

Linked Data Profiles

Paul Davidson, CIO Sedgemoor District Council, UK

Revision 2, 25th June 2013

@paul_davidson

Director of Standards for the UK Local eGovernment Standards Body (LeGSB)

<http://legsb.inetwork.org.uk>

In considering how to assist Local Authorities to publish their information as linked data, it has become apparent that we lack a machine-readable way of expressing the 'profile' of a dataset; or in other words, which classes, properties, URI Sets, SKOS vocabularies etc have been used in a dataset.

To illustrate the problem, the 'org' ontology, described at <http://www.w3.org/TR/vocab-org/>, says ...

Users of the ontology are encouraged to define profiles which strengthen interoperability by specifying particular controlled vocabularies to use for these concepts.

It goes on to encourage defining profiles to describe how 'sub classes' of classes defined in the ontology, have been used.

For example, a Local Authority therefore may describe itself, as

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix ops: <http://data.openpublicdata.com/ops# > .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix org: <http://www.w3.org/ns/org#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<http://data.sedgemoor.gov.uk/id/council/organisation/1>
  a ops:LocalAuthority ;
  skos:prefLabel "Sedgemoor District Council" ;
  org:classification <http://opendatacommunities.org/def/local-government/DistrictCouncil> ;
  dcterms:description "Sedgemoor District Council is a District Council in the County of Somerset, UK" ;
  owl:sameAs <http://opendatacommunities.org/doc/district-council/sedgemoor> ;
.
```

, and be compliant to the 'org' ontology, however, the ontology does not define which properties should be used. For example, whereas here I have used skos:prefLabel to name the council, another council may use dcterms:title, which would be equally valid. The same applies to all of the properties in the above example, and the ops:LocalAuthority class which is a sub-class of org:FormalOrganization.

So how do we communicate that these classes, properties, URI Sets, have been used, and encourage others to use the same 'profile' to describe similar data, so that they can be confidently combined.

This is particularly relevant to Local Authorities where hundreds of local councils and agencies each provide a subset of the same services, and will therefore have the same type of data to publish.

Whilst OWL has the capability to express how a dataset has been constrained, it feels too complicated for this simple purpose.

I see two requirements:

1	A publishable 'Linked Data Profile'. To define how a dataset has used classes, properties, uri-sets etc, from many
---	---

	<p>ontologies/sources.</p> <p>To invite others with similar data to use the same profile, and therefore find that their data can be queried together.</p> <p>To be able to register a dataset so that it can be discovered by its 'Profile'.</p> <p>To be able to 'query', and present a dataset (or across many datasets) by referring to its 'profile'</p> <p>To be able to validate that a dataset does conform to a 'profile'</p> <p>To enable others to create a new profile by adapting an existing one.</p> <p>To be able to create a register of good practice 'patterns' for commonly recurring data structures, that can then be re-used.</p> <p>To be able to define the Classes and Properties that are returned when dereferencing a URI.</p>
2	<p>A publishable Data Entry specification</p> <p>To define a configuration sufficient to enable data entry and data import, to transform a dataset to conform to a published profile.</p> <p>To encourage a market for tools and services that can enter/import/transform data to avoid proprietary lock-in.</p>

I have set up a sandbox at <http://www.openpublicdata.com> to explore how data from many sources can usefully be queried, if they conform to a common 'profile'.