



D3.1 – PROCESS AND METHODOLOGY FOR CORE VOCABULARIES

Deliverable

JOINING UP GOVERNMENTS



This report / paper was prepared for the ISA programme by:

Author's name: PwC EU Services EESV

Company's name: PwC EU Services EESV

Disclaimer

The views expressed in this document are purely those of the writer and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.

All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

This paper can be downloaded from the ISA website:

<http://ec.europa.eu/isa>

TABLE OF CONTENTS

1. Introduction.....	5
1.1 Core Vocabularies	5
1.2 Context	5
1.3 Aims and Objectives.....	6
1.4 Scope.....	7
2. Stakeholder groups	8
2.1 The European Commission	8
2.2 The Members States of the European Union	8
2.3 Standards organisations.....	8
2.4 The global community	8
3. Specification Process.....	9
3.1 Roles.....	9
3.1.1 The European Commission.....	9
3.1.2 The Community	9
3.1.3 Working Group	10
3.1.4 Review Group.....	12
3.1.5 Endorsement Group	13
3.2 Process steps	15
3.2.1 Secure IPR	15
3.2.2 Establish a working environment and culture.....	16
3.2.3 Draft process	20
3.2.4 Review process	21
3.2.5 The Last Call Working Draft	22
3.2.6 Review Group Feedback.....	23
3.2.7 Review Last Call Working Draft.....	24
3.2.8 Gather evidence of actual or expected use.....	25
3.2.9 Endorsement process.....	26
3.3 Managing change	27
3.3.1 Editorial changes only	27
3.3.2 Vocabulary term changes.....	27
4. Methodology.....	28
4.1 Identify a meaningful set of Core Concepts	28
4.2 Research and review existing solutions	29
4.3 Research existing data and services.....	32
4.4 Use cases	32
4.5 Requirements	33
4.6 Terminology and conceptual data model	33
4.7 Naming conventions	34

4.7.1	Technical constraints	35
4.7.2	Conventional constraints	35
4.8	Identifier Conventions	36
4.9	The namespace document	38
4.10	Quality Assurance/Conformance Criteria	39
5.	Conformance Statement	40
6.	Specification Template	41
6.1	Title and Version Control	41
6.2	Abstract	41
6.3	Status of this document	41
6.3.1	Early Working Drafts	41
6.3.2	Last Call Working Draft	42
6.3.3	Final Draft	42
6.3.4	Final Published Version	43
6.4	The body of the document	43
6.5	HTML/CSS	43
6.6	Validation	44
Appendix A	Comparison of Existing Processes and Core Vocabularies	45
Appendix B	References	51
Appendix C	Terminology And Glossary	54
Appendix D	Summary of Process and Method Steps	57

1. INTRODUCTION

This document outlines the process and methodology used for the development of Core Vocabularies in the context of the SEMIC.EU project. The process and methodology steps are summarised at the end of the document in section Appendix D. Throughout the body of the document, several terms appear in bold type on their first use. These are defined in the section 'Terminology And Glossary', beginning on page 54. External references are denoted in square brackets and listed alphabetically beginning on page 51.

1.1 CORE VOCABULARIES

The **metadata** and **reference data** used in electronic public services across Europe most often has a very specific context. Attaining consensus on common metadata and reference data for these electronic services is a *critical* step towards semantic interoperability. Unfortunately, consensus building is hindered by the diverse cultural, multi-lingual, legal, and organisational contexts of these e-Government services. To alleviate this problem, consensus building should start at a higher level of abstraction that surpasses the contexts of individual electronic public services, and thus the cultural, lingual, legal, and organisational differences of individual countries. In particular, consensus can be more easily attained on the semantics of a *small* set of fundamental concepts, for which less divergent opinions exist [EGOV-CV]. These concepts are what we describe as Core Vocabularies.

A **Core Vocabulary** is a simplified, reusable, and extensible data model that captures the fundamental characteristics of an entity in a context-neutral fashion [EGOV-CV]. Well known examples of existing Core Vocabularies include the Dublin Core Metadata Set [DC]. Such Core Vocabularies are the starting point for agreeing on new semantic interoperability assets and defining mappings between existing assets. Semantic interoperability assets that map to or extend such Core Vocabularies are the *minimum required* to guarantee a level of cross-domain and cross-border interoperability that can be attained by public administrations.

1.2 CONTEXT

This document has been prepared in the context of Action 1.1 of the **ISA Programme: Methodologies for the development of semantic assets**, known as the SEMIC.EU project [A1.1]. The specific focus of this action is to support public administrations in Europe addressing semantic interoperability concerns related to the cross-border and/or cross-sector electronic interactions between public administrations that result from the implementation of EU policies and activities.

With the above-mentioned *minimalist approach* to of cross-domain and cross-border interoperability in mind, Action 1.1 of the ISA Programme comprises the following steps, scheduled between October 2011 and March 2012:

1. Raise awareness about the importance to agree on a small set of simplified, reusable, context-free Core Vocabularies [CVOG] as a fundamental semantic basis for interoperable electronic public services;
2. Specify a process and methodology for developing Core Vocabularies (this document);
3. Identify a list of 20-30 Core Vocabularies to be elaborated in the future and agree on two Core Vocabularies to be specified first;
4. Finalise and build consensus on the Core Person Vocabulary [CPV];
5. Encourage Member States to endorse and reuse the Core Vocabularies.

The involvement, active participation, and eventually endorsement and reuse by the ISA Member States are critical success factors for these Core Vocabularies.

1.3 AIMS AND OBJECTIVES

This document has two aims:

To define the process through which consensus can be reached among stakeholders and domain experts so that a vocabulary is recognised as meeting its design goals. Having achieved consensus, the final part of the process is endorsement of the Core Vocabulary by the Member States. We can break this aim into the following objectives:

- achieve consensus among stakeholders and domain experts that the vocabulary is an appropriate method of achieving interoperability between data sets;
- achieve the widest participation in the development of a new Core Vocabulary;
- achieve the widest possible review of the vocabulary during its development;
- endorsement by the Member States.

The process is set out in section 3.

To define the methodology through which a Core Vocabulary is specified. That is, the best practices for how terms should be selected from existing vocabularies or, where necessary, developed, encoded and presented to its intended audience. We can break this down into the following objectives:

- to define clearly the kind of problem the Core Vocabulary will solve;
- to make the overall structure of the Core vocabulary clear, indicating how existing vocabulary terms, and any new ones, work together to comprise the Core Vocabulary;
- to follow technical necessity and gain credibility, by following community convention when naming new terms;
- to follow best practice when publishing the Core Vocabulary.

Meeting this aim is the subject of section 4.

1.4 SCOPE

There are a number of underlying technologies that are well established for sharing data, notably XML and RDF. As section 4 (Methodology) describes, Core Vocabularies should be published in both of these machine-readable formats as well as human-readable formats (such as HTML).

Without prejudice towards other technologies, the process and methodologies set out in this document expressly encourage the use of Linked Data, an application of the Semantic Web for which RDF is the technological foundation. RDF offers better support for the extension of Core Vocabularies than XML. Documents such as *Best Practice Recipes for Publishing RDF Vocabularies* [RDFBP], as well as the work of the Linked Data movement [LDORG], are therefore relevant. However, the Core Vocabularies produced as a result of following this document will be as *technology-neutral* as possible.

Core Vocabularies developed through the process and methodology set out here will themselves be examples of linked data, or, in this case, linked metadata. They will re-use existing vocabulary terms where possible and terms from each other. As a result, the Core Vocabularies will form a graph of connected concepts with a common framework for their URI space.

2. STAKEHOLDER GROUPS

In the process of developing and maintaining Core Vocabularies, the following stakeholders can be identified

2.1 THE EUROPEAN COMMISSION

The need for effective interoperability is a central part of the Digital Agenda, one of the flagship initiatives in the Europe 2020 Strategy. It is in the interests of the Member States, and particularly the ISA Programme, to support the creation of an environment in which cross-border and cross-sector interoperability can be achieved. Fostering consensus among a representative cross section of the many stakeholders is a critical aspect and challenge for the European Commission.

2.2 THE MEMBERS STATES OF THE EUROPEAN UNION

The Member States, represented in the ISA Programme, have an interest in the re-use of shared vocabularies to facilitate cross-borders and cross-sector interactions between public administrations through a common semantics for electronic public services.

The Member States are involved, through participation of the representatives they appoint as members of Working Groups, in the development and consensus-building processes leading to endorsement of the Core Vocabularies.

2.3 STANDARDS ORGANISATIONS

Standards organisations act in the interests of their members to facilitate the emergence of vocabularies that are supported by community consensus. An important concern for these organisations is that the initiatives by the European Commission do not lead to a *fragmentation* of effort in the e-Government standardisation domain. Therefore, standards organisations will be involved as observers and potential contributors in the development of the Core Vocabularies, and may take the results as input in their formal processes.

2.4 THE GLOBAL COMMUNITY

The wider community of actors in the e-Government space has an interest in the examples set by the European Public Sector as it will allow them to build on practical examples developed in Europe. Some members of the global community will be involved as contributors to the development (as invited experts), while the wider community will get an opportunity to review the resulting specifications in a public comment period.

3. SPECIFICATION PROCESS

This chapter specifies the process through which consensus can be reached among stakeholders. It outlines the roles that the different actors in this process play and the process steps that need to be taken.

3.1 ROLES

The goal of the development process is to create, review and, finally, agree and endorse a new specification. These different steps are executed by distinct groups.

3.1.1 The European Commission

The European Commission is the owner of the overall process and of the derived results. In the process, it is assumed that the European Commission (EC) will provide the strategic directions for the work together with staffing, either through internal resources or via contractors, who will assemble and support Working Groups and Review Groups according the process set out below.

3.1.2 The Community

Composition

The EC and/or the **ISA Coordination Group (TIE Cluster)** will identify individual experts as being important members of the community around any Core Vocabulary. They are an important part of the community, however, there is more to consider. Technical interoperability is not in itself the desired result; rather, it is the means by which the desired result of people sharing information can be achieved. In order for this to happen, a broad community of interested parties needs to be brought together (see section 2).

For example, administrations already run a wide variety of data systems, publishing primary data resources with or without metadata. Such organisations may use **Semantic Interoperability Assets** but this is by no means certain. What is certain is that whatever system is used will already meet the immediate needs of the organisation itself. It is critical that people already working in the relevant domain and who wish to make their system interoperable with others are involved in the development of a vocabulary. It is these individuals who are the likely end users and beneficiaries of the new vocabulary so their support is highly desirable.

There are three areas in particular within which stakeholders may be identified, besides those in the ISA Coordination Group. Firstly the policy DGs of the EC, Secondly the Member States that are already working with experts in the fields in which the Core Vocabularies are to be developed. Finally, what can be described as the 'standardisation community' which itself encompasses industry experts, academics, researchers and others who already have an interest in the work and who are likely to want to continue to use the Core Vocabulary in future.

Tasks

Process Step 1. Identify Stakeholders. The EC identifies organisations and individuals that are likely to have the most need for, and benefit most from, the proposed Core Vocabulary. The EC invites them to participate in the process: ideally by joining the Working Group or at least by subscribing to the public mailing list and reviewing published drafts.

Implementation

The stakeholders will be identified based on a list of contact persons that is maintained by the subcontractor responsible for the semantic asset project. Each identified stakeholder will be contacted and asked whether or not he/she wishes to

- Actively participate and contribute as a member of the Working Group;
- Be kept up-to-date of further progress on the project.

Timing

The community must be built before a Working Group can be formed. The speed at which it builds will depend on the existing demand for and interest in the new Core vocabulary. The community that will define a Core vocabulary that does not exist and for which there is already a perceived demand will form more quickly than one centred on an area already well-served by existing vocabularies and/or for which there is little demand.

3.1.3 Working Group

Composition

The Working Group is the subset of the community that actually does the work. Membership of the Working Group requires a commitment to attend meetings and to make contributions to the activity. The group is expected to work remotely via various electronic communication channels. Exceptionally, a physical meeting may be arranged, perhaps to resolve a conflict. However, conflicts are unlikely since Working Groups are explicitly tasked with defining a minimum set of vocabulary terms that facilitate interoperability (see also Dispute resolution page 18).

Working Group members must be able to work without any conflict of interest and their participation must be approved by the EC or recommended by the ISA Coordination Group (TIE Cluster). Representatives of the Member States are entitled to membership in a Working Group and whilst representation of a Member State is not limited to a single place, the EC may limit the number of places taken by a single Member State if it is already well represented. There is no formal minimum number of Member States that must be included in a Working Group, however, it is clearly in the interests of all concerned that as many are represented as possible. The Member State representatives who participate in the Working Group are assumed to represent

the interest and point of view of their Member State, but also to represent other Member States who are not represented in the Working Group.

In addition to the representatives of the Member States, the EC may also appoint external experts to the group. EC staff, or their representatives, will attend Working Group meetings and will have the following roles:

- advising on the process and methodology set out in this document;
- acting as a liaison with the EC;
- acting as the permanent secretariat;
- additional roles to ensure the efficient achievement of the group's objectives.

Like any group of individuals, a Working Group will have its own dynamic with different levels of contribution and engagement. Central to the success of the group is the **chair** whose role is to:

- gather the group and maintain a positive and productive environment;
- organise and chair the meetings;
- distribute an agenda and reminder about each meeting in good time;
- ensure that minutes are taken and archived;
- ensure that decisions are recorded;
- guide the group through the development of the Core Vocabulary;
- ensure that consensus is achieved within the group;
- act as primary point of contact for the Working Group;
- ensure that the comments of reviewers are sought and acted upon when received, and that reviewers are content with the response of the Working Group.

This is clearly an important role and so the choice of chair is critical. It is sometimes possible for the chair's duties to be shared by two co-chairs although a higher number is rarely desirable or successful.

Another key role within the Working Group is that of **editor**, i.e. the person who actually documents the Core Vocabulary. Again, this task can be shared between individuals.

Tasks

Process Step 2. Form Working Group. The EC facilitates the formation of the Working Group and announces its formation to the community and the public at large.

Process Step 3. Identify Chair / Co-chairs. The EC identifies the chair or co-chairs.

Process Step 4. Identify Editor(s). The chair identifies the editor(s) of the Core Vocabulary document.

Implementation

The members of the Working Group will be selected out of the persons that replied to Process Step 1 by indicating they wish to actively contribute. Based on the responses, the subcontractor will create a list of suggested Working Group members and will identify the person most eligible to be the chair.

The person that is identified to be the chair will be contacted and asked whether or not he/she wishes to take responsibility for the working environment and culture that will be installed in Process Step 7. In case the response is negative, the subcontractor will consult with the EC and suggest another chair.

The members of the Working Group will be notified via e-mail of their selection, this message can coincide with the creation of each member's user account on the relevant platforms used by the project (see section 3.2.2, as described in the implementation of Process Step 7). The message will also clarify how the EC will handle IPR on any contributions made.

Timing

The formation of the Working Group marks the beginning of the formal process of defining the Core Vocabulary.

3.1.4 Review Group

Composition

During the development of the Core Vocabulary, the Working Group will make efforts to engage the wider community in reviewing its work. In particular, for the purpose of developing Core Vocabularies to be endorsed and reused by the Member States it is important that *all* Member States have the opportunity to review the Core Vocabulary specification. Therefore, a Review Group should be formed simultaneously with the Working Group. The Review Group comprises ISA Member State Representatives. Evidence of review by this group is mandatory for the progression of the Core Vocabulary. They should be given as much notice as possible of the actual date on which the Core Vocabulary will be available for them to review.

Tasks

Process Step 5. Form Review Group. The EC identifies members of the Review Group and advises them of the projected date by which the Core Vocabulary will be available for their review. If that date changes, then the Working Group chair, working through the EC, will advise the Review Group accordingly.

Timing

The Review Group should be established simultaneously with the Working Group.

3.1.5 Endorsement Group

Composition

The Endorsement group comprises the representatives of the ISA Coordination Group (or alternatively the ISA Trusted Information Exchange (TIE) Cluster Group). Appointment to this group is not part of the remit of the Working Group or the EC.

Tasks

The primary task of the Endorsement Group is *endorsement* of the final draft, as presented by the Review Group. It is important to note that there is no *legal* basis for endorsement. This means that the endorsement of a Core Vocabulary does not make it legally binding in a Member State. In particular, the Member States commit themselves to reuse and extend the Core Vocabularies in the context of their public sector data and electronic publishing services. This can be achieved by different governance approaches:

- incorporation of the Core Vocabularies into (national) legislation;
- inclusion of the Core Vocabularies on a comply-or-explain list of standards and specifications at the national level;
- incorporation of the Core Vocabularies in the national interoperability framework (NIF);
- encouragement of public administrations to reuse the semantic interoperability assets via promotional campaigns.

Timing

A Core vocabulary may only be submitted to the Endorsement Group once it has completed all previous stages (see next section). The EC will advise the Working Group of forthcoming dates when the Endorsement Group may be in a position to consider the proposal. Such dates must be considered as firm target dates for completion of the work by the Working Group and Review Group.

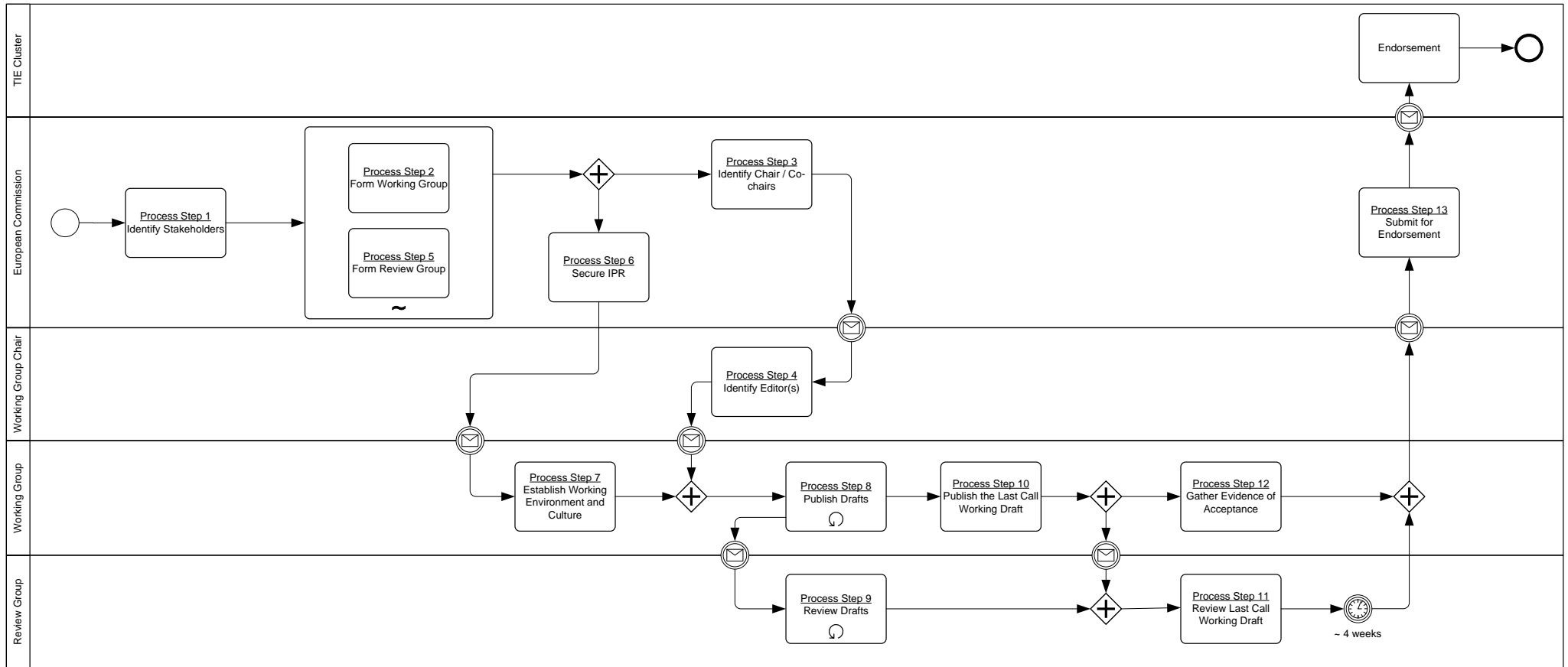


Figure 1 BPMN Diagram - Process for Core Vocabularies - Overview

3.2 PROCESS STEPS

The following sections set out the steps that must be taken by the various groups identified in section 3.1.

3.2.1 Secure IPR

As the discussion with the European Commission legal services about the IPR arrangement was still ongoing when this report was drafted, the current status of the discussion is presented in this section. Amendments to the approach may be needed based on the final recommendations by the legal experts.

Objectives

While the ownership and maintenance of the Core Vocabulary is with the European Union and European Commission respectively, the objective is to ensure that the Core Vocabulary is available for free use and redistribution with or without modifications by all stakeholders including the primary user group, namely the EU institutions that publish Public Sector Information and the organisations that will wish to consume such data. At the same time the link to the original Core Vocabulary must be maintained in all derivative works, and the European Commission should be protected against false statements of endorsement and any liability claims for damage resulting from the use of the Core Vocabulary. To achieve this, the Core Vocabulary and all documentation related to it will be published under an open licence that satisfies the above-mentioned requirements.

To grant the rights of free reuse and redistribution on the outgoing side, the European Union must acquire sufficient rights on the incoming side. For this the European Union must conclude a Contributor Licence Agreement [CLA] with all contributors (Working Group members and Review Group members) who have contributed intellectual property in the process of specifying a Core Vocabulary. This will allow defending the specification if a legal dispute arises concerning the contributed content. The EC will take steps to ensure that individuals and organisations invited to participate in the Working Group and Review Group are made aware of the licensing arrangements for the completed work. All contributors to the Core Vocabulary will be acknowledged, including their affiliation if applicable. Furthermore, the EC will ensure that sufficient evidence of agreement with the Contributor Licence Agreement is collected by asking the contributors to (electronically) sign the Contributor Licence Agreement. The EC will retain attribution of authorship as far as is practical.

Input

- Incoming rights: specific Contributor Licence Agreement [CLA] to be provided by the EC legal services;
- Outgoing rights: specific Licence for publication and redistribution with or without modification) to be provided by the EC legal services.

Tasks

Process Step 6. Secure IPR. On joining the Working Group or Review Group, the EC will advise members that contributions will be according to the Contributor Licence Agreement and that the final specification will be published under an open licence. Members are asked to (electronically) sign the Contributor Licence Agreement.

Implementation

As soon as the final licensing arrangements are in place, invitations to Working Group and Review Group members will include explicit details of the future ownership and licensing of the Core Vocabulary and the required Contributor Licence Agreement. This will also be highlighted in the first Working Group meeting. All Working Group and Review group members will be asked to (electronically) sign the Contributor Licence Agreement.

Output

As soon as the final licensing arrangements are in place, the Core Vocabulary will be published by the EC under an open licence. The EC will retain attribution of authorship as far as is practical.

Timing

At the formation of the Working Group.

3.2.2 Establish a working environment and culture

As the development of a new vocabulary is very much a group effort, it is essential that an effective working environment is setup and a collaborative culture is agreed upon.

Public communication

The group must have a means of communicating its work to the public, i.e. anyone who is interested in the development of the vocabulary. This requirement can be fulfilled minimally with a Web site (or section of an existing Web site) but is better achieved if the site includes a blog and/or a wiki. Community members should be encouraged to use their social media networks to draw attention to the ongoing work.

In addition to the Web site, a publicly archived mailing list should be established to which anyone can subscribe. This is the primary channel through which interested parties will be informed that new drafts of the vocabulary have been published and comments can be received. Mailing lists are made available on the OSOR platform for this purpose,

Drafts are snapshots of an evolving 'live' document. In between the publication of working drafts, the presumption is that the group will also make its 'live' document public. This works towards the aim of achieving consensus among the widest possible community. However, the

Working Group should make it clear that review comments should refer to the static public drafts, not the live document. A Working Group may choose not to make its live document public if that is the consensus of the individuals concerned.

Meetings and private communication

In addition to the public mailing list and Web site, the Working Group must be able to communicate with each other in private. To achieve this, it is important to have an archived private mailing list to which only Working Group members (and system administrators) have access. Furthermore, there must be a means of easily sharing documents, including meeting agendas and minutes. Various free and paid for proprietary tools exist to facilitate this.

It is expected that Working Groups will meet regularly via telephone conference. In addition to the audio channel, it is very helpful to establish a real-time text-sharing system such as a chat room, or shared Instant Messaging system. This acts as both the medium through which meeting minutes can be taken and augmented by any participant, and as a channel through which terms, references and hyperlinks can be communicated easily. The [Arkadin](#) system provides this functionality and more, and is made available for use by Working Groups. A project area in the Commission's Confluence wiki is also available to the Working Group as a means of sharing internal documents and other project management-related materials.

Working groups should meet (via telephone conference) regularly. Typically this will be a weekly call at a regular time although Working Groups may choose a different frequency.

As noted in section 3.1.3, face to face meetings should be seen as the exception. However, if one is required, it should be planned with ample advance notice (typically 8 weeks). Such meetings can provide useful points in the calendar by which time milestones are met. They also act as a focus for community building and Working Group cohesion. These aspects are hard to quantify against the cost in time and money of arranging and participating in a face to face meeting, but experience shows that “a lot gets done” during a face to face meeting. They are generally seen as a good use of resources if held approximately every 4 months.

Transparency and record keeping

Transparency and record keeping is an important aspect of achieving consensus among a community. To this end, Working Groups must take care to document their work, i.e. how the vocabulary was developed.

A first step towards this is to ensure that an agenda is sent before each meeting and that minutes are kept. These should certainly be available to all members of the Working Group and many groups make their agendas public (again to foster transparency and community engagement).

During its work, the group will raise and discuss a variety of issues, each of which should be recorded and assigned an identifier. It is helpful to quote the identifier in any e-mail discussion about the issue and the chair should encourage this. Issues should be resolved by the Working Group and the resolution recorded before the issue is marked as ‘closed.’

Similarly, individual participants will be assigned action items during meetings. These too should be recorded, clearly stating to whom the action has been assigned and the date by which s/he agrees to complete it. **Do not** assign action items to 'everyone' since in reality this is often taken to mean 'no-one.'

Finally the Working Group must establish a system that records comments received and how these have been dealt with. Each comment must be addressed by the Working Group which should record its decision on whether to accept the comment (in part or in full) or reject it. In the latter case, the Working Group must give a reason why the comment has been rejected. Accepting a comment usually means making a change to the vocabulary or the documentation.

Once a comment has been dealt with, the commenter should be informed of the group's decision(s) and reasoning. This may be an iterative process but ultimately, the Working Group needs to be able to show that every comment has been dealt with to the satisfaction of the commenter. The JIRA is made available to Working Groups for this purpose.

Dispute resolution

The goal is to achieve consensus among the community group, including external commentators and this is usually possible. However, human nature plays a part and there may be times when consensus cannot be reached on an issue or a comment received. In such cases, one possible course of action is to seek external guidance. For instance, by posing a question to another Working Group or discussion forum. Alternatively, the Working Group may take a simple vote to settle an issue, with the chair having the casting vote (such votes should be recorded). Ultimately the EC may settle a dispute but this is very much a last resort as it shows that full consensus has not been reached.

Objectives

To set the rhythm and pace for the group's work.

To establish effective communication within the Working Group and between the Working Group and rest of the community.

To operate in an open and accountable way, ensuring that the group's work is fully documented and its decisions can be assessed.

Input

There is no specific input.

Tasks

Process Step 7. Establish a working environment and culture. Establish a working environment and agree on a working culture.

Implementation

The subcontractor responsible for the semantic asset project will establish a working environment as soon as Process Step 1 has started as follows:

- create a new semantic asset project on the Commission's Confluence wiki. This enables the sharing of documents and information between the Working Group members;
- establish two archived mailing lists on the OSOR platform:
 - **A private mailing list with suffix “-wg”**. All Working Group members will be subscribed to this list. This list can be used for internal discussions within the Working Group.
 - **A publicly available mailing list with suffix “-rg”**. Anyone, including project members, can subscribe to this mailing list and send messages to it.
- establish an issue tracking system on the JIRA platform;
- establish a publication route for public documents.

Management of these tools is placed at the discretion of the Working Group. The identified chair of the Working Group will have administration rights on all available tools.

- The subcontractor will request the relevant system administrators to create user accounts on the relevant platforms for each member of the Working Group. The e-mail address of the Working Group members will be added to a private Working Group mailing list. The identified chair of the Working Group will be granted “project owner” privileges. The Working Group members will be invited as soon as Process Step 2 has been completed, which implies that the members of the Working Group are known.
- The subcontractor will request the relevant system administrator to create a user account for each Review Group member, will add them as “members” to the semantic asset project, and add them to the Review Group mailing list. This ensures that each member of the Review Group is notified when a draft is published for review.
- The subcontractor will create the following issues in the issue tracker and assign them to the chair for follow up:
 - set up a meeting schedule and make this available on the Wiki;
 - set up a release schedule and make this available on the Wiki;
 - set up a kick-off meeting and make the agenda available on the Wiki;
 - identify the editor(s) of the specification.

Both schedules should foresee at least 2 meetings and draft releases. This is deemed to be a minimum to receive sufficient feedback from the community and reviews from the Review Group.

These tasks must be completed within a week after appointing the chair.

Output

An effective communication system that supports the community and the Working Group.

A set of rules collaboration rules ensuring a positive working culture.

A complete recording system to keep track of the Working Group's actions and decisions.

Timing

It is important to establish communication channels and practices when the Working Group is first established. At the same time, establishing the meeting schedule is an important agenda item for the Working Group's first meeting.

The recording system should also be established as soon as possible. This way, it can be used immediately to track the tasks that resulted from the Working Group's first meeting.

3.2.3 Draft process

The Working Group is responsible for developing the vocabulary, but at the same time the community is engaged in reviewing the document, thus ensuring the broadest possible consensus.

Each draft must be published at its own stable URI. A 'latest version' URI must also be established that maps to whichever is the latest version. For example, the Working Group might publish drafts on the 1st of February and the 14th of March 2012, in which case the URIs of those drafts might be analogous to:

```
http://example.org/drafts/mycorevoc/20120201
```

```
http://example.org/drafts/mycorevoc/20120314
```

The URI `http://example.org/mycorevoc` would initially be mapped to `http://example.org/drafts/mycorevoc/20120201` but would switch to the later draft when it was published. This 'latest version' URI must be chosen with care as it should be permanent and will often serve as the namespace document which imposes further requirements (see section 4.8).

Each draft should also link to its predecessor so that reviewers can easily navigate back through the vocabulary's development.

Objectives

To enact an iterative process whereby draft versions of the vocabulary are published, subjected to community review, and updated in the light of such comments.

Input

The specification template (see section 6) provides guidance on the structure of the document.

Tasks

Process Step 8. Publish drafts. Publish drafts of the core vocabulary as it evolves, seeking feedback from the community on each occasion.

Implementation

The drafts will be published as Releases for the Asset project. Each published draft will be put up for review using the public mailing list, to which all Review Group members are subscribed, and actively promoted via multiple channels.

A minimum of 2 interim drafts is foreseen:

- A first draft that will already be relatively stable as it will draw on existing vocabularies wherever possible.
- The second foreseen draft is expected to be of such quality that it would only receive detailed feedback and criticisms.

The release schedule must be made available on the Wiki. If the Working Group decides to publish drafts more frequently, it is welcome to do so. However, the minimal amount of time between two subsequent releases must be at least two weeks. This helps ensure that reviewers and the community have ample time to review and comment on the draft.

Output

Published drafts of the core vocabulary.

Timing

The frequency and timing of such publications will be determined by the Working Group members who will need to find a balance between showing that work is actively being carried out and bombarding the community with too-frequent requests for comments.

3.2.4 Review process

Each published draft of the Core Vocabulary will include the e-mail address to which comments should be sent (see section 6). This will be the publicly available, publicly archived mailing list. When comments are received, whether from the global community (section 2.4) or the Review Group, the Working Group assigns an identifier to each comment within each e-mail and must address each one. The Working Group can resolve each comment in one of three ways:

Accepted

This usually means that the comment is acted on fully and will almost certainly entail making changes that will be reflected in the next draft.

Rejected

The Working group decides that the comment should not be acted upon (and therefore no change made to the draft).

Partially accepted

Some of the comment is accepted but other parts are rejected.

The resolution of each comment should be recorded by the Working Group, usually within a meeting, but it is sometimes more practical to deal with some comments through the group's mailing lists. The Working Group must contact each reviewer and inform them of the resolution concerning each of their comments. The reviewer may make further comments on the resolution and the Working Group must make every effort to reach consensus.

Objectives

To obtain feedback from interested parties outside the Working Group

Input

Drafts of the Core Vocabulary

Tasks

Process Step 9. Process Comments. Consider comments received concerning the interim published drafts.

Implementation

The comments will be sent to the public mailing list and it is through this medium that the Working Group reports back to the reviewer. Each comment is logged as an issue on the Asset Issue Tracker. The e-mail sent back to the reviewer should include the original comment, the resolution of the Working Group and the justification for the resolution, particularly where the comment is rejected.

Output

A public record, via the mailing list archives, of the comments received and the actions taken.

Timing

The review starts as soon as drafts are sent out for review.

3.2.5 The Last Call Working Draft

When the Working Group believes that it has done all it can to develop the vocabulary, it should publish a Last Call Working Draft and seek feedback from the Review Group and the wider community. A date must be set by which comments should be made and the comments handled as described under Transparency and record keeping in section 3.2.2.

If the comments received lead to substantive changes being made to the vocabulary, particularly any changes that would entail an existing implementation to be re-engineered, then the Last Call process must be repeated.

If, however, the comments received lead to only trivial or editorial changes, a further call will not be necessary and the vocabulary can progress to the next stage.

Objectives

To ensure feedback has been received from the Review Group and the wider community.

Input

The Working Group should record the decision to go to Last Call in the minutes of the relevant meeting.

Tasks

Process Step 10. Publish the Last Call Working Draft of the vocabulary and contact the Review Group, seeking its feedback. The wider community should also be invited to comment. The call should set a date by which comments must be received.

Implementation

The implementation of this step is entirely similar to Process Step 8 but the chair of the Working Group will indicate this draft to be the Last Call Working Draft and that no additional draft will be published. The chair will also communicate a deadline by which comments have to be received. The announcement will be done through the release mailing list.

Output

The published Last Call Working Draft, messages to the relevant groups.

Timing

The Last Call period should last not less than 4 weeks but shorter periods may be necessary, for example, to aim for a particular meeting of the endorsement group.

3.2.6 Review Group Feedback

As noted in section 3.1.4, the Review Group is formed at the same time as the Working Group and comprises ISA Member State representatives. Members of the Review Group are also members of the global community (section 2.4) and, as such, are welcome to make comments on any aspect of the Core Vocabulary as it evolves. Within the specific context of the Review Group, however, reviewers should consider whether the proposed Core Vocabulary:

- meets the needs of the Member State they represent and others about which they are knowledgeable;
- has a high likelihood of enabling greater interoperability with and between European Institutions;
- conflicts in any way with existing practices;

- may be improved, for example by the addition or subtraction of terms, by refining definitions, by linking to other vocabularies.

The Working Group must carefully consider the feedback provided by the Review Group and record its considerations, noting what changes have been made to the vocabulary or providing justification for why changes have not been made. The Review Group will then be asked to accept or reject the actions taken by the Working Group. This is an iterative process with the clear aim of reaching full consensus between all members of the Working and Review groups. Both the Working and Review Groups should be aware of the date of the Endorsement Group's next meeting and make every effort to be ready to present the final draft by that date (see section 3.1.5).

Implementation

Comments by the Review Group should be made via the public mailing list and will be handled following the process detailed in 3.2.4.

If consensus cannot be reached then the Working Group or the Review Group may refer the matter to the EC for resolution but this is an extreme situation that should not normally occur. Endorsement by the ISA Coordination Committee/TIE Cluster is less likely to be secured in the event of the EC arbitrating between two strongly held but conflicting positions.

3.2.7 Review Last Call Working Draft

There is no difference between the way a comment received on the Last Call Working Draft is handled and any of its predecessors. There is however a greater sensitivity in that the document that emerges from this process has the backing of both the Working Group and the Review Group. Therefore the review process described in section 3.2.4 must be followed rigorously.

Objectives

To show consensus between the Working Group and the Review Group.

Input

The Last Call Working Draft

Tasks

Process Step 11. Review Last Call Working Draft.

Implementation

As detailed in section 3.2.4

Output

A record of the comments received, particularly from the Review Group, the resolutions of the Working Group, and evidence that consensus has been reached.

Timing

Review Group comments must be received and considered before the Core Vocabulary can be submitted to the Endorsement Group.

3.2.8 Gather evidence of actual or expected use

Dealing with the comments received on the Last Call Working Draft provides evidence that consensus has been reached on the terms in the vocabulary. This needs to be augmented with evidence of actual or planned deployment. This can take several forms.

Usage reports

By far the best form of evidence that a vocabulary is fit for purpose and will be broadly accepted is a set of reports from organisations that have already begun to use it. Such reports are most likely to come from members of the community that has supported the work throughout the process. The reports should describe how the vocabulary was used and, crucially, how using it aided interoperability.

Working groups are most likely to secure implementation reports if relevant organisations are active in the Working Group from the beginning and create test implementations throughout the process (see section 3.1.2).

In the event that implementation reports highlight problems or shortcomings with the vocabulary, the Working Group may have to re-draft the vocabulary.

Evidence of intention to adopt

It is possible that an organisation's decision to use a vocabulary to aid interoperability will be dependent on securing endorsement by the Member States. Furthermore, since a vocabulary is not a technology, testing in the engineering sense may not be necessary. However, what *is* necessary as a minimal substitute is evidence that the Core Vocabulary is seen as useful by organisations that publish government data and that those organisations intend to use it.

Objectives

Proving that the vocabulary is seen as fit for purpose.

Input

N/A

Tasks

Process Step 12. Gather evidence of acceptance by the potential users of the vocabulary.

Implementation

The evidence will be stored as a Document on the Asset project on the project wiki. Evidence of acceptance may continue beyond the life of the Working Group.

Output

Evidence of actual or intended use.

Timing

Evidence should be gathered during the Last Call Working Draft phase so that it can be presented, along with the final draft of the vocabulary, for endorsement.

3.2.9 Endorsement process

The final stage in the process is to secure endorsement by the Member States of the European Union, acting through the Interoperability Solutions for European Public Administrations (ISA). The final draft of the vocabulary, together with evidence of actual or intended usage, should be submitted to the ISA Coordination Group/TIE Cluster.

The endorsement process itself is out of the scope of the process and methodology defined in this document.

Objectives

Endorsement by the Member States

Input

Several documents will be collated for this purpose:

- The namespace document in its three formats (human readable, XML and RDF, see section 4.9)
- The use cases and requirements document (sections 4.4 and 4.5)
- A report on the development of the Core Vocabulary that summarises the challenges faced and rationale for the decisions taken. The report should also include the evidence gathered in 3.2.8

Tasks

Process Step 13. The EC to submit documents to the TIE Cluster or ISA Coordination Group.

Implementation

Upon endorsement, the chair of the Working Group will publish the namespace document as version 1.0 of the Asset as well as the report on the Core Vocabulary's development. The release will be announced via the release mailing list.

Output

Version 1.0 of the completed, stable vocabulary, endorsed by the Member States.

Timing

This step cannot be entered into until all previous steps have been completed.

3.3 MANAGING CHANGE

The final "version 1.0" of any vocabulary should be stable. That is, it should not be edited ('the bytes are locked'). Therefore, the documentation must include a link to a separate errata document which can be edited.

3.3.1 Editorial changes only

If minor editorial changes need to be made to the vocabulary, such as correcting typographical errors or clarifying a vocabulary term without changing its meaning, then such changes can be recorded in the errata document by the EC.

3.3.2 Vocabulary term changes

If an error comes to light after publication of the vocabulary that, when corrected, *would* require implemented software to be updated, then this entails a new version of the vocabulary to be developed. Neither the Working Group nor the EC can make changes to the terms in a published vocabulary. The process defined in this document would therefore need to be followed again, up to and including endorsement, to create the new version. The original version 1.0 must remain intact.

Although the vocabulary document cannot be edited in any substantive way, if a new version is created and endorsed, the EC will add a suitable notice to the original vocabulary pointing to the new version which supersedes it.

4. METHODOLOGY

In this section, we set out the technical stages that a Working Group must go through in its development of a new core vocabulary.

4.1 IDENTIFY A MEANINGFUL SET OF CORE CONCEPTS

A necessary first step is the identification of a set of meaningful Core Concepts, within which the Core Vocabulary to be developed can be positioned. Like any vocabulary, Core Vocabularies are related to one another. It is important to be able to see the bigger picture prior to starting any further work on Core Vocabularies.

One difficulty that will arise in constructing a meaningful set of Core Concepts, is that *any* concept might have the potential to be used in the context of an e-Government application. This might lead to an overly large set of potential Core Concepts and Core Vocabularies. It is therefore important to restrict the scope to e-Government. In particular, this means that any candidate Core Concepts must make sense in a large super-context of e-Government, and thus across different sectors of public administration. Core Concepts like “Organisation” or “Person” satisfy this scope criterion. Concepts like “Electronic Component” do not.

Objectives

To prioritise the core vocabularies to be developed and to detect dependencies between core vocabularies and identify coordination needs at an early stage.

Input

Existing libraries of core concepts, such as:

- The Aggregate Core Components in the UN Centre for Trade Facilitation and Electronic Business [UN/CEFACT] Core Component Library (CCL);
- the Schema.org type hierarchy [Schema.org];
- Foundational Ontologies;
- Relevant work from operational EC DGs and/or pan-European projects
- Relevant work from MSs
- Any relevant work conducted at standardisation bodies

Tasks

Method Step 1. Identify the Core Vocabularies that are likely to meet the most pressing needs of the potential users within European institutions and for which the relevant expertise is most readily available.

Output

A list of core concepts.

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.

4.3 RESEARCH EXISTING DATA AND SERVICES

Having highlighted the importance of researching existing vocabularies, it is also necessary to research how existing data and services are currently structured. Of special importance here are the Base Registries maintained by the Member States [ISABR]. The proposed Core Vocabulary must not cause conflicts with legacy systems.

Method Step 3. Research existing published data and services, noting that there should not be any conflicts with the proposed Core Vocabulary. As with the research on existing vocabularies, this research should inform the writing of use cases.

The research carried out into existing vocabularies, data and services should be made public but this does not need to be formalised. Publicly archived e-mails are sufficient.

4.4 USE CASES

Based at least in part upon the output of the preceding research activities, the Working Group should articulate a series of use cases. These are written in prose and are typically of the form:

Business need: *Stakeholder group X needs to find, compare, exchange,... information in an electronic fashion and hereby requires a minimum level of semantic interoperability.*

Usage scenario: *Alice is trying to do X. She finds resources A, B, and C and wishes to combine them to get an overall picture. Noticing that they all use the same terms from vocabulary V she is able to match up the data.*

Derived requirements: *The following terms in vocabulary V need to be agreed on...*

This over-generalised use case will be at the heart of any vocabulary development so the use cases devised by the Working Group should be quite specific about what it is that the vocabulary will enable the user to do that they currently can't do. In other words, the use cases should set out the problem, or problems, that the new vocabulary will solve. Against this, it is important to note the points made in section 4.1: Core Vocabularies are, by definition, very general and not specific to any sector.

It is worth encouraging all members of the Working Group to write use cases, perhaps in a shared workspace such as a wiki. This improves the likelihood that each member's perspective is taken into account.

Method Step 4. Articulate the problem(s) that the Working Group is trying to solve in the form of a series of use cases.

4.5 REQUIREMENTS

Once the use cases have been written, these should be reduced to a set of requirements that the new vocabulary must meet. In complex cases it may be appropriate to assign a priority to each requirement using the familiar RFC2119 Keywords of MUST, SHOULD and COULD however this is unusual for Core Vocabularies as the assumption is that all requirements will be met. At the end of the process, it should be possible to match the final draft of the vocabulary against this list of requirements and ensure that they have indeed all been met.

The use cases and requirements should be published by the Working Group. This is separate from the vocabulary document itself but the latter should refer to the former.

Method Step 5. Derive a set of requirements from the use cases.

Method Step 6. Publish the use cases and requirements in a single document.

4.6 TERMINOLOGY AND CONCEPTUAL DATA MODEL

A Core Vocabulary will describe concepts, relationships and properties.

For clarity in *this* document:

A **concept** is a 'thing', such as a vehicle, a person, a dataset or a document. It might also be an activity such as a meeting or something more abstract such as a particular performance of a music work. In RDF terms, a concept will be represented as a Class.

A **relationship** is something that links two concepts (things) together. In RDF terms, a relationship is a predicate that links two classes.

A **property** is generally a value in a particular dimension such as the weight of an object, the name of an organisation. In RDF terms, a property is a predicate that takes a literal value which may or may not be typed.

It is important that the concepts, relationships and properties that comprise a vocabulary are conveyed clearly and this is usually best achieved through the use of a diagram. This not only makes the vocabulary's framework clear for potential users, it clarifies it for Working Group members too. Although not mandatory, UML [UML] provides an excellent and well understood method of creating such diagrams. As an example, the diagram reproduced in Figure 2 gives an easy to follow overview of the Organisation Ontology [ORG].

Method Step 7. Unless the vocabulary is very simple with no more than a handful of concepts, create a concept diagram. Consider using UML.

Notice that Figure 2 includes terms from other vocabularies. As noted in section 4.1, it is good practice to re-use existing vocabularies and these are likely to be a part of any conceptual model.

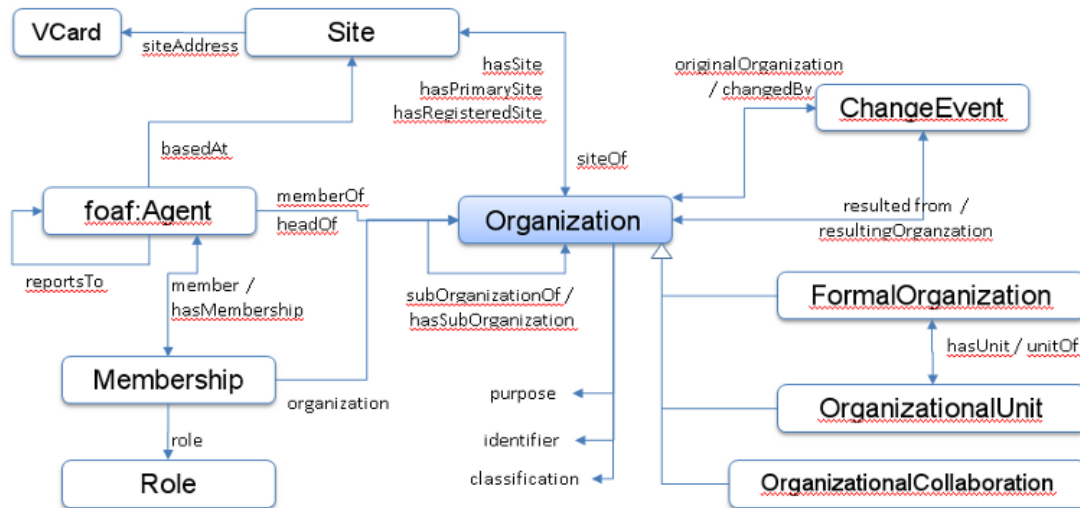


Figure 2: The UML Diagram for the Organisation Ontology designed by Dave Reynolds of Epimorphics [ORG]

When developing a vocabulary it is tempting to be specific about how it should be used: to include lots of cardinality rules for XML implementations and domain and range restrictions for the Semantic Web. This temptation should be resisted. The aim is to create a Core Vocabulary that can be used widely to aid interoperability. Therefore, it is to be hoped that it will be used in situations not foreseen by the Working Group or the EC. Unnecessary restrictions on how each vocabulary term can be used may hinder adoption by users who would otherwise find the vocabulary useful for their purposes.

Method Step 8. Do not impose cardinality rules or domain/range restrictions on vocabulary terms unless strictly necessary.

4.7 NAMING CONVENTIONS

In theory, a Working Group can assign any name it likes to any concept, relationship or property. In practice, there are two constraints on the kind of names that should be used:

- technical;
- conventional.

4.7.1 Technical constraints

The vocabulary will be encoded in at least RDF and XML and both languages effectively require that terms begin with an upper or lower case letter from the ASCII character set, or an underscore. This tight restriction means that, for example, terms may not begin with a number, hyphen or accented character. A formal definition of these restrictions is given in the XML specification document [XML].

Method Step 9. Use words beginning with an upper or lower case letter (A - Z, a - z) or an underscore (_) for all terms in a vocabulary.

4.7.2 Conventional constraints

Beyond the technical constraints there is conventional way of naming terms that, if followed, will increase the readability of the vocabulary and, importantly, confidence in it. In brief:

- concepts (classes in RDF) begin with an upper case letter, e.g. `Person`, `Vehicle`, `Address`;
- relationships and properties use camel case, such as `creator`, `isPartOf`, `familyName` etc.

When creating property names, it is conventional to use simple nouns such as `name`, `weight`, `colour`. This is much preferred over terms such as `hasName`, `hasWeight` or `hasColour` as the 'has' prefix does not add anything to the term's meaning.

Method Step 10. Use simple nouns for property names.

For relationships, use a verb such as `includes`, `replaces`, `manages`. In this situation, it is acceptable to use the 'has' prefix, so that the `hasSite` relationship from the organisation ontology is consistent with conventional practice (see Figure 2).

Method Step 11. User verbs for relationship terms.

Relationships are usually bi-directional. Referring to the organisation ontology again, the concepts of an `Organisation` and a `Site` are defined along with two relationships that are the inverse of each other:

`Organisation` → `hasSite` → `Site`
`Site` → `siteOf` → `Organisation`.

Vocabularies should define such inverse pairs for relationships although this does not extend to properties. So, while Dublin Core includes a property of `dateAccepted`, there is no inverse

property that would link a given date, which is expressed as a simple value, to all the documents accepted for publication on that date.

Method Step 12. For each relationship, include a definition of its inverse.

The organisation ontology also exemplifies the use of the preposition 'of' as a suffix. 'By' and 'To' are also common in this regard such as `embodiedBy` and `conformsTo`. However, if the preposition can be avoided, do so.

Method Step 13. Use prepositions in vocabulary terms only if necessary.

4.8 IDENTIFIER CONVENTIONS

The steps in this section adhere to the Best Practices set out in the Cool URIs for the Semantic Web [COOL].

As noted earlier, Core Vocabularies must be published in multiple formats, including RDF. This entails the assignment of identifiers to each term and, to be useful in the kind of linked data applications envisaged, those identifiers should be HTTP URIs.

At its simplest, the process of minting identifiers for Core Vocabulary terms is straightforward:

1. decide on a namespace;
2. append each term to that namespace.

Once again the Organisation Ontology provides a good example. Its namespace is `http://www.w3.org/ns/org#` so that the identifier for the concept of an organisation is `http://www.w3.org/ns/org#Organization`.

This example conforms to several best practices that must be followed.

First of all, it uses the hash method of providing individual identifiers. As already emphasised, Core Vocabularies developed under this process should take a minimalist approach. This means that existing vocabularies will be re-used wherever possible and that the number of new vocabulary terms to be defined is likely to be relatively low. The alternative approach of using a namespace ending with a slash character is suitable for data sets and large scale vocabularies but is rarely applicable in the case of Core Vocabularies.

Method Step 14. Use a namespace ending with a hash character (#)

The namespace includes a short but easily human-readable portion that makes it easy to see what the vocabulary is. In this case that key phrase is 'org' but notice that it appears after 'ns' (short for namespace). The fact that the namespace is on an Internet domain that ends in .org is coincidental. Furthermore, the namespace is relatively short - as short as it can practically be for the environment in which it is deployed. Use of such mnemonic URIs makes writing them, remembering them, and debugging software that uses them, much easier than long strings of meaningless characters.

Method Step 15. Keep the namespace as short as possible.

Method Step 16. Include a portion that identifies the vocabulary for human readers.

Namespaces should not change. They should remain stable for decades: periods of time in which new technologies will emerge and be implemented. Therefore it is important that the namespace does NOT include any technology-specific component such as .php or .aspx.

Method Step 17. Do not include any technology-specific component in the namespace (except HTTP).

The namespace is politically neutral. Individuals, organisations and countries are, by nature, protective of their identity. A namespace that includes the name of a DG, European institution or Member State is very unlikely to be used by anyone else, no matter how perfect the semantic match. A namespace in the .eu top level domain is very unlikely to be used outside Europe. Such protective attitudes may limit the potential usage of a vocabulary, the exact opposite of the aim of the Core Vocabulary which is to encourage and facilitate interoperability.

Method Step 18. When choosing a namespace, do not restrict the pool of potential users unnecessarily by using a namespace that declares 'ownership' or geographical relevance.

If it is not possible to follow these methods within the planned hosting arrangement for the Core Vocabulary, it may be appropriate to use a persistent URL, or PURL. These provide a level of indirection between a permanent, stable URI and what may be a dynamic resource that the PURL identifies. Dublin Core does this. It's most widely used vocabulary term is `creator` for which the identifier is `http://purl.org/dc/terms/creator`. A request to that URI is currently redirected to `http://dublincore.org/2010/10/11/dcterms.rdf#creator`. Notice the date in the eventual URI. This allows the Dublin Core Metadata initiative to revise the documentation of the `creator` term and publish it at a new URI whilst maintaining the stable identifier.

purl.org itself is a free service that many organisations choose to use but software is readily available that will allow organisations to run their own PURL service under their own domain name and infrastructure [PURLs].

Method Step 19. If necessary, consider meeting Method Step 18 using PURLs

4.9 THE NAMESPACE DOCUMENT

The Core Vocabulary must be available from the namespace. Furthermore, it should be available in at least 3 formats:

- HTML - this is the primary output of the Working Group. Following the template in section 6, it sets out all the terms in the vocabulary and their definitions as well as including additional sections such as the conceptual data model (section 4.6) and the conformance criteria (section 0).
- XML - as an XML schema.
- RDF/RDFS - as RDF/XML and, optionally, also as Turtle.

The HTML representation is the human-readable document. Other formats are machine-readable. The multiple machine-readable formats reflect the different technologies favoured by different engineers.

The representation returned from the namespace must be determined through server-side content negotiation. Each representation has its own URI, however, which is the namespace plus the relevant extension, i.e. {namespace}.html, {namespace}.xml, {namespace}.rdf and, if provided, {namespace}.ttl. Each representation should include appropriate pointers to the others, must be validated and must be returned with the correct content-type at the HTTP layer.

Method Step 20. Create and validate the namespace documents in HTML, XML and RDF/XML. Consider also serialising the RDF schema in Turtle.

Method Step 21. Either the Working Group or the EC must make each one available through {namespace}.ext where .ext is the relevant extension.

Method Step 22. Either the Working Group or the EC must set up content negotiation to handle requests to the namespace itself.

As pointed out in the previous section, the namespace need not necessarily be the actual location of the documents. Redirection from the namespace to a different location provides an indirection layer that might be advantageous to the Working Group and/or the EC.

The final HTML document that describes the endorsed vocabulary must include a link to an errata document in line with section 3.3.

Method Step 23. When publishing the final version of the Core Vocabulary, link the HTML version to an errata document.

4.10 QUALITY ASSURANCE/CONFORMANCE CRITERIA

The vocabulary document must include a conformance statement. That is, a statement that says what an implementation must do in order to be considered conformant with the vocabulary.

In the case of Core Vocabularies this wording is very likely to be sufficient:

A conformant implementation of this vocabulary MUST understand all vocabulary terms defined in this document.

However unlikely, it is possible that the vocabulary will have natural divisions so that it might be appropriate to set different conformance levels. For example, a vocabulary used to describe vehicles may have a group of terms related specifically to motor vehicles that could be used in an implementation that had no need to understand the terms that relate to bicycles and scooters. In such a situation there might be different levels of conformance:

A level M conformant implementation of this vocabulary MUST understand the following terms: Lorry, van, car...

A level B conformant implementation of this vocabulary MUST understand the following terms: bicycle, scooter...

A fully conformant implementation of this vocabulary MUST understand all vocabulary terms defined in this document.

Method Step 24. Include a Conformance Statement.

5. CONFORMANCE STATEMENT

In the spirit of section 0, a conformant Core Vocabulary will have been published after all the Process and Methodology steps itemised in sections 3 and 4 of this document have been fulfilled. These are summarised in section Appendix D.

6. SPECIFICATION TEMPLATE

A template is available to aid the development of the HTML document. It is similar in style and layout to specification documents produced by other organisations, notably the W3C, and therefore will be familiar to many members of the target audience. The sections of the template are as follows.

6.1 TITLE AND VERSION CONTROL

The title of the document should be the title of the vocabulary. This may, but is not expected to, change throughout the development process.

A subtitle that gives the type of document (working draft, final draft or, when published in its stable form, the version number, typically "1.0").

Identifiers and links to the current, latest and previous versions (see section 3.2.3).

Links to the XML and RDF representations of the vocabulary. These do not need to be created until the Final Draft is to be published so that earlier working drafts need not include these links.

The name(s) of the editor(s).

The copyright notice.

6.2 ABSTRACT

The abstract should give a busy reader an immediate over-view of what the vocabulary covers.

6.3 STATUS OF THIS DOCUMENT

The template provides wording appropriate for different stages in the development and publication of the vocabulary. Text wrapped in multiple @@@ symbols must be edited before publication (and the @@@ symbols removed). Working Groups should delete or comment out the unwanted sections for each publication such that the text will be broadly as set out below for each version. Working Groups are encouraged to add extra text to the status section if it aids clarity. For example, the group may be seeking feedback on a particular aspect of the vocabulary.

6.3.1 Early Working Drafts

This section describes the status of this document at the time of its publication. Other documents may supersede it.

This document was produced by the @@@ Name of Working Group, linked to its public Web space@@@, following the @@@Process and Methodology for creating Core Vocabularies@@@.

Publication as a Working Draft does not imply endorsement by the European Commission or its representatives. This is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to cite this document as other than work in progress. The Working Group will seek further endorsement by the Member State representatives in the ISA Coordination Group or the Trusted Information Exchange Cluster.

Comments on the vocabulary are invited via the public mailing @@@ link to [mailto: address@@@](mailto:address@@@) with @@@archive@@@.

6.3.2 Last Call Working Draft

This section describes the status of this document at the time of its publication. Other documents may supersede it.

This document was produced by the @@@ Name of Working Group, linked to its public Web space@@@, following the @@@Process and Methodology for creating Core Vocabularies@@@.

Publication as a Working Draft does not imply endorsement by the European Commission or its representatives. This is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to cite this document as other than work in progress. The Working Group will seek further endorsement by the Member State representatives in the ISA Coordination Group or the Trusted Information Exchange Cluster.

This is the Last Call Working Draft of the vocabulary. Comments are sought via the public mailing @@@ link to [mailto: address@@@](mailto:address@@@) with @@@archive@@@ by @@@ Date @@@.

6.3.3 Final Draft

This section describes the status of this document at the time of its publication. Other documents may supersede it.

This document was produced by the @@@ Name of Working Group, linked to its public Web space@@@, following the @@@Process and Methodology for creating Core Vocabularies@@@.

This document has been reviewed by representatives of the Member States of the European Union, PSI publishers, and by other interested parties. Publication of this Final Draft does not imply endorsement by the European Commission or its representatives. This is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to cite this document as other than work in progress. The Working Group will

seek further endorsement by the Member State representatives in the ISA Coordination Group or the Trusted Information Exchange Cluster.

Comments on the vocabulary are invited via the public mailing @@@ link to <mailto:address@@@> with @@@archive@@@.

6.3.4 Final Published Version

This section describes the status of this document at the time of its publication. Other documents may supersede it.

This document was produced by the @@@ Name of Working Group, linked to its public Web space@@@, following the @@@Process and Methodology for creating Core Vocabularies@@@.

This document has been endorsed by (@@@ISA Coordination Group Members|TIE Cluster@@@) Members. In addition it has been reviewed by representatives of the Member States of the European Union, PSI publishers, and by other interested parties. It is a stable document and may be used as reference material or cited from another document. The European Commission's role in publishing this vocabulary is to promote its widespread deployment to enhance interoperability between public sector data sets.

Comments on the vocabulary are invited via the public mailing @@@ link to <mailto:address@@@> with @@@archive@@@.

6.4 THE BODY OF THE DOCUMENT

The remainder of the document is self-explanatory. Working Groups may include additional sections and sub-sections: the template is provided as an aid, not a limit. Working Groups may wish to include a Change Log at the end of the document but this is not mandatory.

6.5 HTML/CSS

The template is written as a polyglot HTML document [PG]. Use of this particular markup is not mandatory: acceptable alternatives are XHTML Transitional (1.0 or 1.1) or HTML5. If using either HTML5 or the polyglot markup, the HTML5 'shim' must be included as in the template for backwards compatibility with older browsers.

The stylesheet is derived from that used on the joinup.eu platform. Additional styles may be added, perhaps to highlight issues and questions during the document development process, but these must not override styles in the stylesheet provided. Therefore, styles should be declared at document level above the link to the `base_semic_spec_style.css` stylesheet.

6.6 VALIDATION

All documents published by the Working Group must be validated using the appropriate tools:

HTML:	http://validator.w3.org/
CSS:	http://jigsaw.w3.org/css-validator/
XML Schema:	http://www.w3.org/2001/03/webdata/xsv
RDF:	http://www.w3.org/RDF/Validator/
Turtle:	http://www.rdfabout.com/demo/validator/

APPENDIX A COMPARISON OF EXISTING PROCESSES AND CORE VOCABULARIES

The process and methodology set out in this document closely parallels that of many other standards bodies. Before creating a tailor-made process for European PSI data interoperability, an overview was sought of the general structure of development and maintenance processes. We can distinguish the following steps:

1. Inception of an idea and submission for inclusion in a work plan leading to a formal decision to establish a work item
2. Assignment of the work item to a group of people, development of a set of working drafts leading to publication of a public draft
3. Review and comment by an identified community with resolution of issues leading to a final draft, draft recommendation or draft standard
4. Endorsement of the final draft leading to a formal recommendation or standard

This 4-phase approach is reflected in the process described in this document:

1. Inception - the EC sets out the work plan.
2. Assignment - the EC forms a Working Group and Review Group.
3. Review - the Review Group and the wider community comment on the drafts and the Working Group responds.
4. Endorsement - the Final Draft is submitted for endorsement by the ISA Coordination Group/TIE Cluster.

The descriptions below are meant as examples and not intended to be complete descriptions of the processes at the various organisations. Most notably, it omits any processes that take place if drafts are rejected.

For a full description of the processes, please refer to the source material listed at the end of this section.

1. Inception of an idea and submission for inclusion in a work plan leading to a formal decision to establish a work item and assign it to a Working Group

The right to submit proposals for work items may be either fully open or restricted to a particular group. A formal decision to start an activity is usually taken on the basis of a proposed charter by an identified officer of the standards organisation after evaluation by some form of advisory or coordinating committee or by a vote of (a representation of) its membership.

Examples:

- CEN/European Standards: a proposal may be submitted to the CEN Management Centre (CMC) by a CEN Member, a CEN Technical Committee, the European

Commission or EFTA Secretariat, an international organization or a European trade, professional, technical or scientific organization. The CMC circulates a proposed decision to the CEN Technical Board (BT) who then instructs the Technical Committee to set up a Working Group or assign the item to an existing group.

- CEN/CEN Workshop Agreements: any party interested in developing a CWA submits a request to the CEN Member or to the CEN-CENELEC Management Centre (CCMC). The CCMC announces the proposal for a Workshop to its Members who have the right to object. The BT is informed if the subject of the Workshop falls within the scope of an existing Technical Committee
- DCMI: proposals may be submitted by Directorate (e.g. as a result of an outsourced activity), Affiliates, Communities, Task Groups, other organizations external to DCMI, or any individual. The Directorate consults with the Advisory Board before deciding on the further procedure.
- INSPIRE: Proposal can be made by Legally Mandated Organisations (LMO) or the Spatial Data Interest Community (SDIC). New drafting teams that develop implementation guidelines are proposed by SDICs and LMOs and established by the European Commission INSPIRE Team (CT).
- ISO: proposals are submitted by a national member body, a technical committee or subcommittee, a policy level committee, the technical management board, the Chief Executive Officer, a body responsible for managing a certification system operating under the auspices of the organization or another international organization with national body membership to ISO as a whole. The Chief Executive Officer consults various interested parties and circulated to all national bodies. On the basis of the replies the Technical Management Board decides to establish a new technical committee or assign the work to an existing technical committee, which then can establish a specific Working Group.
- OASIS: Any three members can start a technical committee upon submission of a proposal to the OASIS TC administrator. If the proposal meets the documentary requirements, the OASIS TC Administrator post the submission to the membership with a call for participation.
- W3C: activities are started based on interest from the Team or from a Member submission to the Team. The W3C Director solicits review from the Advisory Committee and creates an Activity if there is consensus.
- UN/CEFACT: Any stakeholder, defined as a person or organization that would like UN/CEFACT to do something for them, can express a need by submitting a CEFACT Activity Request (CAR). In case of a project proposal, the chair of a Permanent Group submits this to the Forum Management Group. The Forum Management Group approves or rejects the proposal.

2. Development of a set of working drafts leading to publication of a public draft

On the basis of some form of work plan or development roadmap, the Working Group will develop the idea into a working draft that may be revised and updated based on consensus within the group itself, until a stage is reached where the proposal is sufficiently mature to be

put up for comment by a wider group. That process itself may be closed (restricted to the members of the Working Group) or open (visible to a wider group or even the public)

Examples:

- CEN/European Standards: The Working Group iterates the production of working drafts that are sent to the parent Technical Committee until the TC Convener and TC Secretary agree that the draft is ready for publication in the CEN Enquiry Process
- CEN/CEN Workshop Agreements: The Workshop participants draft the CWA(s) according to the specifications laid down in the business plan.
- DCMI: All working drafts are public documents. The group working on the item keep the community informed of all activity.
- ISO: In the Preparatory stage, a Working Group will work on the work item until it decides to submit a Committee Draft or a Publicly Available Specification
- INSPIRE: The Drafting Team produces a draft. Draft Implementing rules are subject to stakeholder and/or public consultation
- OASIS: The TC may at any stage during development of a Work Product approve the Work Product as a Committee Specification Draft or Committee Note Draft, as appropriate. Approval of these drafts shall require a Full Majority Vote of the TC.
- W3C: a first Working Draft is announced by the Director which is a signal to the community to begin reviewing the document. Subsequent drafts may be issued taking into account comments.
- UN/CEFACT: Internal Drafts are circulated within the parent PG until the PG approves the advancement of a Public Draft.

3. Review and comment by an identified community with resolution of issues leading to a final draft, draft recommendation or draft standard

The public draft will then be made available to a wider group of reviewers, either in a restricted group such as members or to the wider public, in an announcement that solicits comments from the community. After a pre-established time period, the Working Group will gather comments, discuss them and take decisions to resolve them. These resolutions will then be communicated to the reviewing community, together with a final version of the proposal.

Examples:

- CEN/European Standards: the TC Secretary sends the draft to the CCMC who checks formal requirements and dispatches the draft to the National Member Bodies who can vote with comments within five months. This process can be repeated with 2, 3 or 4 month ballot periods. The received comments are analysed and evaluated by the TC (possibly delegated to the WG) and a 'table of decisions' is prepared which, together with a revised draft, is sent to the TC for final decision to proceed to Formal Vote.
- CEN/CEN Workshop Agreements: a draft is made available to the registered Workshop participants. If the Workshop has defined this in its business plan, an external

commenting phase of a minimum of 60 days can take place. This is mandatory for CWAs that are of public interest. Registered WS participants send in their comments. In case of an external comment phase a comments resolution report is created compiling all received comments. The Chairperson decides when consensus is reached. The WS secretariat sends the approved CWA to CCMC for publication.

- DCMI: When consensus is reached, a public comment period is announced on the DCMI Web site. The group working on the item aims to resolve all comments received in a four-week comment period and submits the revised draft to the Directorate. This draft goes through an initial review in the Advisory Board for at least four weeks. If the AB is in agreement, a Proposed Recommendation is published.
- ISO: The committee stage is the principal stage at which comments from national bodies are taken into consideration, with a view to reaching consensus on the technical content. The comment period is open for 2, 3 or 4 months as agreed by the TC or SC. The Secretariat prepares a compilation of comments with proposed resolutions. Consideration of successive drafts shall continue until consensus has been obtained. Within ISO and JTC 1, in case of doubt concerning consensus, approval by a two-thirds majority of the voting members of the technical committee or subcommittee may be deemed to be sufficient for the committee draft to be accepted for registration as an enquiry draft; however every attempt shall be made to resolve negative votes. At the enquiry stage, the enquiry draft (DIS in ISO, CDV in IEC) shall be circulated by the office of the CEO within 4 weeks to all national bodies for a 5 months vote. An enquiry draft is approved if a) a two-thirds majority of the votes cast by the voting members of the technical committee or subcommittee are in favour, and b) not more than one-quarter of the total number of votes cast are negative. When the chairman has taken the decision to proceed to the approval stage or publication stage, the secretariat of the technical committee or subcommittee prepares, within a maximum of 4 months after the end of the voting period and with the assistance of its editing committee, a final text and send it to the office of the CEO for preparation and circulation of the final draft International Standard.
- INSPIRE: Each Draft Implementing Rule is subject to stakeholder and/or public consultation, before the Commission submits the proposals for Implementing Rules to the Regulatory Committee envisaged by the INSPIRE Directive.
- OASIS: The decision by the TC to submit the draft for public review requires a Full Majority Vote, and must be accompanied by a recommendation from the TC of external stakeholders who should be notified of the review. The TC may conduct any number of review cycles (e.g. approval to send a Committee Specification Draft to public review, collecting comments, making edits to the Committee Specification Draft, etc.). The initial public review of a public review draft must take place for a minimum of 30 days, and any subsequent reviews must be held for a minimum of 15 days. The draft may not be considered for approval by the TC as a Committee Specification or Committee Note until it has undergone a review cycle during which it has received no comments that result in any changes. After the public review, the TC may approve the Committee Specification Draft as a Committee Specification or the Committee Note Draft as a Committee Note. If any comments have been received during the most recent Public Review period, that vote may not commence any earlier than 7 days after the last day of

that Public Review. The approval of a Committee Specification or Committee Note shall require a Special Majority Vote.

- W3C: A Last Call announcement is made which is a signal that the Working Group believes that it has satisfied its relevant technical requirements (e.g., of the charter or requirements document) in the Working Draft; the Working Group believes that it has satisfied significant dependencies with other groups; and that other groups should review the document to confirm that these dependencies have been satisfied. The announcement begins a review period that should last at least three weeks but may last longer if the technical report is complex or has significant external dependencies. This process leads to either a Call for Implementations or a Call for Review of a Proposed Recommendation.
- UN/CEFACT: The UNECE secretariat provides links on the UNECE website to the Public Draft and related information. The FMG notifies Heads of Delegation and various e-mail distribution list subscribers that the Public Draft is available for review and provides them with review-process details. The project team processes comments and posts updated Public Drafts and comment logs to the PG website or the UNECE website (through the secretariat). The comment/update/posting cycle continues until the PG approves a project team recommendation to conclude ODP5. While the criteria, evaluation and ultimate decision to conclude ODP5 is left to the PG, the PG must ensure that the project team has met all comment processing requirements (see Annex I). The draft resulting from this step is called a “proposed UNECE Recommendation”.

4. Endorsement of the final draft leading to a formal recommendation or standard

The last stage of the process may involve a formal voting process among the membership of an organisation or could be based on a more informal notion of ‘sufficient community support’ as determined by management of the standards organisation.

Examples:

- CEN/European Standards: The CEN National Members submit their vote and their comments within the prescribed two-month period. The objective is to resolve negative votes by taking into account comments. If more than 71% of weighted votes are positive, the CCMC processes the document as an approved European Standard.
- CEN/CEN Workshop Agreements: No further action.
- DCMI: The Proposed Recommendation is available for final public comment of at least four weeks. If no serious objections are expressed, the Directorate publishes a DCMI Recommendation.
- ISO: At the approval stage, the final draft International Standard (FDIS) is distributed by the office of the CEO within 3 months to all national bodies for a 2 month vote. If the final draft International Standard has been approved, it proceeds to the publication stage as an international standard.
- INSPIRE: Endorsement is done by the Regulatory Committee envisaged by the INSPIRE Directive.

- OASIS: Approval of an OASIS Standard is a three-step process: 1. Submission of a Candidate OASIS Standard to the TC Administrator, 2. Completion of a public review lasting a minimum of 60 days, and 3. A membership-wide ballot. In votes upon Candidate OASIS Standards, each OASIS Organizational Member at the time the ballot is issued shall be entitled to cast one vote. If at the end of the voting period at least 15 per cent of the voting OASIS Membership has voted to approve the proposed standard, and if no votes have been cast to disapprove the proposed standard, it shall become an OASIS Standard immediately following the end of the voting period.
- W3C: The Director announces the call for review of Proposed Recommendation to the Advisory Committee who are given at least four weeks to comment. During the review period, the Working Group requests endorsement and support from Members (e.g., testimonials as part of a press release). The Director publishes a W3C Recommendation when satisfied that there is significant support for the technical report from the Advisory Committee, the Team, W3C Working Groups, and the public. The decision to advance a document to Recommendation is a W3C decision.
- UN/CEFACT: The Project Team submits the Proposed UNECE Recommendation to the UNECE secretariat. The UNECE secretariat presents the Proposed UNECE Recommendation to the UN/CEFACT Plenary for approval.

Source material:

CEN/European Standards: <http://kwz.me/pa>

CEN/CEN Workshop Agreements: <http://kwz.me/pT>

DCMI: <http://dublincore.org/documents/approval/>

ISO/International Standards: <http://kwz.me/pa>

INSPIRE Consultation: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/41>

OASIS Technical Committee Process: <http://www.oasis-open.org/policies-guidelines/tc-process>

UN/CEFACT Open development process: <http://kwz.me/pm>

W3C Process Document: [W3CPROC]

APPENDIX B REFERENCES

- A1.1** Improving semantic interoperability in European eGovernment systems, ISA.
http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm
- ADMS** Asset Description Metadata Schema
<http://www.semic.eu/semic/view/documents/adms-specification-v0.6.html>
- AGVOC** Comprehensive multilingual agricultural vocabulary.
<http://aims.fao.org/website/AGROVOC-Thesaurus/sub>
- ALS** "A Little Semantics Goes A Long Way", Jim Hendler ca. 1997.
<http://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html>
- BIBO** The Bibliographic Ontology, Frédérick Giasson and Bruce D'Arcus.
<http://biblontology.com/>
- CAMSS. 2010.** CAMSS, Common Assessment Method for Standards and Specifications.
webgate.ec.europa.eu/fpfis/mwikis/. [Online] 12 2010. [Cited: 19 06 2011.]
<https://webgate.ec.europa.eu/fpfis/mwikis/idabc-camss/>.
- CLA.** Contributor Licence Agreement
<https://webgate.ec.europa.eu/CITnet/confluence/display/ISACV/Contributor+Licence+Agreement>
- COOL** Cool URIs for the Semantic Web, Leo Sauermann and Richard Cyganiak. W3C Note, 3 December 2008. <http://www.w3.org/TR/cooluris/>
- CC** Creative Commons Rights Expression Language
<http://creativecommons.org/ns#>
- CPV** SEMIC.EU Core Person: A pan-European schema for personal data.
<http://www.semic.eu/semic/view/snav/Conformance.xhtml>
- CVOC** Case Study - SEMIC.EU Core Vocabularies: the Core Person Case
<http://www.semic.eu/semic/view/documents/Case-Study-Core-Vocabularies.pdf>
- DC** Dublin Core Metadata Initiative, <http://dublincore.org/>
- DOAP** Description of a Project, Edd Dumbill.
<http://trac.usefulinc.com/doap>
- EBR** European Business Register, <http://www.ebr.org/>
- EGMS** e-Government Metadata Standard, <http://www.esd.org.uk/standards/egms/>
- EGOV-CV** e-Government Core Vocabularies: The SEMIC.EU approach. Retrieved from European Commission - Directorate-General Informatics:
<http://www.semic.eu/semic/view/documents/egov-core-vocabularies.pdf>
- EIF** European Interoperability Framework (EIF) for European public services. [Online] 12 2010.
http://ec.europa.eu/isa/strategy/doc/annex_ii_eif_en.pdf.
- EUPL** The European Union Public Licence, Patrice-Emmanuel Schmitz
<http://www.osor.eu/eupl>
- EUVOC** EuroVoc, the EU's multilingual thesaurus. <http://eurovoc.europa.eu/>
- FOAF** Friend of a Friend, Dan Brickley and Libby Miller.
<http://xmlns.com/foaf/spec/>
- GR** Good Relations Vocabulary, Martin Hepp
<http://purl.org/goodrelations/>

-
- IEEE.** Develop Standards. *IEEE Standards Association*. [Online] [Cited: 22 August 2011.]
<http://standards.ieee.org/develop/process.html>.
- INSPIRE** Infrastructure for Spatial Information in the European Community, EU Directive.
<http://inspire.jrc.ec.europa.eu/>
- IPSV** Integrated Public Sector Vocabulary. <http://doc.esd.org.uk/IPSV/>
- ISA** Interoperability Solutions for European Public Administrations
<http://ec.europa.eu/isa/>
- ISABR** ISA Access To Base registries.
http://ec.europa.eu/isa/actions/documents/isa_1.2_access_to_base_registers_workprogramme.pdf
- LD** Linked Data: Evolving the Web into a Global Data Space, Tom Heath and Chris Bizer, 2011.
<http://linkeddatabook.com/editions/1.0/#sec:reusingTerms>
- LDORG** Linked Data - Connect Distributed Data across the Web
<http://linkeddata.org/>
- OASIS. 2010.** Technical Committee (TC) Process. *OASIS | Advancing open standards for the information society*. [Online] 28 July 2010. [Cited: 22 August 2011.] <http://www.oasis-open.org/policies-guidelines/tc-process>.
- OC** Open Corporates, <http://opencorporates.com/>
- ORE** Open Archives Initiative Object Reuse and Exchange
<http://www.openarchives.org/ore/>
- ORG** An Organization Ontology, Dave Reynolds, Epimorphics, 8 October 2010.
<http://www.epimorphics.com/public/vocabulary/org.html>
- OWMS** Overheid.nl Web Metadata Standaard <http://standaarden.overheid.nl/owms/4.0/doc/>
- PG** Polyglot Markup: HTML-Compatible XHTML Documents, Eliot Graff. currently a W3C Working Draft, May 2011 <http://www.w3.org/TR/2011/WD-html-polyglot-20110525/>
- PNAW** Personal names around the world, Richard Ishida, W3C, August 2011
<http://www.w3.org/International/questions/qa-personal-names>
- PURLs** software is available from <http://sites.google.com/site/persistenturls/>.
- RDFBP** Best Practice Recipes for Publishing RDF Vocabularies, Diego Berueta and Jon Phipps, W3C Note August 2008. <http://www.w3.org/TR/swbp-vocab-pub/>
- SCHEMA.ORG** Full Type Hierarchy.
<http://www.schema.org/docs/full.html>
- SEMIC** The Semantic Interoperability Centre Europe
<http://www.semic.eu/>
- SKOS** Simple Knowledge Organization System Reference, Alistair Miles and Sean Bechhofer. W3C Recommendation 18 August 2009 <http://www.w3.org/TR/skos-reference>
- TIE** Trusted Information Exchange, an EC initiative under the ISA.
http://ec.europa.eu/isa/actions/01-trusted-information-exchange/index_en.htm
- UKOGL** UK Open Government Licence for Public Sector Information, The National Archives
<http://www.nationalarchives.gov.uk/doc/open-government-licence/open-government-licence.htm>
- UKP** Rt Hon David Cameron's entry on the UK Parliament Web site
<http://www.parliament.uk/biographies/david-cameron/25752>
- UML** Unified Modeling Language, OMG. <http://www.omg.org/spec/UML/>

- UN/CEFACT** Core Component Library (UN/CCL). *UNECE - United Nations Economic Commission for Europe*. [Online] 2011.
http://www.unece.org/cefact/codesfortrade/unccl/CCL_index.htm.
- vCARD** The Internet Mail Consortium, <http://www.imc.org/pdi/vcardoverview.html>
- W3CDTF** Date and Time Formats, Misha Wolf and Charles Wicksteed, September 1997.
<http://www.w3.org/TR/NOTE-datetime>
- W3CPROC** World Wide Web Consortium Process Document Ian Jacobs, 14 October 2005.
<http://www.w3.org/Consortium/Process/>
- W3CPP** World Wide Web Consortium patent Policy 2004. D. J. Weitzner
<http://www.w3.org/Consortium/Patent-Policy/>
- WGS84** Basic Geo (WGS84 lat/long) Vocabulary, Dan Brickley
<http://www.w3.org/2003/01/geo/>
- XML** Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C Recommendation 26 November 2008. Tim Bray et al. <http://www.w3.org/TR/xml/>

APPENDIX C TERMINOLOGY AND GLOSSARY

This document uses various terms to describe types of resource and refers to European institutions and activities. These are defined or explained in alphabetical order below.

Core Concepts. Core Concepts are simplified, reusable, and extensible data models that capture the fundamental characteristics of an entity in a context-neutral and language-neutral fashion. They are the starting point for agreeing on new semantic interoperability assets and defining mappings between existing assets.

Core Vocabulary. A Core Vocabulary is a Core Concept, represented in a particular representation language, such as an RDF or XML schema.

Descriptions of semantic interoperability assets are the asset descriptions contained in and made available from semantic interoperability repositories. ADMS, the Asset Description Metadata Schema, is useful in this regard [ADMS].

Human-readable semantic asset. A semantic asset is *human-readable* if it is documented in natural language, such that the meaning of the concepts it defines (e.g. classes, properties, relationships, codes, etc) can be understood by humans.

ISA Programme. The Interoperability Solutions for European Public Administrations programme was launched by the European Commission to run from 2010-2015. Its objective is to support cooperation between European public administrations by facilitating the efficient and effective cross-border and cross-sector interactions between such administrations, enabling the delivery of electronic public services supporting the implementation of Community policies and activities.

Machine-readable semantic asset. A semantic asset is *machine-readable* if it is documented in a format that can be meaningfully processed by a computer application, such that the concepts it represents (e.g. classes, properties, relationships, codes, etc) can be identified by the computer application.

Metadata in this context means schemata or generic data models.

Open semantic asset. Similar to open-source software (as defined by the Free Software Foundation), open semantic assets are semantic assets with an open license that prevents anyone from exploiting it exclusively. An open license guarantees a number of principles, such as the freedom to use the semantic asset, the freedom to extend or adapt the semantic asset, and the freedom to redistribute the semantic asset. Full access to the documentation of the semantic asset is often a precondition for this.

Primary data resources include documents, services, software, data in spreadsheets, maps etc. This is a very general term for what might be considered 'e-Government data.'

Reference data are value vocabularies (e.g. code lists, taxonomies) or instance vocabularies.

Semantic asset catalogue/registry. A semantic asset catalogue/registry is a infrastructure where *metadata* about semantic assets are documented and retrieved, whereas the semantic assets themselves are located elsewhere.

Semantic asset management. Semantic asset management refers to the good practice of adopting policies, processes, and systems to plan, perform, evaluate, and improve the reuse of semantic assets. It ensures that metadata and reference data are treated as valuable assets, stored or documented in a central location, easily retrievable, and accessible in a reusable format by third parties.

Semantic asset repository. A semantic asset repository is a storage infrastructure where semantic assets can be stored, documented and retrieved.

Semantic interoperability. The European Interoperability Framework [EIF] defines semantic interoperability as *the ability of organisations to process information from external sources in a meaningful manner*. It ensures that the precise meaning of exchanged information is understood and preserved throughout exchanges between parties. According to EIF, semantic interoperability encompasses semantic and syntactic aspects. Semantic interoperability is referred to the meaning of data elements and the relationship between them. It entails developing a vocabulary to describe data exchanges, and ensures that the data elements are understood in the same way by communicating receivers. On the other hand, syntactic interoperability is about describing the exact format of the information to be exchanged in terms of grammar, format and schemas.

Semantic interoperability asset. A semantic asset is a collection of highly reusable metadata (e.g. xml schemata, generic data models) and reference data (e.g. codelists, taxonomies, dictionaries, vocabularies) which are used for e-Government system development. The following reference models and reference data are considered to be semantic assets:

- **Core Vocabularies** (e.g. Dublin Core [DC]) ;
- **metadata specifications and schemas** (e.g. UK e-GMS [EGMS], NL OWMS [OWMS], INSPIRE [INSPIRE]);
- **controlled vocabularies or taxonomies** (e.g. subjects, domains, document types, security levels; examples are Eurovoc [EUVOC], Agrovoc [AGVOC], UK Integrated Public Sector Vocabulary [IPSV]);
- **Mapping specifications** (conversions across controlled vocabularies and metadata schemas).
- **Reference collections of organisations and persons** (e.g. registers of government agencies, business registers, citizen databases)
- **References to geographic locations and jurisdictions** (e.g. cadastre, spatial planning, map information)
- **Code lists** (e.g. language codes, country and region codes such as ISO 639, ISO 3166, NUTS)

- **Syntax specifications** (e.g. W3C Date and Time format [W3CDTF], ISO 8601)

Trusted Information Exchange (TIE) cluster. The TIE Cluster is a cluster of actions within the ISA Programme which aim at facilitating Trusted Information Exchange between Public Administrations in Europe.

APPENDIX D SUMMARY OF PROCESS AND METHOD STEPS

Process Step 1.	Identify Stakeholders. The EC identifies organisations and individuals that are likely to have the most need for, and benefit most from, the proposed Core Vocabulary. The EC invites them to participate in the process: ideally by joining the Working Group or at least by subscribing to the public mailing list and reviewing published drafts.....	10
Process Step 2.	Form Working Group. The EC facilitates the formation of the Working Group and announces its formation to the community and the public at large.	11
Process Step 3.	Identify Chair / Co-chairs. The EC identifies the chair or co-chairs.....	11
Process Step 4.	Identify Editor(s). The chair identifies the editor(s) of the Core Vocabulary document.	11
Process Step 5.	Form Review Group. The EC identifies members of the Review Group and advises them of the projected date by which the Core Vocabulary will be available for their review. If that date changes, then the Working Group chair, working through the EC, will advise the Review Group accordingly.	12
Process Step 6.	Secure IPR. On joining the Working Group or Review Group, the EC will advise members that contributions will be according to the Contributor Licence Agreement and that the final specification will be published under an open licence. Members are asked to (electronically) sign the Contributor Licence Agreement.....	16
Process Step 7.	Establish a working environment and culture. Establish a working environment and agree on a working culture.	18
Process Step 8.	Publish drafts. Publish drafts of the core vocabulary as it evolves, seeking feedback from the community on each occasion.	21
Process Step 9.	Process Comments. Consider comments received concerning the interim published drafts.	22
Process Step 10.	Publish the Last Call Working Draft of the vocabulary and contact the Review Group, seeking its feedback. The wider community should also be invited to comment. The call should set a date by which comments must be received.	23
Process Step 11.	Review Last Call Working Draft.....	24
Process Step 12.	Gather evidence of acceptance by the potential users of the vocabulary. .	25
Process Step 13.	The EC to submit documents to the TIE Cluster or ISA Coordination Group.	26

Method Step 1.	Identify the Core Vocabularies that are likely to meet the most pressing needs of the potential users within European institutions and for which the relevant expertise is most readily available.	28
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).	31
Method Step 3.	Research existing published data and services, noting that there should not be any conflicts with the proposed Core Vocabulary. As with the research on existing vocabularies, this research should inform the writing of use cases.....	32
Method Step 4.	Articulate the problem(s) that the Working Group is trying to solve in the form of a series of use cases.	32
Method Step 5.	Derive a set of requirements from the use cases.	33
Method Step 6.	Publish the use cases and requirements in a single document.	33
Method Step 7.	Unless the vocabulary is very simple with no more than a handful of concepts, create a concept diagram. Consider using UML.	33
Method Step 8.	Do not impose cardinality rules or domain/range restrictions on vocabulary terms unless strictly necessary.	34
Method Step 9.	Use words beginning with an upper or lower case letter (A - Z, a - z) or an underscore (_) for all terms in a vocabulary.	35
Method Step 10.	Use simple nouns for property names.	35
Method Step 11.	User verbs for relationship terms.	35
Method Step 12.	For each relationship, include a definition of its inverse.	36
Method Step 13.	Use prepositions in vocabulary terms only if necessary.	36
Method Step 14.	Use a namespace ending with a hash character (#)	36
Method Step 15.	Keep the namespace as short as possible.	37
Method Step 16.	Include a portion that identifies the vocabulary for human readers.	37
Method Step 17.	Do not include any technology-specific component in the namespace (except HTTP).	37
Method Step 18.	When choosing a namespace, do not restrict the pool of potential users unnecessarily by using a namespace that declares 'ownership' or geographical relevance.	37
Method Step 19.	If necessary, consider meeting Method Step 18 using PURLs	38

Method Step 20.	Create and validate the namespace documents in HTML, XML and RDF/XML. Consider also serialising the RDF schema in Turtle.....	38
Method Step 21.	Either the Working Group or the EC must make each one available through {namespace}.ext where .ext is the relevant extension.	38
Method Step 22.	Either the Working Group or the EC must set up content negotiation to handle requests to the namespace itself.	38
Method Step 23.	When publishing the final version of the Core Vocabulary, link the HTML version to an errata document.	38
Method Step 24.	Include a Conformance Statement.	39