

Vocabulary Application Profiles: Bridging Islands of Interoperability

Submitted on 21 Sep 2015 by Martin Alvarez-Espinar

An important underlying principle of the open data initiative is that it should be made easier for organisations to share data and create a network effect by which efforts invested by one organisation can create benefits for others. **The increased use of common vocabularies and application profiles will contribute to realising the potential of such network effects.**



In an open world there is a need for basic agreements on how to describe information. This drives the development of base standards, such as [Dublin Core](#), [FOAF](#) and [SKOS](#) to provide a vocabulary of commonly understood descriptors. This is a bit like a dictionary that defines what terms mean, also called the semantics.

In practice, it turns out that this is not enough. Just like a grammar book is used in conjunction with a dictionary to describe how words are combined to create understandable sentences, **Application Profiles** give the rules and recommendations regarding the way the terms from the semantic standards can be used and combined to provide useful services. An example is the [DCAT Application Profile for Data Portals in Europe](#) (DCAT-AP) that defines the way the base W3C standard [DCAT](#)—which itself is based on the more general Dublin Core standard—should be used to create a network of interoperable open data portals.

Both types of initiatives, development of base standards and definition of Application Profiles, share a challenge which is that there are many base standards and many potential service environments. To see an example of similar, overlapping and sometimes competing sets of vocabularies, go to Linked Open Vocabularies ([LOV](#)) which contains [over 500 different vocabularies](#). Searching for properties or classes that are related to names of things, LOV returns 1994 results. The **proliferation of vocabularies and Application Profiles makes it hard for people to find the specification that**

fits their needs.

As a result, we see **'islands of interoperability'**, consisting of groups of organisations that share a common goal in a given environment. This is what drove the development of the European DCAT-AP among organisations that wanted to exchange data and to contribute descriptions of their datasets in a central **European Open Data Portal**.

Based on adoption to support the cross-border application, national initiatives across Europe are also adopting application profiles for national services that are adapted and extended versions of the DCAT-AP. Even if the national extensions may not be understandable between countries, the use of a common 'core' application profile makes interoperability across borders easier because the application profiles share the same semantics, or use the same 'dictionary' in the analogy of human language. Others may not all the words of a different dialect but the main message can be successfully conveyed.

An important underlying principle of the open data initiative is that it should be made easier for organisations to share data and create a network effect by which efforts invested by one organisation can create benefits for others. **The increased use of common vocabularies and application profiles will contribute to realising the potential of such network effects.**

Revision of DCAT and DCAT Application Profile

Recently, the DCAT-AP specification [was reviewed](#) and updated with comments from the community. Technically, the revision of the DCAT-AP led to a number of discussions about what the vocabulary does and does not offer. For example, an obvious aspect that is **missing from core DCAT is any support for versions** of (abstract) datasets and their (concrete) distributions. In other words, there's no way to say that various datasets are a time series –how many edits can you make to a dataset before it becomes a new version?

Due to the flexibility of the Semantic Web, it's possible to make statements by using Dublin Core terms (i.e., '[replaces](#)', or [isVersionOf](#)), or any other suitable vocabulary such as the [Provenance, Authoring and Versioning vocabulary](#) (PAV), but these are not part of the core DCAT.

Should they be?

Could they be?

The answer to the first of those questions is for the community to decide. The answer to the second is yes. The [W3C Community Group](#) process allows people to come together and discuss topics around standards. Anyone may participate in these groups that don't create formal standards (Recommendations) but a Community Group could be formed to maintain and improve DCAT over time. Some of the things it might want to do are:

- add those demanded new terms;
- add new translations of the labels and comments, adding to the existing labels in English, French, Spanish, Japanese, Greek and Arabic;
- clarify existing terms;
- provide examples of DCAT in use.

The one thing it wouldn't be able to do is to delete terms or to break the existing model but extending it would be possible, all in the same namespace. This extension may work in the same way as the DCAT-AP extended the core version of DCAT.

But there is no magic. For this to happen, the group needs a sufficient number of people interested

and committed enough to work on it. If you're interested in pursuing this, just [contact Phil](#).

Written by [Makx Dekkers](#), [Phil Archer](#) and [Martin Alvarez](#) for ePSI Platform

Photo credit: [fab_a_paris](#)