Exporting as RDF is not enough

Alexander Coley (alex.coley@defra.gsi.gov.uk)
Department for Environment, Food and Rural Affairs (Defra), London, UK
and Chair of the UK Government Linked Data Working Group (UKGovLD)

The huge potential of Linked Data is that it will help break down many of the current information silos and add rich and diverse data directly onto the web, gaining and benefiting from the infrastructure of the web itself.

We are seeing many examples of this across many topic areas, countries and organisations. In the public sector in the UK we have examples of publishing reference data, environmental observations, legislation, statistics, geographies, community data and much, much more through linked data approaches. Some of this data is hugely valuable to certain communities and powerful when combined with other data.

When as a community we talk about the benefits of Linked Data and areas that it can help improve the effectiveness or efficiency of the Public Sector in the UK a common response is that of "yes, but" followed by "I don't like.." or "I don't need Linked Data". In the geospatial community this has been especially common. However, when you talk about this in more detail the discussion moves though a real need to see the benefits of Linked Data but negative experiences of early provision of RDF.

A number of early examples of Linked Data services have been static (almost never updated) data with data available in RDF and possibly with a SPARQL endpoint. The learning curve is huge and the available tools to reduce that step are limited.

Examples of the integration of spatial data are similarly of differing approaches and the learning curve maybe even greater.

There are other ways of doing Linked Data that have demonstrated many of the benefits while also adding other options that reduce the negatives.

For example, many of the current UK Public Sector Linked Data services are provided through api's that enable multiple ways to get the that that you want in multiple formats. This reduces the burden while packaging the data in a way that meet multiple needs. There are examples with simple api access and others with more complex access options. Example data explorer style applications help start people off with the data.

However, there are not yet common practices for doing this. There are gaps and there are conventions of practice to establish:

- What is the minimum that you would expect from a data service
- What sort of data should be provided through these services

- What types of functionality should we expect
- What are the api choices
- How should URIs be patterned and structured

Some of these are beginning to have conventions of approach form while others remain very much still as gaps. In the more geospatial side of requirements there are gaps such as:

- How should geometry be provided
- How are more advanced data options exposed
  - for example: complex geography vs simple representations or large collections of observations
- What geospatial api choices are there
- and more

There are mismatches in standards across a number of areas. These problems are probably more notable in the geospatial community where the simple representation versus availability of the detailed geometry for complex geospatial analysis poses conflicting requirements. There have been examples of differing approaches to this problem but nothing near a common practice.

We have a fairly well developed set of experiences of publishing data traditionally these have not yet developed for publishing the data in way that many of us want to see. These questions, their answers and more will help build some conventions that make that task easier.

There may be simple steps that we can take to make data more accessible. In particular about making aspects available across specialisms within organisations.

What is good enough in short term to make data available to the web of data and what is needed in longer term to get wider benefits?