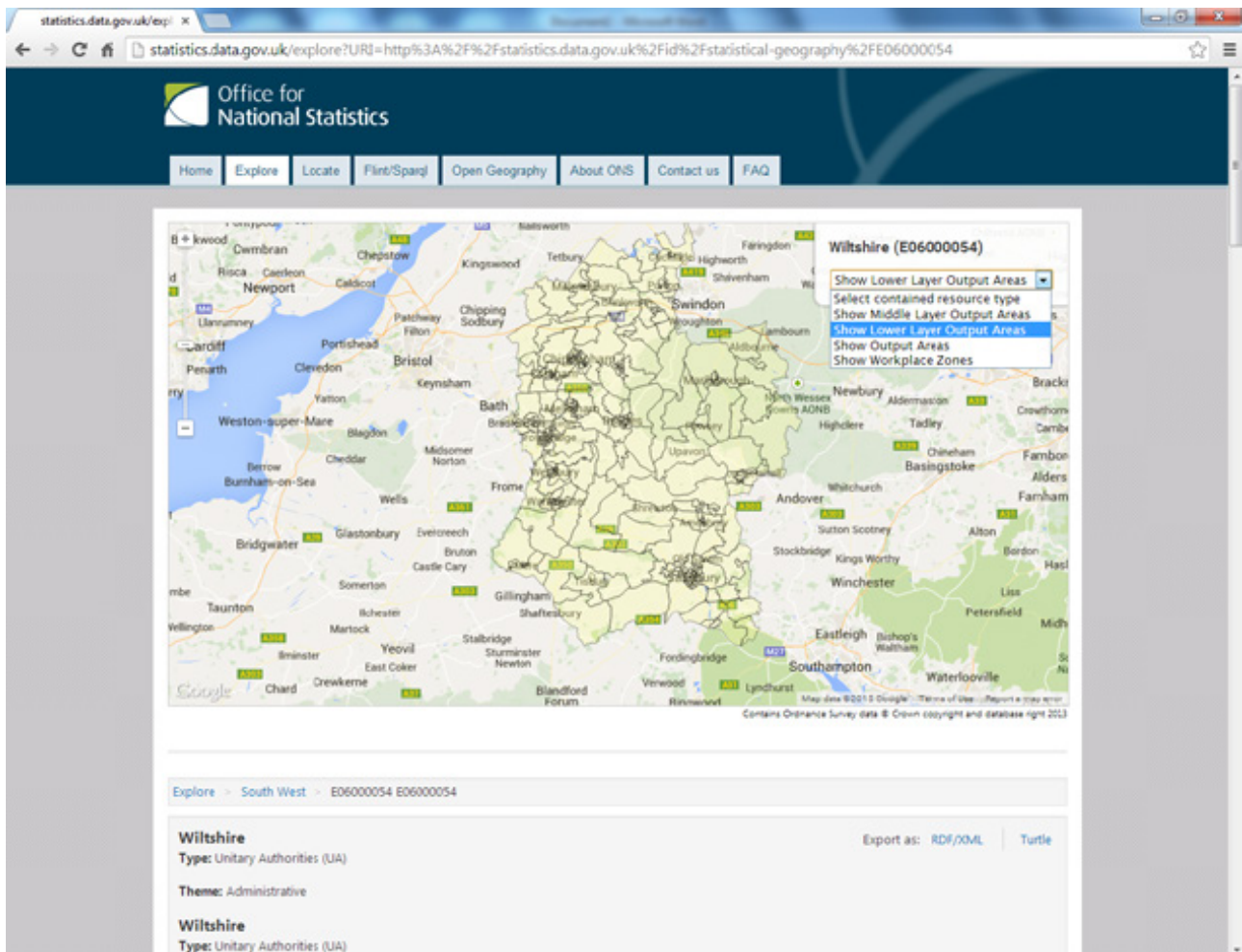


Figure 1. Drop-down menu for selecting a nested hierarchy

Results

A URI was developed for the NSOs geostatistics.

The various products that were published were: a Code History Database, the Register of Geographic Codes, the National Statistics Postcode Lookup and geographical boundaries.

In addition, user tools were developed facilitating not just automatic (machine) interaction with the data, but also facilitating data use by individual people.

Once completed, the system moved from being a special project to being integrated within the normal working environment of the NSO which, among other things, involved developing a sustainable maintenance process and developing skills within the organisation.

Having completed this stage of the development work, the focus subsequently moved on to linking more statistics to the geographic data so that a wider range of statistics was made available for any particular geography. The statistics disseminated through the tools developed for linking data were also disseminated through traditional products. In the future these tools might be discontinued as the data for different geographies migrate to a single data dissemination platform.

Linked data is a feasible format for delivering the requirements of INSPIRE. Providing data in RDF format makes it possible for data from several disparate sources to be connected together and delivered through a single user application.

Methodology

- [Final report](#)
- [ONS geography linked data site](#)
- [ONS open geography portal](#)
- Use of open standards:
 - [SPARQL 1.1](#)
 - [Turtle — terse RDF triple language](#)
 - [RDF](#)
 - [URI sets for the public sector](#)

External links

- [Office for National Statistics \(ONS\)](#)