Timing

This is an early task that should be started prior to forming the Working Groups, but that is likely to require several iterations to agree on a suitable hierarchy.

4.2 RESEARCH AND REVIEW EXISTING SOLUTIONS

The generic entities that Core Vocabularies describe — people, organisations, publications and so on — are likely to be covered by existing vocabularies but there are other considerations that apply specifically to a *Core Vocabulary*. What do community members use at the moment? Have they already identified vocabularies as being relevant? For a given vocabulary one can ask three key questions:

- does it encompass the required scope?
- is it stable and/or subject to a formal change process?
- is it already widely used?

If an existing vocabulary passes all these tests then one can achieve interoperability simply by using it. The following non-exhaustive list shows vocabularies that pass all three tests and are relevant to e-Government.

- The Dublin Core Metadata Initiative (DCMI) Metadata Terms vocabulary defines general metadata attributes such as title, creator, date and subject [DC].
- The Friend-of-a-Friend (FOAF) vocabulary defines terms for describing people, their activities and their relations to other people and objects [FOAF].
- The Description of a Project (DOAP) vocabulary (pronounced "dope") defines terms for describing software projects, particularly those that are Open Source [DOAP].
- The Good Relations Ontology defines terms for describing products, services and other aspects relevant to e-commerce applications [GR].
- The Creative Commons (CC) schema defines terms for describing copyright licences [CC].
- The Bibliographic Ontology (BIBO) provides concepts and properties for describing citations and bibliographic references (i.e., quotes, books, articles, etc.) [BIBO].
- The OAI Object Reuse and Exchange vocabulary is used by various library and publication data sources to represent resource aggregations such as different editions of a document or its internal structure [ORE].
- The Basic Geo (WGS84) vocabulary defines terms such as lat and long for describing geographically-located things [WGS84].
- The vCard vocabulary underpins many electronic business card systems [vCARD];
- The INSPIRE Directive [INSPIRE]
- The European Business Register [EBR]
- Open Corporates [OC]

(List derived from Linked Data: Evolving the Web into a Global Data Space [LD]).

Re-inventing terms from these vocabularies will *reduce*, not increase, interoperability between different data publishers. It should be noted that any list in a document such as this will become out of date over time and should not be seen as definitive or complete.

The UN Centre for Trade Facilitation and Electronic Business [UN/CEFACT] maintains a Core Component Library (CCL) - a highly sophisticated collection of basic core components and aggregate core components. The approach taken is the opposite of the minimalist approach set out for the development of Core Vocabularies in that the CCL provides an exhaustive set of terms, for example:

UN00003667 Accompanying Person. Details
UN00003668 Accompanying Person. Relationship To Patient. Code

The CCL does, however, provide an internationally agreed set of terms, with identifiers. Where a Core Vocabulary developed under this process matches a term in the CCL, a direct reference should be made. The Simple Knowledge Organization System [SKOS] provides a number of suitable relationships for this purpose.

If a stable vocabulary in widespread use covers some, but not all, of a domain then there is work to do in defining *additional* terms so that taken together, the base vocabulary (or vocabularies), and the new terms, are sufficient for the task. For example, the FOAF vocabulary includes terms for a person's full name, given name and family name. It does not include terms for a patronymic name, or multiple family names (both common features of names in Europe) [PNAW], nor does it include a term for how one should address an individual (as does, for example, the UK Parliament Web site [UKP]). Likewise, if existing vocabularies cover more than required by the minimalist approach emphasised in this document then a *subset* of those existing terms might form the Core Vocabulary.

This highlights an important difference between a general vocabulary and a Core Vocabulary. In general vocabulary development it is common to define a term as a specialisation of an existing one. For example, the Dublin Core vocabulary includes the term `identifier`, the value of which should be 'An unambiguous reference to the resource within a given context.' In the context of a library, the consensus within the community might be that a more useful term would be `catalogueNumber`. In a case like this, `catalogueNumber` might be a better term to use, however, it would be defined as a specialisation of the Dublin Core term (in RDF terms this would be a sub-property). In that way, systems that understand Dublin Core terms may still be able to consume and process the data, even if they cannot directly understand the `catalogueNumber` term.

Core Vocabularies are, by definition, not specialised. Like Dublin Core, they provide the super-classes that more specialised vocabularies can refer to for greater semantic interoperability.

It is possible that a vocabulary may be found that appears to fit the Working Group's needs perfectly but that does not appear to be in widespread use. In this case, the research task is to identify why this is so. Common reasons include:

- the vocabulary is the work of a small number of individuals who did not succeed in building a community around their effort;
- the vocabulary has not been promoted sufficiently;
- the vocabulary is subject to restrictive licence terms and is effectively unavailable for reuse;
- a more popular vocabulary is already in use that does a similar job.

In the first two cases, the authors may welcome an approach from a Working Group to stabilise and promote their work. In the third case, developers of open standards should be wary of infringing copyright (this highlights the importance of section 3.2.1). If the fourth case applies then this is a signal that the more popular vocabulary would likely be a better choice. This can be frustrating as it does inevitably lead to compromises being made, however, the aim is to make data interoperable and for that reason *large scale deployment trumps semantic correctness*.

Given the growing number of stable and well-used vocabularies, it is possible that a Working Group will be able to recommend that a concept is covered entirely by terms in existing vocabularies and not create any new terms at all. This is as much an acceptable, useful and successful outcome as one that defines a large number of new terms. Indeed, endorsement by the Member States may be more easily secured as interoperability and deployment will have largely been proven.

Objectives

To follow best practice by re-using existing, well-defined, well-used vocabularies, rather than re-inventing them.

Input

Existing vocabularies such as, but not limited, those listed above.

Tasks

Method Step 2. **The Working Group should research existing vocabularies, their provenance, usage and stability. This research should inform the writing of use cases (see section 4.4).**

Output

A list of terms from existing vocabularies that may be useful in the domain of the proposed Core Vocabulary.

Timing

This is an early task for the Working Group and one that should be carried out before any new terms are minted.