ECP 2007 EDU 417008

ASPECT

ASPECT Approach To Multilingual Vocabularies, Including Automated Translation Services

Deliverable number	*D2.3*
Dissemination level	*Public*
Delivery date	*23 February 2009*
Status	*Final*
Author(s)	*Mike Collett (VMG),* *Peter Collins (VMG),* *Neil Smith (VMG),* *Rob Tice (VMG)*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

0 Introduction

For the purposes of this document, we define a vocabulary as a set of terms (being either individual words or phrases). They may be explicitly defined, both in regards to their own meaning, and in the relationships they hold with other terms, and thereby organized into some form of structure. A vocabulary is therefore a structure containing terms, varying in complexity from simple lists and hierarchical taxonomies, to more extensive networks of relationships and associations such as in thesauri and ontologies.
Relationships are provided in the descriptive metadata for each term that can be structured in one of several standardized exchange formats.

In the Dublin Core Metadata Initiative Abstract Model (http://dublincore.org/documents/2007/06/04/abstract-model/) a vocabulary is also defined as "a set of one or more terms" where a "term is a property (element), class, vocabulary encoding scheme, or syntax encoding scheme." This does not conflict with the broader ASPECT view.

Part of the Vocabulary Management Group's (VMG's) contribution to the ASPECT project is the selection of appropriate formats for representing vocabularies based on their suitability in supporting interoperability between controlled vocabularies themselves and in the systems that utilize them for classification, indexing and retrieval purposes. Of particular interest, of course, is the manner in which these formats support and enhance multilinguality within and between vocabularies, and semantic interoperability in general.

VMG will also be developing and implementing the ASPECT Vocabulary Bank for Education (VBE), which is defined as a repository in which multilingual terms and vocabularies can be published and disseminated. The VBE will be a web application, written in Java using the struts and tiles framework, that will enable participants to search or browse vocabularies to determine information about individual terms, their relationships and their context, and to create new mappings between terms. The VBE will be designed to facilitate import and export of vocabularies in a range of standard exchange formats, and will manage and retain all historical information relating to terms and vocabularies and their revisions.

Although the ASPECT service model does not explicitly allow for the provision of tools for the production and editing of vocabularies, it is recognised that such tools may be useful for some ASPECT partners. A number of tools are reviewed and their utility analyzed with respect to the requirements of the bank service.

1 State of the art analysis

1.1 Vocabulary Representation & Exchange

This section reviews the current position within the target market with respect to the availability of controlled vocabularies. A clear distinction is made between formats which have been specifically designed for the representation of controlled vocabularies and may be used for two way exchange of information (i.e. both into and out of the VBE) and those which are suitable for one way exchange (i.e. for loading vocabularies into the VBE but not for dissemination outwards). The goal is to support the ingestion of as wide a variety of formats as is practical and required but to limit support for dissemination formats to a small number of standardised formats which meet all or most of the requirements set out in section 2.1.

1.1.1 Vocabulary formats in use today

There are a number of formats currently in use within different communities for the representation and exchange of vocabularies.

Within the related domain of bibliographic services, OCLC are developing experimental terminology services which to some extent parallel what the VBE intends to provide for the European Schools elearning domain. The pilot service, available at http://tspilot.oclc.org/resources, provides a useful reference point for VBE. In their initial analysis, OCLC considered four vocabulary formats:

- Z39.19
- MARC21 Authorities
- SKOS
- Zthes

Of these, Z39.19 is a NISO standard with limited European uptake, although it is broadly equivalent in format and coverage to ISO 2788/ BS 5723. MARC21 is a specifically bibliographic format with limited relevance outside of this sector. It is also a pre-web binary format (although there is an XML equivalent – MARC XML). The OCLC pilot service supports retrieval of concepts/headings in SKOS and Zthes as well as MARC XML and human readable HTML.

A similar parallel development has been the HILT project which focuses on the UK Higher Education sector. Phase III of this project, which ended in February 2007, focussed on specifications for machine to machine interfaces and recommended SKOS as the major interchange format with the service being built in such as way that support for Zthes and MARC could be added at a later stage. More information on the HILT project is available at http://hilt.cdlr.strath.ac.uk/.

Consultation with ASPECT project partners along with desk based research has revealed that the main formats for representing and exchanging vocabularies which have relevance to the European schools sector are:

- XVD – this was developed as part of a CEN Workshop Agreement (CWA). It has been used to provide EUN vocabularies but is not widely used elsewhere.
- Zthes – this originated from work related to the Z39.50 search protocol and is a widely used specification internationally, especially in library communities. Several versions of Zthes exist – for example an extended version of Zthes v1.0 is used extensively in the UK as part of the Becta Vocabulary Management Service.
- CEF – A new European Standard for a Curriculum Exchange Format (currently at a draft stage) supported by a CWA. This is being developed specifically to support use cases such as those of ASPECT. It is based on an extended version of Zthes.
- VDEX – an IMS specification that has been adopted by several educational stakeholders across Europe.
- SKOS – a semantic web exchange format. This has fairly wide adoption and support from W3C. Created to be simpler than the OWL Web Ontology Language that also represents the meaning of terms and the relationships between them.
- XTM – a topic map exchange format used particularly in Norway and Sweden to express curricula information.
- DD8723 part 5 – this is a draft for development suggested by BSI in the UK. It can be considered for import or export once there are vocabularies published in this format or it becomes an approved and stable standard.

There are several standards that have been produced that describe thesauri from a structural and display perspective. These do not provide an exchange format but can be useful when constructing vocabularies. Whilst these do not provide an exchange format they can be useful when constructing vocabularies (e.g. ISO 2788-1986 and ISO 5964-1985 which are guidelines for the establishment and development of monolingual and multilingual thesauri respectively). Recently there is new work planned with a committee draft submitted to ISO (CD 25964) that is based on the five part standard BS8723 and is intended to extend and replace ISO 2788-1986 and ISO 5964-1985.

In addition to these and other ‘standardised’ formats, it is clear that the majority of stakeholders within the ASPECT community do use controlled vocabularies within their systems to some extent. However, these vocabularies are normally held within particular software products (e.g. Content Management Systems) in a native or proprietary format. Since most systems provide some form of export function it would also be useful if the Bank service could support a range of ‘generic’ import formats in order to lower the barriers for stakeholders to easily provide relatively simple controlled vocabularies. Such formats might include:

- delimited text files (e.g. .csv)
- spreadsheet files (e.g. .xls, .xlsx, .ods)
- proprietary xml formats
- rich document formats (e.g. .rtf, .doc, .docx, .odt)

1.1.2 Advantages and disadvantages

The short list of formats identified in section 1.1.1 above has been analyzed in terms of the requirements of the ASPECT Vocabulary Bank described on section 2.1 below. Table 1 summarises the outcomes of this analysis.

Format	Polyhierarchy	Mapping between terms	Multilingual	data transfer optimisation	Specific Curriculum information	Identifiable non-preferred terms	Controlled references (facets)	Compound relations	Extensible
XVD	yes	yes	yes	no	no	yes	no	no	yes
Zthes 1.0	yes	yes	no	yes	no	yes	yes	no	yes
CEF	yes	yes	yes	yes	yes	yes	yes	yes	yes
VDEX	yes	no	yes	no	no	yes	no	no	yes
SKOS	yes	yes	yes	no	no	no	no	no	yes
XTM	yes	no	yes	no	no	yes	part	no	yes
DD BS8723	yes	yes	yes	no	no	yes	no	part	yes

Table 1 *Vocabulary formats vs ASPECT requirements*

Although mapping between terms in Zthes and CEF is not specifically mentioned as part of the specification it can be accomplished as follows

- Reuse of terminology providing a mapping by default
- Creation of specific mapping vocabularies which re-use existing terminology and are purely designed to show inter-vocabulary relationships

As can be seen from the analysis above, the CEF extension to Zthes is the best fit for the requirements of ASPECT. This is not surprising as it was developed under the auspices of CEN-ISSS WS/LT with these requirements in mind. Of the other formats, SKOS would appear to be the next priority as it is widely used by other services such as OCLC-TS and HILT followed by XTM and Zthes 1.0 (for backward compatibility).

1.2 Search, Retrieval and data access

The VBE needs to be searched and browsed by both machines and human visitors. It is important that searching is standards based and it is also important that all vocabularies, authorities, terms, properties etc are available using logical and persistent URL syntax. It should be possible to search system wide or to restrict searches to particular vocabularies.

Management features should be accessible using restful interfaces and it should be possible to ascertain everything about the history of a term and/or vocabulary via these interfaces. The application itself should utilise the same interfaces that are available to third parties. Terminology should be searchable using a variety of access points which are configurable based on the needs of the service and these access points should be able to be configured in response to ingested data formats. The searching structure should be scalable as it is envisaged that the VBE bank may evolve to contain large amounts of data. The machine interface must utilise recognised open standards with wide cross domain uptake.

There are a wide range of standardised query interface specifications in use at the moment. In 2007, a consortium including European Schoolnet and several members of the VMG team involved in ASPECT produced a wide ranging review of Federated Resource Discovery Services for Becta. As part of this research, 16 query specifications were analyzed:

- Search Service Specifications
- SRW (Search/Retrieve Web Service)
- SRU (Search/Retrieve via URL)
- Z39.50
- SQI (Simple Query Interface)
- OpenSearch
- NISO Metasearch Specifications
    - NISO Z39.92-200x, Information Retrieval Service Description Specification
    - NISO RP-2006-02, NISO Metasearch XML Gateway Implementers Guide
- Google (Ajax) and Google Base
    - Google AJAX Search API
    - Google Base
- Google Scholar
- Yahoo!
- Amazon
- Vivisimo
- Scholar SFX
- WebFeat
- LIMBS
- IMS DRI (ECL implementation)
- ebXML

These specifications are described in detail in Annex 1 of deliverable D-2.1 within WP2 (in an extract taken from a Becta report). Analysis of the relevance of these specifications to the requirements of the VBE leads to three main specifications being considered for implementation:

- SRU
- SQI
- OpenSearch

Of these, SRU has the best fit with the requirements of the VBE in terms of being a truly open standard with widespread support and rich search semantics. It is important to remember that VBE should be considered as a service to be utilised by all sectors and across domains and the primary interface should reflect this. If there is a project requirement for the VBE to be accessed using SQI, this could be addressed using the SQI to SRU bridge as described in section 4.2.1 of deliverable D-2.1 within WP2. This is described as an existing part of the

infrastructure and would seem to be the most sensible approach as it requires no changes to the VBE.

In terms of integration with publisher systems, a case could be made for provision of an OpenSearch interface which is lighter weight but has a greater uptake within the publishing domain.

1.3 Vocabulary Upload/ Ingest

There is a requirement for a machine interface to enable vocabularies to be uploaded and ingested into the VBE. Where possible, this should be based on standards compliant protocols.
Section 3.4 in deliverable D-2.1 within WP2 lists the following publication options:

- OAI-ORE
- RSS
- SWORD
- SPI

Of these, the most appropriate for use as part of VBE vocabulary upload are:

- A profile of the Atom Publishing Protocol (as currently profiled in SWORD or CMIS for example).
- Current and emerging work on SPI (Simple Publishing Interface)

As part of ongoing consultation with SPECT partners, possible use cases which might lead to a case being made for other mechanisms (e.g. OAI-PMH harvesting) will be elicited and considered for inclusion at a later stage.

See section 3.4 in deliverable D-2.1 within WP2 for a description of the protocols.

1.4 Current awareness

A number of similar services provide mechanisms by which third party systems can be alerted to changes, e.g. the addition of new vocabularies or updates to existing vocabularies. Mechanisms for alerting include ‘pull’ based protocols (e.g. ‘news feeds’ in RSS and/or ATOM syndication format and ‘push’ based mechanisms (e.g. email). The disadvantage of ‘pull’ based mechanisms is that they generally require the third party to register some details (interests, email address, etc.) with the host service.

See section 3.4 in deliverable D-2.1 within WP2 for a description of the RSS protocol.

1.5 Existing Services & Features

1.5.1 Vocabulary publication tools and services

This section reviews existing tools and services for the publication of controlled vocabularies. In addition to briefly describing each service in Annex 1, a summary is given of the extent to which the standardised vocabulary formats and search interfaces described in section 1.2 above are supported.

Our survey into vocabulary publication tools currently in use indicates that the majority have been developed to fulfil sector-specific terminology services, providing centralized hosting, terminology mapping and dissemination of terminologies or vocabularies specific to their participants or industry sector. Indeed, this list has been limited to those systems that appear to hold reasonably generic vocabulary exchange capabilities; several other systems that are solely applicable to the sector they serve (with notable examples being in the field of medical terminology) have been excluded. This focus on specific industry needs, or the research nature of some examples, is likely to have led to the selection of the vocabulary exchange format(s) employed, with no one format becoming predominant. A small number of systems identified are being supplied commercially, placing greater emphasis on compatibility with a wider range of generic import / export formats. Two systems identified have been developed to perform vocabulary publication services in the Education sector, both in the United Kingdom, and yet both responding to different needs for exchange standard compliance.

This survey was limited to an English language based search in this field, but nevertheless illustrates a cross section of international projects and services in vocabulary / terminology hosting and management, in which deployments are still typically in their earliest iteration and generally fulfilling monolingual vocabulary requirements. In basing the VBE on VMG's Lexaurus Bank, which is itself a 'second generation' bank service, ASPECT will benefit from existing advances pertinent to vocabulary management in the Education sector, and be well placed to address the broad requirements in exchange formats and multilinguality of its European audience.

1.5.2 Editing tools

The provision of tools to edit vocabularies is not within scope of the VBE service. In the majority of cases, vocabularies will be downloaded from the VBE and used 'as is' within a software tool such as a CMS. In the event that editing a downloaded vocabulary is required, this will normally take place within that software environment. Similarly, tools for creating vocabularies to be uploaded to the VBE will normally exist already.

Nonetheless, it would be useful to ensure that tools are available to support editing of some or all of the supported vocabulary formats to promote wide re-use and re-purposing. Whilst it is possible to edit any XML based format using a generic XML (or even text) editor, specialist editing tools are far more user friendly and less prone to errors.

A survey of the market place has identified a range of 'stand alone' editing tools as being in general use, including both commercial and open source tools. These are summarised briefly in Annex 2, including a brief description of the tool and an indication of the extent to which the tool publishers claim adherence to standards.

This survey (via English language based searching) illustrates that while a far greater number of products of this category are in use internationally (and this list is certainly not definitive) the adoption of import / export formats in editing tools is even wider than is the case with publication tools. This obviously reflects the broad range of tasks achievable through the creation of controlled vocabularies in all their various forms, and the growing maturity of this product type. All of the above listed products are examples of 'stand alone' editors, although some provide pairings with co-developed publication banks.

Product suites of this kind benefit from common operational approaches and direct data exchange, and are an obvious choice where a publication bank is already established. In this respect, VMG's own editing tool, Lexaurus Editor, will be a logical choice for VBE users

without an existing editor. VMG will therefore make its Lexaurus Editor available, to ASPECT partners that require it, for the duration of the project, in cases where this will lead to early development and publication of multilingual vocabularies into the ASPECT Vocabulary Bank.

2 ASPECT Vocabulary Bank for Education – technical specification

2.1 Requirements

These requirements are based on the requirements defined in the CEN/ISSS WS-LT draft CWA - Curriculum Exchange Format, CEF.

2.1.1 Polyhierarchy

It shall be possible for terms (curriculum items) to exist at more than one place in a structure or vocabulary.

2.1.2 Terminology Reuse

It shall be possible for terms (curriculum items) to be reused in multiple vocabularies.

2.1.3 Identifiable non preferred terms

When mapping between different terminology sets and managing change between revisions it must be possible to identify relationships to and changes which result in a non-preferred term being created.

2.1.4 Multilingualism

The format shall support multilingualism. Multilingualism can be considered to have two separate requirements. There is a need to indicate linguistic equivalence between different terms as well as the requirement to provide multilingual 'labels' for metadata properties of either terms or vocabularies (e.g. term name, description etc).

2.1.5 Optimal data transfer

Data transfer:
Can be required for a variety of use cases and these use cases can themselves have an impact on the verbosity of the interchange format used. It is important that the interchange format (and therefore the service) is configurable to allow the correct choice of content for a particular use case. As an example, when transferring a complete thesaurus the requirement is only to reduce the amount of redundant information to minimise the file size. However, if the data transfer is to support browsing it may be acceptable to allow a small overhead for each data packet in order to minimise the number of separate calls to the service.

Partial transfer:
When transferring data verbatim between producing and consuming systems it is essential that the modelling of this transfer allows for large scale data interchange. In these situations the physical size of the data files can have an impact on the transfer medium used (e.g. http) and the capability of the receiving system to actually ingest the data. It is thereof a

requirement that the system can transfer data in chunks of ‘n’ terms (where n is governed by the receiving system)

2.1.6 Controlled references (facets)

It is essential that the proposed solution allows different contributors to describe facets of their terminology. As an example in the de-facto UK interchange format there is the concept of ‘curriculum type’ which is used to differentiate between the data types in the curriculum structure. Each contributing authority must be allowed to contribute their own controlled references even if these references are later consolidated.

2.1.7 Compound relationships

It should be possible to express compound relationships between items. (i.e. a relationship which requires more than one target). Examples of this include;

- A competency relationship may have parts that include both an action and a topic.
- A linguistic equivalence may require a combination of terms to fully describe a term in a different language.

2.1.8 Extensibility

One of the most difficult areas to predict is that of extensibility. Many approaches to extensibility have been investigated but the fact remains that it is difficult to promote absolute extensibility at the same time as easy interoperability. The distinction between rigorous description and easy interchange has been historically difficult to reconcile. It is therefore essential that there is a controlled mechanism for adding local extensions to the format of terms and relationships without breaking basic interoperability.

Metadata considerations aside, it is a fundamental requirement that the internal model of the terminology structure is format-independent. It is impossible to predict what may be available in the future therefore it is essential that the solution is able to model as many different formats as possible.

2.2 Input formats

Following on from the previous section, the choice of input formats should be driven by a need to be representative of the most currently supported formats. As the project progresses the chosen formats should be able to react in response to new standards, domain uptake and vocabulary availability.

The following have been selected from a combination of their domain usage and their relevance to the project. Whilst it must be remembered that the bank should be format independent and should not present a barrier to those who wish to participate, this must be tempered with the previously mentioned recognition of the most widely used formats.

The choices for inclusion (in no specific order) are as follows (together with their proposed implementation stage):

- Zthes (Bank 1)
- CEF (Bank 1 – dependent on agreement of the standard and profiles)
- VDEX (Bank 1 - with an agreed profile)
- XVD (Bank 1 – subject to data supply and description)
- SKOS core (Bank 1 – subject to data supply and description)

- XTM (Bank 2 – subject to data supply and description)?

As consultation with publishers in WP3 has shown, many publishers utilise their own locally controlled vocabularies which are not created in any standardised format. Rather than see this as simply a barrier, it is proposed that the project recognises that this may be representative and utilises this disparate data. It may be possible to create an 'import wizard' to allow these bespoke vocabularies and formats to be imported into the bank without specific configuration or transformation and this will be investigated throughout the duration of the project with the corpus of data provided.

2.3 Validation and conformance

Within the data model for the service vocabularies can be loosely coupled to schemas which describe the metadata about a term or a vocabulary (e.g. identifier, term name, authority etc). This allows validation to be performed after ingestion against whichever schema is subsequently used to define to element templates.

In the initial VBE release validation is limited to ensuring that document is well formed, however for subsequent releases it will be possible to run a report for an ingested vocabulary to indicate any non-conformances.

2.4 Output formats

The following formats are selected for the vocabulary bank to be able to **export**. They have been selected from a combination of their domain usage, their relevance to the project and the mechanisms available for viewing.

The choices for inclusion (in no specific order) are as follows (together with their proposed implementation stage):

- Zthes (Bank 1)
- Nested xml (Bank 1)
- CEF (Bank 1 – dependent on agreement of the standard, schema and an ASPECT profile of it)
- VDEX (Bank 1)
- XVD (Bank 2 subject to investigation of representation in CEF)
- SKOS (Bank 2)
- XTM (Bank 2)

The decision on having XVD as an input and/or output format depends largely on whether the LRE thesaurus can be represented in any of the other formats.

Other platform specific formats should be considered, e.g. mindmap, VISIO, Omnigraffle etc and their suitability can be ascertained for Bank 3 in conjunction with uptake and perceived need.

Most formats which are designed for efficient transfer replace any implied hierarchy or network with a flat transfer format to avoid repetition. This is efficient for interchange but less so for visualisation and application of local formatting. It is therefore important that a nested format is provided for easy transfer into document formats (e.g. html etc)

2.5 Data Augmentation

One of the key areas in terminology management is identification and the adoption of consistent and non-conflicting identification mechanisms. It is tempting to suggest that everything should simply be a resolvable URL. However, this assumes persistence of the domain which has been proven to be a premature assumption (at least within UK education).

To simplify the process of the allocation of non-conflicting identifiers, an identification service should be provided which communicates non-conflicting identifiers to the VBE. This environment should allow any participating organisation to plug in their own service so that allocation is consistent with their own internal requirements as well as providing the required uniqueness.

Given the multi-lingual requirement of the bank and the predominance of mono-lingual vocabularies, the ability to reference an external (pluggable) translation service should be included. The access should be API independent but one will be decided on and used initially. Although it is recognised that auto translation is a difficult area when applied to a wide-ranging, uncontrolled domain its application to if to the targeted education domain and specifically to controlled terminology within the VBE may well be beneficial.

2.6 Identification scheme/service

When new terminology is created in many different locations by many different participants it can be important to ensure that identifiers are unique. The VBE system only requires uniqueness of identification within each participating authority, however, providing a service which allows participants own identification services to plug in to the VBE allocation service can be beneficial. As well as brokering to participants services it can provide generic id allocation for those who wish to use it (e.g. a simple numeric id).

We therefore propose that the deployment of a plug-in based service which brokers id allocation is provided as part of the VBE in release 2. The initial deployment in release 2 will provide a simple numeric service with plug-ins provided as per participants' requirements for release 3.

2.7 User Interface

The human interface should follow the latest w3c accessibility criteria for web based applications. The initial VBE release will utilise the current interface as provided by the chosen system with scope to modify the interface during the project in response to evolving requirements.

2.8 Machine interfaces

2.8.1 Query

As previously described, the initial query interface will be based on SRU. It is important that all elements of ingested terminology are searchable but it is suggested that the initial release be limited to:
Terms:

- Title
- Description/scope
- Term Type

- Curriculum type

Vocabularies:

- Title
- Description
- Authority

Adding additional access points will simply be governed by choosing the correct balance between inclusivity and ease of use.

In addition static URL's should provide access to the following

- Term types
- Relation types
- Term and vocabulary metadata properties
- Term and vocabulary metadata schemas
- Profiles
- Authorities
- All vocabularies
- All vocabularies in an authority
- Specific vocabularies
- Specific terms within a vocabulary
- Top terms within a vocabulary

2.8.2 Download

Vocabulary downloads (and partial vocabulary downloads), deltas (differences between versions) and revision numbers should also all be accessible using restful calls.

2.8.3 Upload

The upload mechanism which will be implemented in the initial release is a simple file upload protocol using http POST. As part of the ongoing development, suitable standards based options and protocols for upload will be investigated (and implemented if there is a consensus).

2.8.4 Other

Browse controls should utilise de facto standards (e.g. JSON).

For current awareness/ alerting, we propose an investigation of mechanisms by which registered users can express an interest in particular individual vocabularies (or sets of vocabularies) and be alerted when changes occur. As well as providing a current awareness feed which systems can consume to be alerted of changes. The proposed mechanisms at the moment are

- Email alerts to registered users
- RSS feed to indicate which vocabularies have changed

2.9 Workflow

This section describe how vocabularies are expected to be entered into the bank and how they are accessed and modified. The precise requirements may be modified in light of experience and feedback.

The application should distinguish between registered and unregistered users.

The application will make different features available to registered users depending upon their role and will differentiate between users who are registered and logged in someone who is not logged in. In addition registered users will be allowed to request a new role within the system.

As you navigate down the following list each different role receives all the features of the previous roles plus the specific features added.

User roles:

- Unregistered
  - Browse all published vocabularies
- User
  - Download published vocabularies in all formats
  - Download deltas (differences between vocabulary revisions)
  - Request an elevated role within the system
- Editor
  - Upload new revisions of vocabularies that they have permission to edit
  - Delete vocabularies that they have permission to edit.
- Organisational Administrator
  - Publish vocabularies under the control of the organisation
  - Create authorities under the control of their organisation
  - Request the ability to share an authority currently used by another organisation
  - Manage role assignment requests within their organisation (i.e. a user registers and wishes to be allocated the role of editor)
  - Manage user roles within their organisation
  - Assign vocabulary edit permissions to editorial users
- System Administrator
  - Manage role assignment (up to administrator level) system wide.
  - Manage role assignment requests (up to System administrator) system wide.
  - Allocate sharing of previously created authorities.
  - Manage organisations

3 Timetable for the Aspect Vocabulary Bank

This section provides a development timetable with tasks and deliverables, within the overall framework of ASPECT Work Package 2 plan. The development plan is based around the 3 versions of infrastructure and services as specified in D2.5 – D2.7.

Based on the technical specification set out in section 2, it is proposed that an initial version of the Aspect Vocabulary Bank, VBE v1.0, is released as part of Deliverable 2.5 (Infrastructure and Services v1.0) with enhancements delivered according priority as part of Deliverable 2.6 (Infrastructure and Services v2.0) and Deliverable 2.6 (Infrastructure and Services v3.0).

The proposed content of each version is set out below. The content of VBE v1.0 should be regarded as ‘fixed’ on acceptance of this deliverable. However, minor changes and re-prioritisation affecting the content of VBE v2.0 and VBE v3.0 will be possible as a result of feedback from participants during the course of the project and managed through normal change control procedures.

3.1 VBE v1.0 (D2.5.1 – M9)

- installation of baseline Bank service
- population with agreed subset of vocabularies
- configuration to allow import of vocabularies in agreed formats (see section 2.2)
- configuration to allow export of vocabularies in agreed formats (see section 2.3)
- configuration of permissions/ access privileges to allow development of required workflows for vocabulary lifecycle management (see section 2.7)

3.2 VBE v2.0 (D2.6.1 – M18)

- additional export formats (see section 2.3)
- automated translation function (see section 2.4)
- 'simple' import format tool / wizard (see section 2.4)
- enhanced workflow management (see section 2.7)
- current awareness (see section 1.4)
- conformance reports (see section 2.3)
- id allocation initial deployment (simple numeric)

3.3 VBE v3.0 (D2.7.1 – M27)

- further enhancements to workflow (see section 2.7)
- additional enhancements / support for import/ export formats based on feedback from v1 and v2
- bespoke integration into resource discovery infrastructure
- possible standards based upload depending upon evaluation (see section 1.3)
- id allocation plug-ins as required

4 Supporting activities

Documentation and training materials, in the format specified in D2.4, will be produced to accompany each version of the Bank. These materials will be added to the infrastructure (wiki, Moodle course, etc) developed as part of D2.4. In addition, specific training events and/or workshops focussed on use of the Bank may be held in conjunction with other planned ASPECT events according to demand.

The documentation and training materials are likely to include the following:

4.1 Why use controlled vocabularies?

This will be presented as a document and as a narrated video. It will provide a background to how vocabularies can be used to support tagging, searching, browsing and filtering.
It will include a User Case from the ASPECT project.

4.2 How to use ASPECT Vocabulary Bank

This will be presented as a document, an interactive tutorial, online help and as narrated video demonstrations of key functions. It will include help for general users that will include:

- browsing the ASPECT Vocabulary Bank
- searching for terms
- downloading vocabularies
- how to change the preferred language

4.3 Administering ASPECT Vocabulary Bank

This will be presented as a document and an interactive tutorial. It will include help for administrators and editors that will include:

- how to load a vocabulary in a chosen format
- guidance on producing vocabularies in supported formats
- how to change a vocabulary to a new version
- how to revert to a previous version of a vocabulary
- how to delete a vocabulary
- how to administer users

4.4 How to integrate ASPECT Vocabulary Bank

This will be presented as a document and an interactive tutorial. It will include help for developers that wish to integrate ASPECT Vocabulary Bank into their systems, such as tagging tools or content management systems, so facilitating 'semantic interoperability'. The content will include:

- typical architectures
- issuing queries
- using version and history information
- getting deltas
- guidance on how to connect to a CMS
- examples of integrating vocabularies into a tagging tool

It is acknowledged that certain items of Documentation and Training Materials may require translation by other project partners.

5 Annex 1: Publishing Services and Tools

5.1 DCSF Vocabulary Bank (http://bank.vocman.com)

Description
UK, Free to use within Education sector, pre-cursor to Lexaurus bank, no longer actively developed
Standards Compliance
Zthes v1.0

5.2 Lexaurus Bank (http://www.vocman.com)

Description
UK, Commercial
Standards Compliance
Zthes v1.0, SKOS, SRU

5.3 LexBig (http://informatics.mayo.edu/LexGrid/index.php?page=lexbig)

Description
US, Open Source, used in the Health sector
Standards Compliance
OWL, RRF

5.4 HILT (http://hilt.cdlr.strath.ac.uk/)

Description
UK, JISC project and possible future service (aimed at UK HE community)
Standards Compliance
SKOS (possibly Zthes in future), SRW

5.5 Ocean OTS (http://www.oceaninformatics.biz)

Description
Australian, commercial, aimed at health sector
Standards Compliance
XML format(s) not known

5.6 OCLC Terminology Service (http://www.oclc.org/research/projects/termservices/)

Description
USA, Research, possible future service, terms of future use unclear
Standards Compliance
Zthes v1.0, SKOS, MARC, SRU

5.7 Semaphore Ontology Service ((http://www.smartlogic.com)

Description
UK, Commercial
Standards Compliance
ISO 2788, Import / Export formats: MultiTES; csv; XML

5.8 WebChoir (http://www.webchoir.com)

Description
US, formerly known as TCS-8 and TermChoir, commercial
Standards Compliance
Z39.19, Import / Export formats: ASCII, XML, MARC

6 Annex 2: Editing Tools

6.1 a.k.a.® Classification Software (http://www.a-k-a.com.au/aka_classification)

Description
Australian, Commercial
Standards Compliance
ISO 2788, ISO 15489

6.2 APELON TDE (http://www.apelon.com/solutions.htm)

Description
US, Commercial
Standards Compliance
Format(s) not stated

6.3 Data Harmony (http://www.dataharmony.com/)

Description
US, Commercial
Standards Compliance
ISO 2788, ISO 5964, Z39.19

6.4 Knoodl (http://knoodl.com)

Description
US, Web based SOAS (hosted on Amazon EC2), Free to use
Standards Compliance
OWL, RDF

6.5 Lexaurus Editor (http://www.vocman.com)

Description
UK, Commercial
Standards Compliance
Zthes, SKOS, other xml formats

6.6 Mondeca Intelligent Topic Manager T3 (http://www.mondeca.com)

Description
France, Commercial
Standards Compliance
URI, RDF, OWL, SKOS, BS8723, ISO 25964.

6.7 MultiTes (http://www.multites.com)

Description
US, Commercial
Standards Compliance
XML format(s) not stated

6.8 Protégé (http://protege.stanford.edu/)

Description
US, Open Source
Standards Compliance
OWL, SKOS

6.9 Semaphore Ontology Manager (http://www.smartlogic.com)

Description
UK, Commercial
Standards Compliance
ISO 2788, Import / Export: MultiTES; csv; XML

6.10 Soutron Global Thesaurus (http://www.soutron.com)

Description
UK, Commercial
Standards Compliance
XML format not stated

6.11 Stratify (http://www.stratify.com)

Description
US, Commercial
Standards Compliance
XML format not stated

6.12 STRIDE (http://www.questans.co.uk/p100l2.html)

Description
UK, Commercial
Standards Compliance
ISO 2788, Z39.19, ISO 5964

6.13 Synaptica (http://www.synaptica.com/djcs/synaptica)

Description
US (Dow Jones), Commercial
Standards Compliance
Zthes, Z39.19; ISO 2788 / 5964, Export formats include Microsoft Word and Excel, XML, csv, RDF, OWL

6.14 Term Tree 2000 / One2one (http://termtree.com.au/)

Description
Australia, (formerly Hierarch) Commercial
Standards Compliance
ISO 2788, ISO 15489

6.15 Tema Tres

Description
Spain, Open Source
Standards Compliance
XTM (possibly Zthes, SKOS)

6.16 Teragram Taxonomy Manager TK240 (http://www.teragram.com)

Description
US, Commercial
Standards Compliance
Format(s) not stated

6.17 TB Thesaurus Builder (http://www.thesaurusbuilder.com

Description
The Netherlands, Commercial
Standards Compliance
ISO 2788

6.18 Triga TheMa for Oracle (http://www.moving-objects.de)

Description
Germany, Commercial
Standards Compliance
XML (format(s) not stated)

6.19 WebChoir (http://www.webchoir.com)

Description
US, formerly known as TCS-8 and TermChoir, Commercial
Standards Compliance
Z39.19, Import / Export formats: ASCII, XML, MARC

6.20 Wordmap (http://www.wordmap.com)

Description
US, Commercial
Standards Compliance

ISO 2788, .csv, XML (format(s) not stated)