![EPA logo]

Terminology Services

BEST PRACTICES IN TERMINOLOGY DEVELOPMENT AND MANAGEMENT:

A GUIDE FOR EPA EDITORS AND STEWARDS

Data Standards Branch
EPA / Office of Environmental Information

April 2014
Final

TABLE OF CONTENTS
(Use the table of contents to navigate through the print version)

Figures

1.0   Intended Audience and Purpose

This manual is intended for vocabulary Stewards and Editors. A Steward is the owner of an existing vocabulary managed in Terminology Services. Editors support Stewards by developing and/or managing vocabulary content in the Synaptica software product, the EPA's enterprise terminology management tool. In many cases, the Editor may also be the Steward. Regardless of the role, access to Synaptica requires a user name, password, and training which may be obtained by contacting the Terminology Services Coordinator in the EPA Data Standards Branch (etss_admin@epa.gov).

Guidelines in this manual are not mandatory but are intended to help vocabulary Editors and Stewards develop and maintain vocabularies of high quality and usefulness.

2.0   Background

"Say what you mean and mean what you say" is the key to successful communication. Communication is at the heart of the benefits from managing concepts – the things, processes, and activities an enterprise cares about – and the terminology used to express those concepts. Managing terms and their meanings; explaining acronyms and abbreviations; and clearly defining the terms that are used for others within EPA, for partners and collaborators, and for the public, are key to performing EPA's science, regulatory, and information mission.

In practice, managed terms, organized into managed vocabularies, can be used to tag content of various types, from data sets to documents. Managed vocabularies support metadata creation, records and content management, and search. Managed vocabularies, such as thesauri and ontologies, can work behind the scenes to support search and computer-support for decision making. This improves the user's ability to find, access, and use EPA's content.

In recognition of the benefits of managing vocabulary, the Data Standards Branch/Office of Environmental Information provides EPA Terminology Services to the EPA enterprise and its partners. Terminology Services is, in part, a resource for creating, maintaining, searching, and publishing vocabularies. Terminology Services consists of five components – content, governance, tools, machine-to-machine services, and people services. The content component is an online repository of terms of importance to EPA, its stakeholders, and partners. The commercial product, Synaptica, is available to create, store, maintain, and distribute vocabularies. Collaborative governance approaches have been developed to promote search and re-use and to bring communities of interest together to create terminology, definitions, and mappings, and to document differences. Machine-to-machine services such as APIs (Application Programming Interfaces) and web services allow other systems to access vocabularies stored in the system. The Coordinator and the System Support Staff (etss_admin@epa.gov) provide training as needed; answer technical questions; and provide guidance regarding governance structure, workflow, business rules, monitoring procedures, and general consulting

support to help ensure the development of quality vocabularies. Additional information about the Coordinator's role is included in the process descriptions below.

3.0   Introduction

This manual is organized into several sections around the recommended process for planning, implementing, and maintaining a vocabulary including how to create a new vocabulary and its content incorporating best practices related to the selection of terms and relationships, the issue of governance, and collaborative vocabulary development.

Throughout this document, references are made to other Terminology Services manuals as appropriate. These manuals are available from the Terminology Services web site (http;//www.epa.gov/ts). Please note that login is required to access some of these manuals.

3.1   What is a vocabulary?

A vocabulary is a set of terms or symbols that have been selected and organized for a specific purpose. There are several basic vocabulary types.

- Pick lists, keyword lists, acronym/abbreviation lists, or term lists are simple lists of terms, preferably with definitions that do not have a hierarchical structure. They are often used as drop-down menus in user interfaces to online systems.

- Glossaries are lists of terms with definitions that are related to a specific, subject, discipline, or information product such as a report. Generally, glossaries are arranged alphabetically. They can include alternate names for terms such as acronyms.

- Taxonomies and classification schemes are used to categorize content; therefore, they must point both the person applying the terms and the users (which may be systems or people) to broader or narrower terms depending on the rules to be applied and to the results that are needed. There is an emphasis on establishing preferred terminology. A taxonomy can also be used to organize content in a Web environment in an enterprise's information architecture to support browsing and navigation.

- Thesauri are also used to categorize content, but the structure is enhanced by adding additional relationships such as related terms, preferred terms, and synonyms to the broader and narrower relationships included in taxonomies and classification schemes.

- Authority files are developed to control the form of a narrow set of terms. For example, an author name authority file establishes the preferred form of an author's name. An organization authority file establishes the

preferred form of an organization or agency name. The authority file may also provide pointers from abbreviations, acronyms, or variant names to the preferred form. Generally, authority files are used to control types of terms, like proper names, that are not routinely included in taxonomies or thesauri.

- Ontologies are models of the relationships between objects in a particular domain. They are similar to other vocabulary types in that they select entities that are important to the community. However, they are much richer in terms of relationships and may include rules or constraints that aid computers in using this information to support automated processes.

Vocabularies may be formal or informal. The formality of the vocabulary has to do with the official nature of its use. For example, if a glossary is to be made available to the public from the EPA web site or through the EPA Terminology Services within the System of Registries, the glossary is a formal vocabulary requiring a more rigorous review and governance structure. Controlled vocabularies such as keyword lists or taxonomies used to support key business functions such as tagging web pages or document content are also formal, as are vocabularies tied to laws and regulations. An informal vocabulary is one developed for personal, internal small group, or project use. Examples of informal vocabularies would be those developed by individuals using the MyGlossaries function of the Terminology Services web site. Informal vocabularies are not official, can be maintained as the user desires, and do not require a governance structure. However, even an informal vocabulary must have a named point of contact (a position that doesn't have the formal governance responsibility of a Steward) to be included in Terminology Services.

3.2   What is a well-formed or high quality vocabulary?

A quality vocabulary is, first and foremost, a vocabulary that meets the needs of its audience, the business purpose, and the system it is intended to support. However, there are also some principles for well-formed vocabularies that span EPA disciplines and user groups. Well-formed vocabularies have an explicit and recognizable scope, a consistent approach to the form of the terms, and, as much as possible, clear distinctions between the meanings of terms. The criteria for determining quality depend on the scope of the terminology project, which is addressed in Section 4.0. Section 5.0 outlines the steps for a terminology project that will help to ensure a well-formed and high quality result.

4.0   Scope of Terminology Projects

Vocabulary projects come in many shapes and sizes. They vary from the selection of a few controlled terms for a pick list on a user interface to larger enterprise-wide authority files that span multiple systems. They include glossaries that reflect specific regions or EPA offices and enterprise-wide systems. They range from development of a personal glossary to a taxonomy used to organize content in the EPA content management system. Regardless of the level, when considering a new vocabulary, it is important to develop a

plan for the project. The decisions to be made around such a project vary depending on the formality of the project, the type of vocabulary, and whether the project will create a new vocabulary or update an existing one. These issues are briefly discussed below, and a checklist is provided in Section 5.0.

Because of the official nature, public release and breadth of use of formal vocabularies, the process for the planning, development, and maintenance of them is more rigorous. Development will likely include a charter; a formal project plan; resource allocation; a review or advisory committee to aid in the selection of preferred terms and the forming of definitions; official review, release and versioning procedures; and ongoing governance.

It is important to note that an informal vocabulary may migrate into a formal vocabulary as it is used and accepted in a more official capacity. In this case, the Steward should re-evaluate the process for developing and maintaining the vocabulary and, in particular, the governance and review processes, to ensure that they meet the needs of a formal vocabulary.

The type of vocabulary structure may also affect the way the project is approached. Terminology Services has established several vocabulary types as described in Section 3.1 above. Project management, particularly the attention to workflow and people resources, increases as the type of vocabulary becomes more complex. For example, scheduling Subject Matter Experts, coordinating their input, and building consensus for the development of relationships in a thesaurus or definitions in a glossary will require more formal project management than a simple keyword list. Ontologies are the most demanding because they require more input and consensus in order to transfer human knowledge of the relationships and rules within a domain into a standard format that can be used by a computer.

5.0   Vocabulary Project Steps

This section provides best practices for conducting vocabulary projects. The steps are outlined in Figure 1 below. The remainder of this section describes the various steps that should be considered when planning such a project. The numbers in parentheses in the Figure 1 key the steps to the text below. Please note that these are guidelines and the needs of a particular project may vary. In addition, many of these steps may be accomplished simultaneously or in an overlapping manner.

Informal vocabulary development can begin with a review of the processes below. If the goal is to start out with an informal vocabulary and progress to something that is more formal, this section should still be reviewed to keep these practices in mind as the development process continues.

Govern Vocabulary Using Established Structure and Principles

Plan Project (5.1)
- Perform Business Needs Analysis (5.1.1)
- Identify Stakeholders (5.1.2)
- Identify Vocabulary Type (5.1.3)
- Analyze Alternatives (5.1.4)
- Determine Project Scope (5.1.5)
- Identify People Resources & Roles (5.1.6)
- Arrange Technology Support (5.1.7)
- Develop a Charter (5.1.8)
- Determine Level of Effort & Secure Funding (5.1.9)

Set up Vocabulary (5.2)
- Supply Metadata (5.2.1)
- Establish Vocabulary Structure (5.2.2)
- Establish Dev. Environment & Workflow (5.2.3)
- Establish Business Rules & Guidance (5.2.4)

Create Vocabulary Content (5.3)
- Gather Potential Terms & Concepts (5.3.1)
- Select Terms (5.3.2)
- Apply Style Rules (5.3.3)
- Add Term Metadata (5.3.4)
- Create Definitions (5.3.5)
- Create Relationships (5.3.6)

Test Vocabulary Content (5.4)
- User Input to Term Selection & Organization (5.4.1)
- Test against Content/Use Scenarios (5.4.2)
- Test Results with Users (5.4.3)

Monitor & Control (5.5)

Publish (5.6)

Figure 1. Terminology Development Project Steps

Close out Dev. Project (5.7)

Update & Maintain Existing Vocab (5.8)
- Establish Maintenance Workflow (5.8.1)
- Establish Ongoing Governance (5.8.2)
- Use Term. System Maintenance Tools (5.8.3)

While there may be variations on the process depending on the scope, type, and complexity of the vocabulary, the majority of these activities will be needed to ensure a well-developed, interoperable, and quality vocabulary product.

5.1   Plan Project

As with any well executed project, a successful terminology development project starts with good planning. While many of the activities will mirror those for general project management, there are some differences and nuances for a terminology project which are outlined below.

5.1.1   Perform Business Need Analysis

Before embarking on a vocabulary project, it is important to determine why a vocabulary is being considered. Why do you need a vocabulary? Is it to control or provide validation for the contents of a field, (i.e., a limited value domain)? Is it to organize material? If the material self organizes or can be organized naturally based on a very visible characteristic, a controlled vocabulary may not be needed. Is the purpose to enhance the capabilities of a search engine or spell checker; for example, to expand searches using synonyms?

5.1.2   Identify Stakeholders

The audience and purpose will impact the subject, geographic, or activity scope of the terms that are collected and the approval process. It will also determine the degree to which synonyms and related terms should be included and whether the focus of the vocabulary will be on jargon, popular terms, scientific terms, etc. In addition, the project may have specific sponsors or funders who should be considered. Sponsors and funders will impact the resources available for the project. In addition, they may present particular needs for reporting or dictate certain governance structures that should be considered.

5.1.3   Identify the Vocabulary Type Needed

Although the lines between these traditional structures are increasingly breaking down, such that glossaries include hierarchies and thesauri have definitions, it is important to initially determine what type of vocabulary is really needed to satisfy the business needs identified above. A list of vocabulary types with definitions can be found in Section 3.1.

5.1.4   Analyze Alternatives

It is also important to look at existing vocabularies, both inside and outside EPA, to determine if the work has already been accomplished, in whole or in part. The content in the EPA terminology system can be searched to identify vocabularies

that might meet the needs of the project. Actively managed vocabularies are available from Terminology Services. For a list of all active and archived vocabularies, contact the EPA Terminology Services Coordinator (etss_admin@epa.gov). See Section 5.3.1 for more information on alternative resources.

5.1.5   Determine Project Scope

The scope of the project should consider the resources available for the original development as well as the resources that will be needed for ongoing maintenance. A very well-formed vocabulary that is not maintained can quickly become outdated and useless. It may be preferable to undertake a more modest project with a narrower scope and then continue to grow the vocabulary after implementation.

5.1.6   Identify People Resources and Their Roles

While the development of vocabularies can be supported by automated tools, ultimately, decisions need to be made by human beings. Within the context of a given vocabulary project, one or more of the following roles and responsibilities may be needed. These roles and responsibilities should be outlined in the charter or project plan. It is important that resources be formally allocated to these positions in order to ensure that the project is accomplished.

Steward

The Steward is the equivalent of the project manager. In a formal vocabulary, the Steward is responsible for obtaining the financial and people resources needed, setting schedules, establishing metrics and monitoring, determining how terms will be collected, and working with the other team members to establish the workflow, business rules, and guidance. Within the general guidance provided in this manual and the Editor's Training Manual, the Steward directs the project toward its stated objectives. In the case of an informal vocabulary project, the Steward establishes the scope of the vocabulary project. Once the vocabulary is completed, the vocabulary will be maintained by the Steward or turned over to someone else who has been assigned the maintenance role. Glossaries created with the MyGlossaries functionality are maintained by a Point of Contact who has responsibility for the scope and selection of terms for his or her own glossary. However, a Point of Contact does not have the formal governance responsibilities of a Steward.

Editor

The Steward may also be supported by one or more Editors of the vocabulary, responsible for actually making the changes to the vocabulary using the terminology system. The Steward may also serve as Editor. The Editorial process is documented in the Editors' Training Manual.

Lexicographers and terminologists may be employed to assist Editors or to serve as Editors. These professionals understand the rules for constructing different kinds of terminology resources. They also generally have an understanding of the collaborative process that is involved in developing these types of resources. The Data Standards Branch and the Terminology Coordinator can be consulted for a list of support contractors (etss_admin@epa.gov).

Subject Matter Experts

In formal vocabularies, the Steward may be supported by one or more subject matter experts. This is especially important in large-scale vocabularies that span multiple topics, projects, programs, offices, partners, or regions. Subject matter experts bring expertise on the topic or potential use cases for the vocabulary to the team. These experts are generally not experts in vocabulary development; rather they are experts in the subject being addressed by the vocabulary.

Governing Board

A Governing Board may be developed to oversee the vocabulary project. The Board is responsible for representing the broader sponsoring organization, providing advice, particularly with regard to the use of the vocabulary, and giving an "official" stamp of approval. While they may include subject matter experts, the Board's responsibility is to take a broader view of the vocabulary.

Users

There may also be a broader Community of Interest (COI) focused on the project, discipline, or application for which the vocabulary is being or has been developed. The end users must be considered throughout the vocabulary development process since the goal is to make the vocabulary reflect the world view of the users as much as possible.

5.1.7   Arrange for Technology Support

Large formal vocabularies, in particular, will benefit from collaborative development and review of the vocabulary. This approach will share the work, ensure appropriate subject matter expertise, increase the visibility of the effort, and promote the acceptance of the vocabulary.

Collaboration can be supported by a number of technology approaches which are described below. A final decision regarding the best approach for creating the vocabulary should be made in consultation with the EPA Terminology Services Coordinator (etss_admin@epa.gov).

EPA Terminology System

The main tool within EPA's Terminology Services is Synaptica from Synaptica LLC, a commercial-off-the-shelf software product for creating, storing, maintaining, and distributing vocabularies. As a vocabulary management product, Synaptica supports workflow, including input from multiple Editors, and term statuses that support a process for review and governance. The relevant functions provided by Synaptica are outlined in

Section 5.3 below. More information is available from the Editors' Training Manual which is available from Terminology Services.

EPA Portal Technology

The System of Registries (SOR) collaboration pages or other collaborative workspaces such as the Office of Research and Development's Environmental Science Connector can be used to support collaborative work on a vocabulary by creating a project area for the vocabulary. Important documents, such as the charter, lists of committee members, guidelines, and versions of the vocabulary, can be posted to the area for review by a COI within the portal. A COI is created by registering members to the portal and the appropriate project area or inviting participants to register themselves.

The terminology documents are checked out, modified and commented upon, and then checked back in through the portal. These changes and comments are then visible to other participants. If managed properly, the portal can provide a trail of the discussions that have taken place in reaching consensus about a particular term. This type of approach is particularly valuable when working on definitions for terms.

Wikis

Similar to the use of the portal, it is possible to use EPA wiki technology. The Terminology Services Semantic MediaWiki allows for collaborative efforts to create, edit, and review terminology by EPA staff and partners. This approach also uses the concept of a COI to organize a collaborative project. Within the wiki, vocabulary work group members are granted access to their vocabulary's work space. Relevant documents can be posted to the wiki. In addition, a wiki allows others to edit and comment directly on the terminology. Like the portal approach, the wiki tracks what was edited and by whom. To request workspace via the wiki, please contact the Terminology Services Coordinator (etss_admin@epa.gov). You will need an EPA portal account in order to log-in.

5.1.8   Develop a Charter

Similar to any well managed project, developing a vocabulary benefits from a formal charter or statement of purpose, the setting of roles and responsibilities within the group, business rules and guidance, and monitoring and metrics to gauge progress. A charter can be helpful in gaining management support.

The charter states the scope and purpose of the project. It also formalizes the roles and responsibilities. It may also reference pre-existing business rules and guidance and set processes for monitoring and gauging progress against the goal. For large vocabulary projects, this may be developed further into a formal project plan and schedule. A sample charter is included in Appendix A.

5.1.9 *Determine Level of Effort and Secure Funding*

The cost of developing a vocabulary can range from several thousand dollars for a small glossary to several hundred thousand dollars for a large Research and Development (R&D) taxonomy. For a large vocabulary, it is important to consider funding for maintaining the vocabulary to ensure it remains current.

Even small projects may require considerable effort on the part of an individual, project team or small group. The level of effort required on the part of the Steward and the Editor is especially critical. If the vocabulary is to be used and accepted beyond a small group of people, it is important to ensure adequate review in a collaborative environment. Gaining funding support or in-kind effort to involve subject matter experts and Governing Board members therefore, becomes critical to the eventual success of the project. Finally, the level of effort does not end with the publication or dissemination of version 1.0. Ongoing support and funding is required to ensure that the vocabulary remains a viable asset.

For more help in estimating the appropriate level of effort and funding requirements, contact the Terminology Services Coordinator in the Data Standards Branch (etss_admin@epa.gov).

5.2   Set Up the Vocabulary in the Terminology System (for formal vocabularies)

Once the planning for the vocabulary has been completed and resources secured, the Steward should contact the Terminology Services Coordinator (etss_admin@epa.gov) to set up the vocabulary in the System. This involves supplying descriptive and other metadata (information about the vocabulary), establishing the vocabulary structure, identifying roles and permissions for the people who will be working on the vocabulary, and establishing business rules and guidance. The Coordinator will add the vocabulary name and metadata to the *Vocabularies* file.

5.2.1 *Supply Metadata to the Terminology System*

When setting up a vocabulary, metadata for the vocabulary must be provided to the Terminology Services Coordinator (etss_admin@epa.gov). See Appendix E for an example of the vocabulary metadata that will be requested. The metadata elements highlighted with * are those that are required for all vocabularies. Those indicated with a + are required only for Active vocabularies to ensure consistency and compliance with the terminology system's read-only interface. However, all the metadata listed in the appendix may not be available for all vocabularies. Informal vocabularies may have limited metadata based on the use of the vocabulary and whether it was developed from other sources.

5.2.2 *Establish Vocabulary Structure*

The Coordinator will also need to know any additional information that will accompany each term of the vocabulary, such as definitions, scope notes, etc. These additional elements, known as extended attributes, must be created in the vocabulary's structure by the Terminology Services Coordinator or the System Support Staff prior to entry or import of terms into the vocabulary. All imports are done by the Terminology Services Coordinator with support from System Support Staff.

5.2.3   Establish a Development Environment and Workflow

Most formal vocabularies will require collaboration among people or across projects, programs, offices, or partners in order to ensure that the vocabulary terms and definitions are well understood by potential users, represent the expertise of the community, and, therefore, can be recommended for official approval and release. Working with the Terminology Services Coordinator, the Steward of a formal vocabulary should set up a workflow process within the Terminology Services tool, Synaptica. In Synaptica, workflow is established through roles and permissions in combination with status flags. These components are described below.

Roles and Permissions

A user's Synaptica profile dictates which vocabularies the individual can access and what the individual can do with each vocabulary. An individual can have Editor's access (creation and modification privileges) to one vocabulary and read-only access to another. Levels of access and permission ensure that the changes to the vocabularies in Synaptica are performed by the appropriate people and as part of an agreed upon workflow. These profiles are created by the Terminology Services Coordinator.

The most common role is that of Editor. An Editor can add, deactivate and delete terms; add and modify relationships; and add, delete and modify other content such as definitions. In addition, a Read-Only reviewer can review the vocabulary as it is being developed without making changes. Note that the public Read-Only interface cannot be used to view records that have not been released or published. The terminology system supports a variety of additional roles and permissions. See Appendix C for a complete list.

The Terminology Services Coordinator will need to know who should have access to the vocabulary and with which permissions. If there are new users who do not currently have access to the EPA terminology system, the Coordinator will request their contact information and a statement about the reason they will be accessing the system in order to set up their user accounts. More information about obtaining access to the terminology system is included in Appendix D.

Status Flags

Status Flags identify the approval status of the term as candidate, pending, machine loaded, approved, or unapproved. The status flags can be used to govern the terms added to the system or terms that are under review in the system. Candidate is the default

approval status for terms that are added. Editors may change the status term-by-term or globally across the vocabulary.

Model Workflows

The permission levels, in combination with the status flags, determine the workflow. For example, there may be Editors who can add new terms but the status of the term from "candidate" to "approved" may only be changed by a more experienced Editor. The workflow can be set so that a Governing Board or subject matter expert must review the candidate terms before they are approved in the vocabulary. Alternatively, the Governing Board or subject matter expert may be given read-only access to look at the input by the Editor, without being able to make any changes to the terms or their status. Examples of workflow processes used at EPA to develop current vocabularies in the System are included in Appendix E.

The Steward should contact the Terminology Services Coordinator (etss_admin@epa.gov) at the beginning of the project to discuss the desired workflow. This will allow the Coordinator to determine the permission levels and status flags. Once the team has been assembled, the names of the team members and their roles and responsibilities should be provided to the Coordinator so that access can be granted and individual user permissions established based on the user's need to view, add, modify, and/or review and approve the terms within the vocabulary.

5.2.4   Establish Business Rules and Guidance

While many aspects of vocabulary development can be controlled by the permissions and workflow described above, there are other more manual aspects that will need to be developed and included in a guidance document. The guidance document should cover the creation, review, and maintenance processes. It should also document the rules for selecting sources for terms and definitions, criteria for selecting individual terms, and the style rules and input conventions. The more formal the vocabulary, and the more collaborative and distributed the environment, the more important these rules and guidance become to ensure a successful outcome to the project. (See Section 5.3 below for selection, format, and style elements that should be covered in the guidance.)

Guidance is not as critical for informal vocabularies. However, it can be helpful if the vocabulary will be developed over time or shared with others. For example, it will be helpful to document the basis for the selection of terms, decisions about sources of terms and definitions, and general style rules. Having these documented will help ensure consistency and save time.

5.3   Create Vocabulary Content

While there are many ways to create a vocabulary, the guidance below suggests general good practices in this area. The exact order and time spent in any of these steps will depend on the type of vocabulary, the scope of the project, the number of concepts to be handled, and other decisions made during the planning stage.

Terms can be entered into the vocabulary individually through the Editor's interface. Instructions for adding terms, modifying or deleting terms, setting up relationships, etc., are available from the Editors' Training Manual.

Terms may also be entered via a batch import of a tagged CSV (comma-separated variable) or XML (Extensible Markup Language) file. This is the preferred method if an existing vocabulary in electronic form is serving as the basis for the vocabulary. It can also be used if the new vocabulary has been developed in Excel. The specifications and instructions for importing these files are available in the Editors' Training Manual. Imports must be performed by the Terminology Services Coordinator or the System Support Staff.

5.3.1 Gather Potential Terms and Concepts

When beginning a new vocabulary, the first step is to collect terms that reflect the concepts of interest based on the scope, purpose, and audience identified in the planning phase. Terms can be gathered from a variety of sources including, but not limited to:

- Vocabularies already in Terminology Services, especially those that are still actively managed, published by a related organization, or are otherwise authorized. For direct access, contact the Terminology Services Coordinator (etss_admin@epa.gov), or use the public read-only access via the System of Registries (http://www.epa.gov/ts).
- Relevant published materials including journal articles, books, white papers, and technical reports. Review articles and text books are also beneficial since they may present many relevant terms in a single resource and are often reflective of preferred terminology determined by an established community or discipline.
- Web pages, strategic plans, research plans, or other materials for the project or organizational units represented.
- Other authoritative terminology resources such as those available through the CENDI Science Terminology Resources at:
  http://www.cendi.gov/projects/proj_terminology.html

In all cases, it is important to check with subject matter experts or stakeholders to determine the relevance of these materials, as the material, and, therefore, the terminology may have a particular bias or be outdated. The best terms are those that appear in a number of sources.

5.3.2 Select Terms

Once terms have been collected, and as new concepts are added, it is important to select the appropriate terms to include in the vocabulary. While the selection of the terms will depend, in part, on the type of vocabulary, there are several general guidelines.

- Base the selection on usage for the served community. The selection of one term over another based on usage, particularly in published sources, is often called "warrant." It may be helpful to document the results of searches for each term as the number of occurrences of the term over the number of occurrences of all the terms that could be used for the concept (e.g., 5/25).
- Select terms that are important to your audience (known as user or organization warrant)
- Select terms that occur in the content to be tagged or indexed using the vocabulary (known as literary warrant)
- Select terms that are unique, wherever possible avoiding the use of the same term for multiple meanings, unless you have no other choice. In this case, scope notes should be included and definitions written to distinguish the terms. The non-unique term should be modified in parentheses by identifying the context, group or domain to make the meaning unambiguous; e.g., cranes (lifting equipment) and cranes (birds). However, multi-word terms are generally preferred over the use of modifiers.
- Variants of selected terms can be added as non-preferred terms.

Instructions for adding terms in the EPA terminology system can be found in the Editors' Training Manual.

Acronyms, Abbreviations, and Codes

In general, acronyms, abbreviations, and coded values should not be included as the preferred terms in vocabularies. However, it is also the case that acronyms, abbreviations, and codes are extremely common in the EPA environment.

Caution is needed since many acronyms, abbreviations, and codes can stand for multiple concepts. It is important to document these terms and to link them to the expanded form through special relationship types. These relationship types are used by the read-only interface to display abbreviations and acronyms for terms and can be used to create lists of abbreviations and acronyms.

The EPA terminology system has created two special relationship types to distinguish abbreviations and acronyms from other terms. Acronyms and abbreviations should be linked to the full-length terms with the Abbreviation For and Abbreviated By (AF-AB) relationship. For example, TMDL is linked to Total Maximum Daily Load with an AF-AB relationship. The relationship from Total Maximum Daily Load to the acronym TMDL appears as the reciprocal relationship, AB-AF. This special relationship type allows special interfaces or reports to be developed to produce acronym and abbreviation lists.

Similarly, the Code For and Coded By (CF-CB) relationship has been developed to link terms with their coded values. This is used in the Web Taxonomy for the Business Reference Model terms and codes and for substances and their Chemical Abstracts Service (CAS) Registry Numbers. For example, there is a CB relationship between

"polychlorinated biphenyls" and its CAS number, 1336-36-3. The entry, CAS 1336-36-3, has a CF relationship to "polychlorinated biphenyls."

For information on how to create these relationships, see the Editors' Training Manual.

Phrases versus Separate Terms

Whether a phrase representing joined concepts is preferred over two concepts represented by individual terms depends on the common usage and the way in which the vocabulary will be used. For example, in one vocabulary, children and health could be treated as separate concepts. In others, such as the EPA Web Taxonomy, the joined concept of children's health is so prevalent that it makes more sense to combine the terms.

The decisions about using pre-coordination versus post-coordination require a balance. The system or systems that will ultimately provide access to the content that is tagged with the vocabulary must be considered. Most taxonomies that are used to provide browse capabilities for Web sites favor the combination of multiple concepts in commonly used phrases, since there is no easy way to join the contents under the one concept with those under the other concept. Vocabularies that are being used with search engines may favor two separate concepts, understanding that the decisions should still be weighed based on if the users will create phrases versus using terms individually.

In general, terms should be noun phrases composed of an adjective (or in some cases an adverb) followed by the noun. For example children's health is preferred over health of children. Again, the term depends on the common usage within the targeted community. Some nouns modified by prepositional phrases are in such common usage that they represent the concept itself; e.g., "cream of tartar."

Verbs and Other Process Terms

In general, terminologies are composed of nouns and noun phrases; e.g., fermentation versus fermenting or catalysis versus catalyze. However, there are concepts that deal with processes that are difficult to express without the use of verbs. The decision whether to use a noun form or a verb form will depend on the user group and the purpose of the vocabulary. For example, in a taxonomy used for browsing, err on the side of the user warrant to ensure that users resonate with the terms that are being used. If the vocabulary is being used behind the scenes and the users interact with the terminology via a search engine, the noun phrase may be more appropriate.

5.3.3   Apply Style Rules

In addition to selecting the term for the concept, there are conventions about the style of such a term. Style or Editorial rules should be documented for each vocabulary in order to ensure consistency as the vocabulary develops over time. Adhering to the general rules will support the interoperability of the contents in the terminology system. Variants, such as acronyms, abbreviations, variant spelling, and terms from other languages, can be included as non-preferred terms (tagged in the EPA terminology system as non-preferred terms). See Section 5.3.5, Non-Preferred Terms below.

The following are general style rules that are applicable across many vocabulary types.

Singular versus Plural Terms

The selection of singular terms versus plural terms depends on the type of vocabulary. Glossaries and thesauri generally include the singular forms unless the term is a collective noun or is in such common usage that the user will expect to find the plural version. For example, it is more correct to use the term biphenyls instead of biphenyl. In the case of taxonomies, the more common rule is to use the plural form except for the name of specific programs. For example, Health Effects would be used instead of Health Effect, but Aging Initiative would be used instead of Aging Initiatives while Air is the same in singular and plural. Again, the final decision depends on common usage and the EPA context.

Special Characters

Special characters, including accent marks and special symbols, should be retained as long as they reflect the common usage. The EPA terminology system is able to handle UTF-8 character encoding so most scripts can be included. However, if the user is unlikely to be able to reproduce the characters from a standard keyboard, a written-out form may be preferred with the original version included as an entry term.

Terms composed of multiple terms in a series should use commas to delimit the terms; e.g., Policy, Guidelines, and Standards and not Policy / Guidelines / Standards. Always leave spaces between words and special characters, such as Policy / Guidelines / Standards.

In some cases, special characters such as the ampersand and parentheses may be used. The Web Taxonomy uses the ampersand (&) to make the terms more readable and shorter for display purposes.

Capitalization

Title capitalization should be used; i.e., Air Pollution Monitoring rather than AIR POLLUTION MONITORING, Air pollution monitoring, or air pollution monitoring. This is particularly important if the terms will be used as a taxonomy for web browsing.

When a hyphen is used in the term, the term following the hyphen should be capitalized; e.g., Non-Government Organization rather than Non-government Organization.

Term Order

Generally, use the direct order for terms rather than an inverted order; i.e., Food Safety, rather than Safety, Food. However, exceptions may be made in a taxonomy or term list where terms are being grouped into categories in alphabetical order; e.g., Solid Waste, Nonhazardous might be used instead of Nonhazardous Solid Waste in order to present the term in alphabetical order with other types of solid waste.

Language

American English should be used in most vocabularies. Additional languages can be added in the future as variant terms or as a parallel vocabulary. The exception is vocabularies which are intended to provide foreign language terms or which are in multiple languages.

5.3.4   Add Term Metadata

Synaptica incorporates a broad range of metadata at the term level as a best practice. When selecting a term, document its source in the Source Note field, any qualifications about its use in the Scope Notes field, and any additional comments regarding the rationale in the Editorial Note field. Note that the scope note field is displayed in the read-only interface of the terminology tool, while the Editorial note is not.

A source note and definition are recommended for each term in the terminology system. The Synaptica system automatically supplies the date the term is added/changed, the name of the user ID for the person who performed the action, and English as the default language. Figure 2 shows the term record display in Synaptica with the required and recommended field highlighted. Appendix F lists all fields available in Synaptica and their definitions. Note that additional fields can be added by the System Administrator if needed.
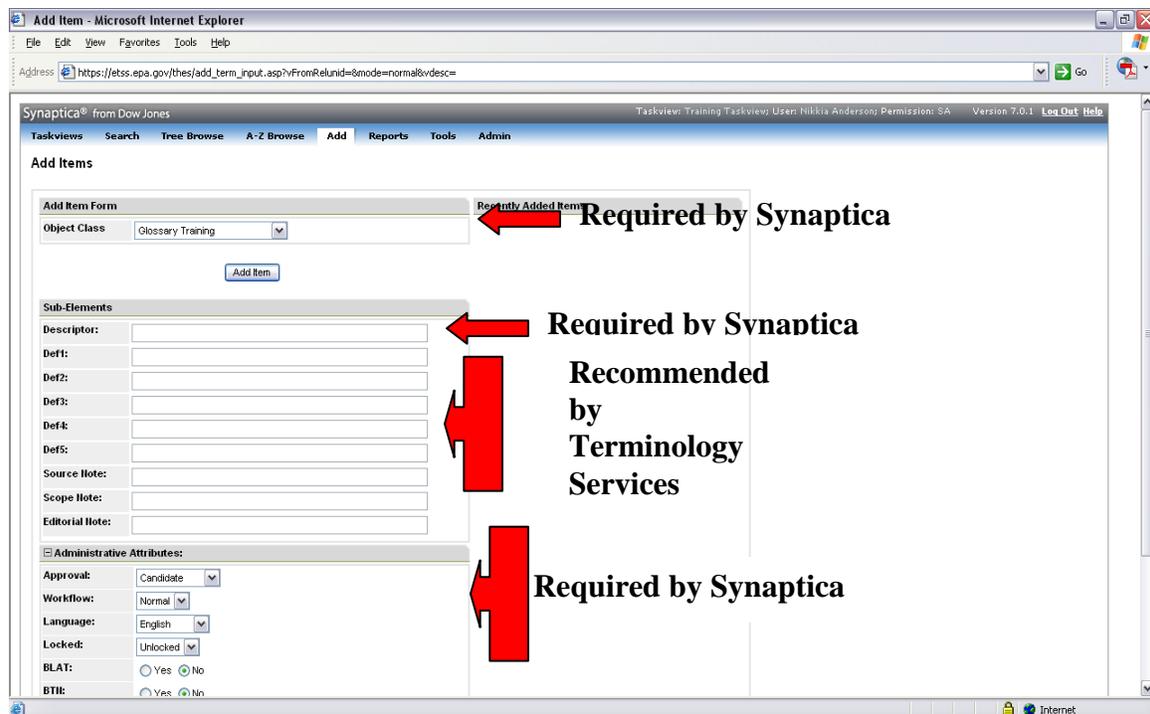

Figure 2: Terminology Services Term-Level Display with Required and Recommended Fields

5.3.5 Create Definitions

It is a best practice for vocabulary developers to include definitions for their terms. Definitions help ensure that the meaning in the vocabulary developer's mind at the time of term creation is the meaning understood by the user. Definitions are also helpful in fine-tuning the vocabulary and can remind the developer of the specific context of the term selected.

Well-formed definitions are based on the following guidelines:

- They do not include the term itself in the definition, thus avoiding a cyclical definition.

- They identify the class of objects to which this concept belongs and identify characteristics that distinguish objects represented by the term from others. The general structure of a definition is to name the group to which the term belongs followed by the distinguishing characteristics. In the following example, the term being defined is presented in bold, the group is italicized, and the distinguishing characteristic(s) is underlined. In the EPA terminology system, the term in bold would not appear in the definition field since it is already in the term field of the record.

  Corn [is] a *grain*, usually <u>white or yellow</u> in color that grows in <u>kernels on long ears</u>.

  Alternatively, in the context of a food cannery:

  Corn [is] a *vegetable* consisting of <u>individual white or yellow kernels</u>.

In addition:

- Definitions are written in active voice and first person singular, using or assuming words such as "is." However, if the term being defined is plural or a collective noun, the definition should also be worded in the plural.

  Children [are] *people* from <u>newborn through 12 years of age</u>.

  Polychlorinated Biphenyls [are] *chlorinated hydrocarbons …..*

- Definitions may be constructed of multiple sentences, but they should be succinct and easily understood by the target audience, focusing on the identification of the group and the distinguishing characteristics rather than an encyclopedic description.

Congener [is] one of the *209 different PCB compounds*. A congener may have between 1 and 10 chlorine atoms, which may be located at various positions on the PCB molecule.

- Definitions should avoid unnecessary jargon, especially if being written for the public.

- Generally, formal glossaries should not include abbreviations or acronyms since they may be made available via the Terminology Services interface in the System of Registries. However, if abbreviations or acronyms must be used, they should be germane to what is being defined. For example, the "congener" definition above is in a PCB Glossary. Therefore, using the acronym PCB would be acceptable and understood by the users. When using definitions from other sources, it may be necessary to modify or enhance definitions to avoid acronyms and abbreviations that may not be understood in the new vocabulary.

Sources of definitions include:

- Other resources in the EPA terminology system
- Terminology Resources from other agencies and organizations (being careful of issues of scope and audience), scientific dictionaries, review articles, relevant standards, and text books
- A definition agreed upon by a group of subject matter experts in the discipline

When writing definitions based on other sources, it is necessary to be mindful of copyright issues. Generally, it is better to create a definition from scratch or based on a paraphrase of another definition, unless the original source of the definition is itself in the public domain and has not made any indication of another source for its definition. Definitions from scientific dictionaries, review articles, standards, books, and journals should generally be paraphrased since they may be copyrighted.

The Definition Source should be enclosed in brackets following the definition. New definition sources should be forwarded to the Terminology Services Coordinator (etss_admin@epa.gov) for inclusion in the Source Definition File, along with abbreviations and source information. The Terminology Services Coordinator will then add the new source, abbreviation, and any relevant metadata to the Definition Source file in the EPA terminology system. The Definition Source file provides a list of sources that others can use and ensures that unique abbreviations or short names are created for each source.

Instructions for adding definitions in the EPA terminology system can be found in the Editors' Training Manual.

5.3.6 Create Relationships

Once the terms have been collected, they may be organized by creating relationships between terms. The standard relationships supported by the EPA terminology system include broader term (BT), narrower term (NT), synonym (SY), and related term (RT). A full list of relationship types is included in Appendix G. The relationship types that are used depend, in part, on the vocabulary type being created. Taxonomies have broader term (BT)/narrower term (NT) relationships. Thesauri are also organized into hierarchies using broader terms (BT) and narrower terms (NT), but they may also have related term (RT) relationships. In general, glossaries are not organized into hierarchies, but they may have related term (RT) relationships and synonym (SY) relationships for acronyms and abbreviations. The relationship types are described in more detail below.

Hierarchies

Thesauri and taxonomies are based on the idea that terms are organized in hierarchies, which are based on broader term (BT) and narrower term (NT) relationships, creating a tree structure. Thesauri have much stricter definitions of what constitute broader term (BT) and narrower term (NT) relationships than taxonomies. Fully standards-compliant thesauri require that the concept of the broader term (BT) subsumes all the concepts of the narrower terms (NTs) underneath it. This is generally reflective of two types of sub-relationships – whole-part and parent-child. In whole-part relationships, the narrower term (NT) is totally a part of the larger whole. For example, the states of the United States are part of the larger concept of United States. In parent-child relationships, the parent is completely encompassing of the children given a particular characteristic. For example, the concept motor vehicle includes trucks, cars, vans, etc.

The placement of terms in relationships depends on the scope, purpose, and needs of the vocabulary's audience. It is obvious that these terms above could be placed in other hierarchies based on other characteristics. For example, trucks, cars, and vans could also be placed under a heading called Things with Tires, which might also include bicycles. Landmasses Bordering the Great Lakes would include some of the states of the United States but also relevant Canadian provinces.

Taxonomies are looser hierarchies that are generally developed for browsing and navigating content. The tendency is to develop these taxonomies based on a number of facets, often in conjunction with the development of metadata schemes for a particular collection of material. A good example is EPA's Web Taxonomy which has individual vocabularies for Substances, Information Types, Audiences, Industries, etc. While most of these vocabularies have hierarchies, they are focused on how people would find information rather than the deep subject indexing for which most thesauri are developed.

Some key principles for creating relationships are:

- Try to put the terms in a single place in the hierarchy, duplicating terms only if it is the same concept. Duplication is acceptable if it makes sense to the users.
- Unless you are developing a browse taxonomy where the hierarchy is geared toward browsing, it is best to connect terms as related terms (RT) rather than strict hierarchies.
- There are few real synonyms, particularly when a separate relationship type is used for Acronyms/Abbreviations and for Codes. (See Section 5.3.2 above for a discussion of these types.)

The instructions for creating hierarchical relationships using the EPA terminology system can be found in the Editors' Training Manual.

Non-Preferred Terms (Use/Use For)

Adding non-preferred (NP) terms to a vocabulary can be a very powerful enhancement, particularly if the vocabulary will be used to expand searches behind the scenes. In addition, non-preferred terms, also known as entry terms, give guidance to users of the vocabulary, particularly those selecting terms for tagging information content. Non-preferred terms provide entry into the vocabulary resource itself, helping users to find the preferred term to use for a concept. Non-preferred terms may be synonyms or word variants, which are not used for tagging but help to guide the user or the system to use the preferred term. Preferred terms and non-preferred terms are joined by Use-Use For (USE-UF) relationships in the EPA terminology system.

Non-preferred terms may be added as the vocabulary is developing, reflecting decisions that are made about term preference and form. In addition, non-preferred terms will be added as the vocabulary is used. A good source for additional non-preferred terms is the search log provided by the search engine. The search logs should be reviewed periodically to identify variants or non-preferred terms that may be barriers to users finding content.

The instructions for creating Use-Use For (USE-UF) relationships and assigning non-preferred terms can be found in the Editors' Training Manual.

Related Terms

Adding relationships between related terms (RTs), terms that are associated in some way but not as formal broader term (BT)/narrower term (NT) relationships, may also enhance a vocabulary. These "see also" references can guide the user to other related concepts that might be of interest. Generally, these relationships are added over time, after the main vocabulary work has been completed. Related terms (RTs) can be valuable in linking content automatically to present related content on a results page. However, it is important to guard against over assignment of related terms (RTs); otherwise, every term becomes related to every other term.

The instructions for relating terms with RTs can be found in the Editors' Training Manual.

5.4 Test Vocabulary Content

At various points throughout the vocabulary development life cycle, it is advantageous to review and test the vocabulary content for usability. There are two general approaches to doing this kind of testing. Both types should be done at some point during the development.

5.4.1 User Input to Term Selection and Organization

Review and testing can occur from the very early stages of the vocabulary development process. At various milestones, perhaps with the completion of a particular facet, tree, or a certain number of terms, the terms can be reviewed by a COI or by end users. This helps to confirm that the terms being selected are actually those that are in common usage by the target user group. This collaborative approach is especially valuable when creating or modifying definitions and when establishing the broader term (BT)/narrower term (NT) relationships in taxonomies.

One common approach is to create index cards or post-it notes that contain the terms. Users, alone or in a group, are asked to sort the cards or post the terms on a wall or whiteboard in an order that makes sense to them. A strawman can also be presented that they can then "move around". It is helpful if users are encouraged to "talk aloud" about why they are making certain placements.

5.4.2 Test Against Content or Use Scenarios

Testing against content involves identifying examples of the content to be organized or accessed using the vocabulary. A good practice is to use the examples of the content to be organized that were collected during the term gathering phase as the first pass. An additional set should be assessed that was not used in the term gathering step but that is related to the topics covered in the selected terms. Depending on the size and the project's assessment of risk, a formal testbed of materials that includes samples of various types of materials and subjects to be covered may be developed as a blind test.

Members of the development team can attempt to use the vocabulary for its stated purpose by running use scenarios. In the case of taxonomies and thesauri, this would be for content organization and access. In the case of glossaries, it would be for definition look-up. This type of testing for authority files would include use scenarios around the control of variants found in natural language materials. In the case of pick lists, the goal would be to test them in the system or systems where they will be used.

5.4.3 Test Results with Users

Another approach to testing, which can be used in combination with the testing described above, is the testing of the vocabulary by those who will eventually be using the system. This type of testing may involve a sample of material organized using the vocabulary to see if the users can find the appropriate materials. It may involve asking users to navigate through a series of web pages using a taxonomy. In the case of a simple glossary, it may involve a review to see if terms are missing, stated in commonly understood ways, and if the definitions are understandable. The test should involve real users and real content. While this type of testing is similar to usability testing of user interfaces, it should be performed on the vocabulary well in advance of its inclusion in any system or product.

5.5   Monitor and Control: Report Progress

An important responsibility on the part of the Steward is to monitor and report on the progress of the project and ultimately on the maintenance of the resulting vocabulary. Keeping metrics on the vocabulary will help determine the level of progress made on adding terms and making necessary modification to improve the vocabulary and keep it up to date.

It is important to first determine what level of monitoring is needed. This is based, in part, on the expected degree of activity. In a major project with near-fulltime resources, it will be important to monitor the number of terms added, approved, or changed on a more frequent basis than if activity occurs less frequently. The level of monitoring will likely be less once the vocabulary is in maintenance mode. The need for monitoring and metrics should be discussed with the Terminology Services Coordinator (etss_admin@epa.gov), who can advise on how this can best be done through existing report functionality.

The terminology system has the ability to create reports to determine what activities, such as addition or deletion of terms, addition of definitions, etc., have occurred within a given vocabulary and/or taskview. In addition, reports can be created based on the Editor or reviewer associated with the change, on statuses, and date ranges. Many of these reports can be saved and run periodically.

The Terminology Services Coordinator also produces metrics for all vocabularies on a periodic basis. These reports can be made available to the Steward or Governing Board. The metrics present information about the vocabularies within a taskview. Common metrics include term and relationship counts, the number of changes to a vocabulary, the number of added terms, the number of saved reports, and the number of times each report is run in a given span of time.

5.6   "Publish" Your Vocabulary

A vocabulary can be compiled and saved on a hard drive for publishing, sharing, or record-keeping purposes. Publishing a vocabulary requires two steps—creating a report and outputting the report in the desired format. The file can be saved to your hard drive from which it can be modified, distributed, printed, etc.

Other file formats can be used to distribute the report. For example, the comma-separated variable (CSV) format can be saved and then imported into an Excel spreadsheet or an Access database. XML can be selected to share the output with another application or for use with style sheets. SKOS/RDF (Simple Knowledge Organization System/Resource Description Framework) format can be selected to interchange the vocabulary with others and to make it accessible via SKOS-oriented web services. Additional information about export can be found in the *Synaptica Import Guide* available from the EPA Terminology Services Resources web site.

Instructions on publishing vocabularies of various types and in various formats are available in the Editors' Training Manual.

In addition, the vocabulary can be "published" through the Terminology Services of the System of Registries. This is a public web-based read-only interface that provides for the search, display, and download of terms from publicly released vocabularies. The decision as to whether a particular vocabulary should be made available through this interface should be discussed with the Governing Board and the Terminology Services Coordinator (etss_admin@epa.gov).

The EPA Web Glossary Standard requires that all web glossaries be displayed via a link to that glossary from Terminology Services. The link can be shared and/or placed on individual project pages while the glossary owner or steward can still manage and maintain the glossary in the terminology system. Additional information for further maintenance and management can be found in the Glossary Management Guide.

Ultimately, the decision about "publishing" or releasing the vocabulary is determined by the publishing or sponsoring organization. Most formal vocabularies are considered EPA products and must undergo the regular review and approval process for products in that particular office, program, or region. When interfacing with these management review processes, it will be important to provide information about the guidance, review, and governance that were used in creating or maintaining the vocabulary, since these activities have been used to ensure the quality of the vocabulary's content throughout the process.

5.7   Close-Out Development Project

As with most projects, it is advantageous to assess the results at the end of the process. A final meeting with the team and the Governing Board should focus on lessons learned, what should be duplicated in similar projects in the future, and what could be improved.

Plans should also be made at this point for the ongoing maintenance activity and how the transition will be made, particularly with regard to roles and responsibilities, as the effort moves from development to maintenance. Issues such as maintenance schedule, funding, and resources, particularly for subject matter experts and other members of the Governing Board, must be considered and communicated to Management.

If collaborative sites have been developed for working on the project, their disposition should be considered. Sites that have been truly collaborative provide a good history of the decisions made regarding the vocabulary. Archiving the site, perhaps retaining it for read-only access or limiting access to a core group of people may be considered.

If a contractor has been used to support the development, routine close-out processes should be followed, including a review of contract deliverables and communication from the contractor to the COTR of all relevant documents, including guidelines and procedures that are specific to this project.

5.8   Update and Maintain an Existing Vocabulary

Planning for the maintenance of an existing vocabulary is as important as the development of a new vocabulary. While a vocabulary may be useful for many years after its development, the standard terminology in a field or long-term project may change. Eventually, the vocabulary will become obsolete, or, at best, confusing.

An existing vocabulary, even one of a formal nature, may not have a pre-existing, well-defined maintenance procedure and governance structure. Therefore, many of the considerations in the checklist should also be addressed when taking on the update to an existing vocabulary. Even informal vocabularies can benefit from documented maintenance procedures.

Maintenance procedures should include the frequency with which terms will be reviewed and added. For example, terms may be added on a monthly or yearly basis depending on the expected number of modifications. Alternatively, changes may be made when the number of changes exceeds an agreed upon threshold. Guidelines are needed for the deletion of obsolete terms. Some vocabularies may delete these terms so that they are no longer visible to users. Others, depending on the use of the vocabulary, may retain these obsolete terms as non-preferred terms with appropriate scope notes. In all cases, the warrant or usage required to trigger these changes should be documented in the maintenance procedures in order to ensure that changes are made in a consistent fashion. If these terms have been used in other documents or systems, it is important to consider the impact of maintenance on these systems.

Ongoing management of the vocabulary involves controlling changes, adding to and modifying the terms to keep them up-to-date, and otherwise reviewing the content against its stated purpose. Maintenance also involves working with the Terminology Services Coordinator to ensure that new Stewards, Editors, and subject matter experts are identified, added to the system, and granted appropriate permissions.

5.8.1 Establish a Maintenance Workflow

Maintaining the vocabulary is an important process to keep the vocabulary actively available through the terminology system. As part of the governance of the vocabulary, a maintenance workflow can be established. A maintenance workflow involves the processes or steps, as well as the roles and responsibilities for the people involved in reviewing updates to terms, extended attributes, and relationships. As an example, the change process for the Web Taxonomy is shown in Appendix E. Change processes may also be triggered by changes to published reference sources used to create the vocabulary. For example, if a glossary is based on another glossary published by a scientific society, a change in that glossary may trigger a review to determine if the change should also be made to the glossary in the terminology system.

5.8.2 Establish Ongoing Governance

While the most intense effort is required during the initial development process, it is important once the initial development is completed to determine the nature of ongoing governance. The formality and extent of the ongoing governance structure and process will depend on the formality of the vocabulary, whether it is to be updated on a routine basis or only periodically (every so many years or with the passage of new legislation or regulations), and the policies and procedures of the publishing organization.

The role of the Steward remains beyond the initial development process. The Steward manages any day-to-day changes, coordinates with the Terminology Services Coordinator, and performs ongoing monitoring and review of Synaptica metrics as needed.

For formal, public vocabularies that are updated on an ongoing basis, the team of Subject Matter Experts may be retained or new experts identified, depending on the scope of new terminology to be entered or reviewed. For example, if a glossary will include new terms specific to a topic such as nanotechnology, the subject matter experts needed for the maintenance would be drawn from experts in this specific area. However, if the maintenance will be a periodic review of terms across the vocabulary, the original team experts may be retained or a new team formed that can cover the scope of the review. Planning for replacement of subject matter experts in advance will help to ensure fresh thinking and allow members to retire from this task.

The form and structure of the Governing Board may also change over time. Depending on the number and types of changes to a vocabulary and the desires of the publishing organization, the Governing Board may review only new terms, review only new definitions, review major changes, or review all changes.

Just as with the development of a new vocabulary, it is important that there be a process for achieving the level of consensus needed. Since the selection of terms and the creation

of definitions can be a very subjective process, establishing the final authority is important, even in an ongoing maintenance environment.

5.8.3 Use EPA Terminology System Tools for Maintenance

There are several services available in the EPA terminology system that support the management of any vocabulary. These include:

- Standard alphabetical and hierarchical reports that can be produced with different filters to allow activities to focus on only pieces of the vocabulary; such as, a specific facet or newly added definitions
- The creation of event logs that identify changes made to terms in the vocabulary
- The creation of customized reports that can filter or include or exclude various extended data elements in the vocabulary
- The creation of these reports as saved reports that can be executed periodically or to some schedule
- The export of the content of a report for inclusion in other tools such as Excel or Word

Access to these reports or to the tools to develop these reports is dependent on the permissions given through the EPA terminology system. Most report capabilities are available to Editors. Instructions on how to create reports are provided in the Editors' Training Manual. The EPA Terminology Services Coordinator should be contacted to discuss report needs (etss_admin@epa.gov). A list of pre-designed, saved reports is also available.

6.0   Integrating the Vocabulary with Other Systems

There are two ways to integrate vocabularies with other systems, whether manual or automated. The first is by exporting the vocabulary into a file which is then brought into the other system. The second is to use web services to interact automatically with the vocabulary in the EPA Terminology System.

6.1   Exporting the Vocabulary

Vocabularies to be used in other systems can be exported to a file for import into the other system. This method may be preferred if an approved or authoritative version of the content is to be released or published at certain intervals. This method is also advantageous when dealing with simple pick lists that must be tightly integrated with the other system. The use of exported vocabularies may also be preferred in cases where the other system does not support open services architectures, where the other system is on a network that is firewalled from the EPA terminology system, or where performance may be an issue.

6.2   APIs and Web Services

In addition to exporting a vocabulary to a file for transfer, display, or printing, a vocabulary can be accessed via a series of APIs or web services. Access to the APIs and web services is available from the Automated Services tab within the Terminology Services (Portal login is required). Over time, additional web services will be added to facilitate the use of the vocabularies in machine-to-machine environments.

A developer using a web service connection between an external system and the terminology system must have a special password. To obtain a web service password, contact the EPA Terminology Services Coordinator (etss_admin@epa.gov).

7.0   Getting Started

Now that you've learned the steps and recommended best practices for planning, developing, and maintaining a quality vocabulary, we encourage you to get started on your project. For consultation on your project or to schedule training, contact the Terminology Services Coordinator (etss_admin@epa.gov). Additional resources are also available from the Terminology Services web site.

As you go through the process of developing your own vocabulary, please consider sharing lessons learned (what worked/what didn't) and your approach to governance with others. Terminology Services can serve as a hub for access to this information. Contact the Terminology Services Coordinator (etss_admin@epa.gov) for details.

As EPA and its partners continue to develop vocabularies and make them available from Terminology Services, these vocabularies will serve as valuable resources to EPA, its partners, and the public, making it easier to find, manage, and share relevant scientific knowledge, and, thereby, improving and advancing EPA's efforts to improve human health and the environment.

Appendix A: Sample Charter

EPA Quality Glossary Governance Council Charter

Background

EPA's Quality Staff is responsible for coordinating the implementation of the Agency Quality System and ensuring that Agency quality issues are communicated effectively. To this end, the Quality Staff recognized the need for an enhanced and improved Agency Quality Glossary which, when complete, will facilitate a shared understanding of quality terminology across disciplines of quality management within the Agency.

For a glossary to be effective, comprehensive, and useful to its intended audiences, experts in quality terminology and management must be involved in the development and governance of the glossary. An oversight body known as the Quality Glossary Governance Council was formed on March 15, 2007, to serve this purpose.

Mission

The Quality Glossary Governance Council:

- **Makes critical decisions about the scope and organization of terminology contained within the quality glossary**
- **Determines the business rules for any additions or changes to the quality glossary Establishes the framework in which the glossary will be managed**
- **Acts as an arbitration board, or determines an arbitration process, to make decisions when multiple definitions exist for a single term**
- **Assigns authority to individuals responsible for governing changes and additions to terms in specific areas**

Membership

The following organizations are represented on the Governance Council:

- **Quality Staff**
- **Office of General Counsel (OCG)**
- **OEI Lead Region**
- **Office of Research and Development (ORD)**
- **OEI Back-Up Region**
- **Office of Environmental Information (OEI) (multiple members)**

In addition, the Office of the Chief Financial Officer (OCFO) and the Office of Acquisition Management (OAM) serve as advisors to the Governance Council. OEI's Office of Technology Operations and Planning (OTOP) serves as a subject matter expert in the area of information technology and provides related advice to the Governance Council.

Representation and participation is critical to the successful operation of the Governance Council. Members are responsible for attending all meetings of the Governance Council, participating in discussions, and contributing to the decision making process. Governance Council members are responsible for naming a delegate in the event of an absence from a meeting, discussion, or decision vote.

Length of Service and Succession

Governance Council members will serve for a minimum of three years. Permanent membership on the Council is required with a representative from: Quality Staff; OEI/OIAA; OEI/OIC; OGC; ORD; and the current OEI Lead and Back-Up Region. Representatives from these organizations may appoint successors from within their organization after a 3 year period. The Council can choose to appoint a successor for a departing member, or can assign the departing member the responsibility of naming a successor. The Council's Chairperson is selected through a nomination process and a Council vote. Council representatives from other non-permanent organizations may rotate after an 18-month period.

Operations and Activities

The Quality Glossary Governance Council communicates via email, meetings, and phone conferences. They may also operate using a portal collaboration environment.

At the introduction of each topic that requires a final decision from the Governance Council, voting and discussion rules will be established. Voting and discussion rules will be determined based on the impact of the decision. The Governance Council Chairperson has the final authority over any voting and decision rules, and is responsible for communicating these rules to the Governance Council before a decision is reached on the issue in question.

Council members will have a minimum of three (3) business days to review materials prior to a meeting, and a minimum of 10 business days to review documents that require detailed comments and/or a decision vote. A quorum or majority of votes is needed for actions requiring a decision.

Charter Review

The Governance Council will review this Charter after three years from the date of approval to re-evaluate the mission and role of the Governance Council.

Appendix B: Vocabulary Metadata

SUB-ELEMENT NAME	DEFINITION
Vocab Short Name*	The short name used for pick lists. This field is limited to 30 characters.
Vocabulary Name	The full name of the vocabulary.
Display Name	The name that will be displayed to identify the vocabulary on the search results or tree browse.
Vocabulary Type*	The classification type of the vocabulary. See table in section 3.0.
Description +	Information that describes the vocabulary and its purpose.
Vocabulary Unique ID (UID) *	Contains the Unique Identifier (UID) for the vocabulary taken from the Object Class Management page. Allows changes to be made to the name of the vocabulary without impacting the link between the Vocabulary metadata and the vocabulary itself.
Taskview *	Indicates the taskviews in which the vocabulary appears, separated by semicolons.
Steward	The person responsible for managing the vocabulary.
Steward E-mail	The e-mail address of the Steward.
Publishing Organization +	The organization responsible for publishing the vocabulary. This is often the organization in which the Steward resides.
External Organization	(Used as a FLAG) – Indicates those vocabularies that are published by an organization other than EPA. E=External Organization; blank=EPA Publishing Organization
Version	The edition of the vocabulary.
Source Creation Date *	The date the vocabulary was created.
Source Modification Date +	The modification date of the vocabulary.
Language	The name of the language or languages of the terms in the vocabulary.
Original Source	The name of the original source of the vocabulary. This may be the same as the publishing organization.
Namespace/URL for Source	The URL of the original source, if it is available online.
Program Related Information +	The URL of program related information that supports the vocabulary.
Terminology Services Vocabulary URL +	The URL for direct access to the vocabulary in the repository via the Terminology Services interface.
Alternative Name	Another name by which the vocabulary is known.
Client System	The system or systems that use the vocabulary.
Terms & Conditions	Disclaimer of use, including copyright notices or limitations on secondary distribution of all or part of the vocabulary.
Active Flag +	Indicates those vocabularies that are active and should be provided via the interface. Archival, in-process, or migrating vocabularies which have not yet been checked by quality control processes should be kept from displaying by the Inactive option. At this point, it will be the same as those vocabularies that are in the Active Task View. A=Active; blank= Inactive
Catalog Only	(Used as a FLAG) – Indicates those vocabularies that are not loaded into the Terminology Services repository but for which a record of the vocabulary is cataloged in the Vocabularies-Vocab Catalog metadata records. C=Catalog Only; blank=Exist in Terminology Services Repository or Inactive vocabulary

SUB-ELEMENT NAME	DEFINITION
Editorial Note	Any notation(s) that should be made about the status of the vocabulary.
Categories +	The topic or topics of the vocabulary selected from a controlled pick list.

*Required field for All Vocabularies
+Required field for Active Vocabularies

Appendix C: Permission Levels

General roles and functions	Synaptica description of roles and functions which are available via pick list user permissions in the Taskview Manager

"Public"
   Read Only Access
   Default Level Access
   Comment via standard comment box
	Read Only
	   Read Only Access

Editor
   Add and delete terms
   Add and modify relationships
   Add, delete, and modify other content   such as definitions
   Public Permissions
	Editor III
	   Merge items
	
	   Clone items
	
	   Reassign Items
	   Editor I, II Access
	   Indexer I, II, III Access
	   Lexicographer I, II, III Access
	   Read Only Access

System Administrator
   Set up vocabulary metadata
   Replicate vocabulary models
   Add data elements to vocabularies
   Control taskviews (user profiles)
   Set up user accounts
   Set permissions
   Editor Permissions
	Administrator I
	   Edit the taskview profile (TVP) metadata (For TVP's user is privileged to access)
	
	   Editor III Permissions
	
	Synaptica System Administration
	   Access to Admin Menu

Other permission that are available but currently not being used	

	Indexer I
	   Add INTER-object relationships (but not INTRA-object relationships)
	   Read Only Access
	Indexer II
	   Edit/Delete INTER-object relationships (but not INTRA-object relationships)
	   Indexer I Access
	Indexer III
	   Add/Edit/Delete INTRA-object relationships
	   Indexer II Access
	Lexicographer I
	   Add items (except for approved items)
	   Indexer III Access
	Lexicographer II
	   Edit Items
	   Lexicographer I Access

Other permission that are available but currently not being used	

	Lexicographer III
	    Delete items (except locked records)
	    Lexicographer II Access
	Editor I
	    Add items with approved status and promote unapproved terms to approved
	    Lexicographer III Access
	Editor II
	    Edit/Delete locked records
	    Lexicographer II Access

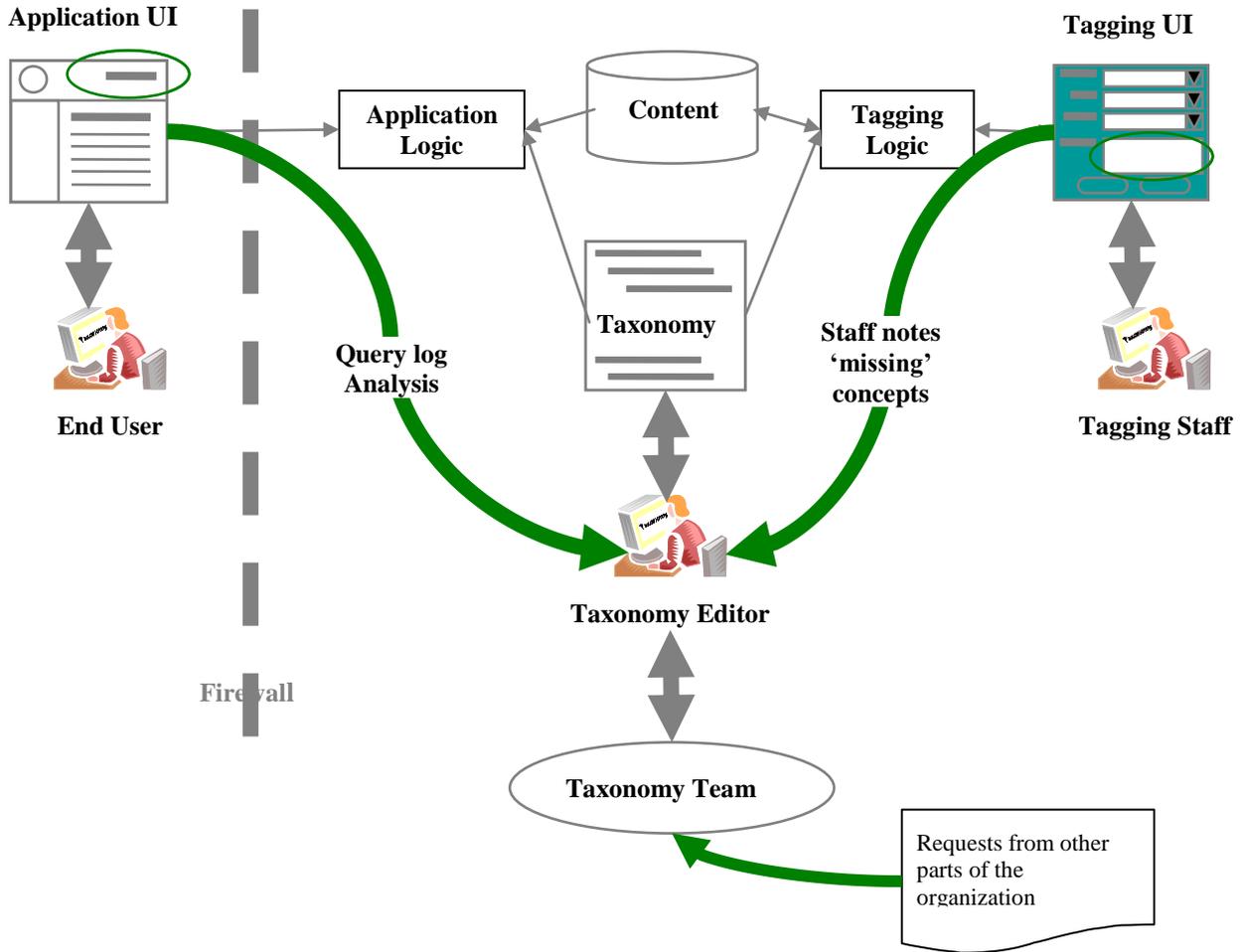# Appendix D: Access Rights in the Terminology Services Environment

Terminology Services is enabled by a three-tiered information technology environment within the EPA System of Registries. Levels of permission for user access are as follows:

- Public EPA Internet Terminology Services – No login required for searching, retrieving, and downloading EPA vocabularies that have been approved for release. The site also includes access to an international environmental thesaurus, a web taxonomy indexing feature, and other information on terminology groups, best practices, and how to obtain support.

- EPA and Partner Extranet – EPA Portal login and password (which is the same as EPA Webgate or Network login) are required for access. This includes all the capability of the public site plus a tool to allow users to create and maintain personalized glossaries from existing or new terms. The Extranet also supports groups developing or updating vocabularies by providing collaborative, member-only discussion forums such as the Terminology Services Semantic Media Wiki.

- Synaptica Tool – Synaptica-specific login, password, and training are required for Stewards and Editors creating or deleting vocabularies and managing terms and relationships within those vocabularies. If you have an EPA Portal login and password, a user account can be issued to you for the Synaptica software tool. Please note that if you are a contractor, you must have approval of the appropriate government vocabulary Steward and you must be supporting that Steward on a current, active contract.

Please note that remote access is also possible for those working off-site or those partners and contractors who are not on the EPA network. For additional information or to obtain access to EPA's Terminology System, contact the EPA Terminology Services Coordinator in the EPA Data Standards Branch by accessing the Terminology Services web site (http://www.epa.gov/ts) and submitting a request.

Appendix E: Sample Governance and Approval Structures

Governance Example 1: Taxonomy Committee Interactions
for the EPA Web Taxonomy

Application UI

Application Logic

Content

Tagging Logic

Tagging UI

End User

Query log Analysis

Taxonomy

Staff notes 'missing' concepts

Tagging Staff

Taxonomy Editor

Firewall

Taxonomy Team

Requests from other parts of the organization

Change Process for the Web Taxonomy

Legend
O – Originator
TE – Taxonomy Editor
TC – Taxonomy Committee

O ---> Submit Change Request

TE — Simple? — Yes → Change as REQUESTED ← **TE**

No

TE — Research/complete Change Request form

TC — Change? — Yes → Immediate? **TC**

No — Yes

TE — Inform Originator

Assign Priority **TC**

No
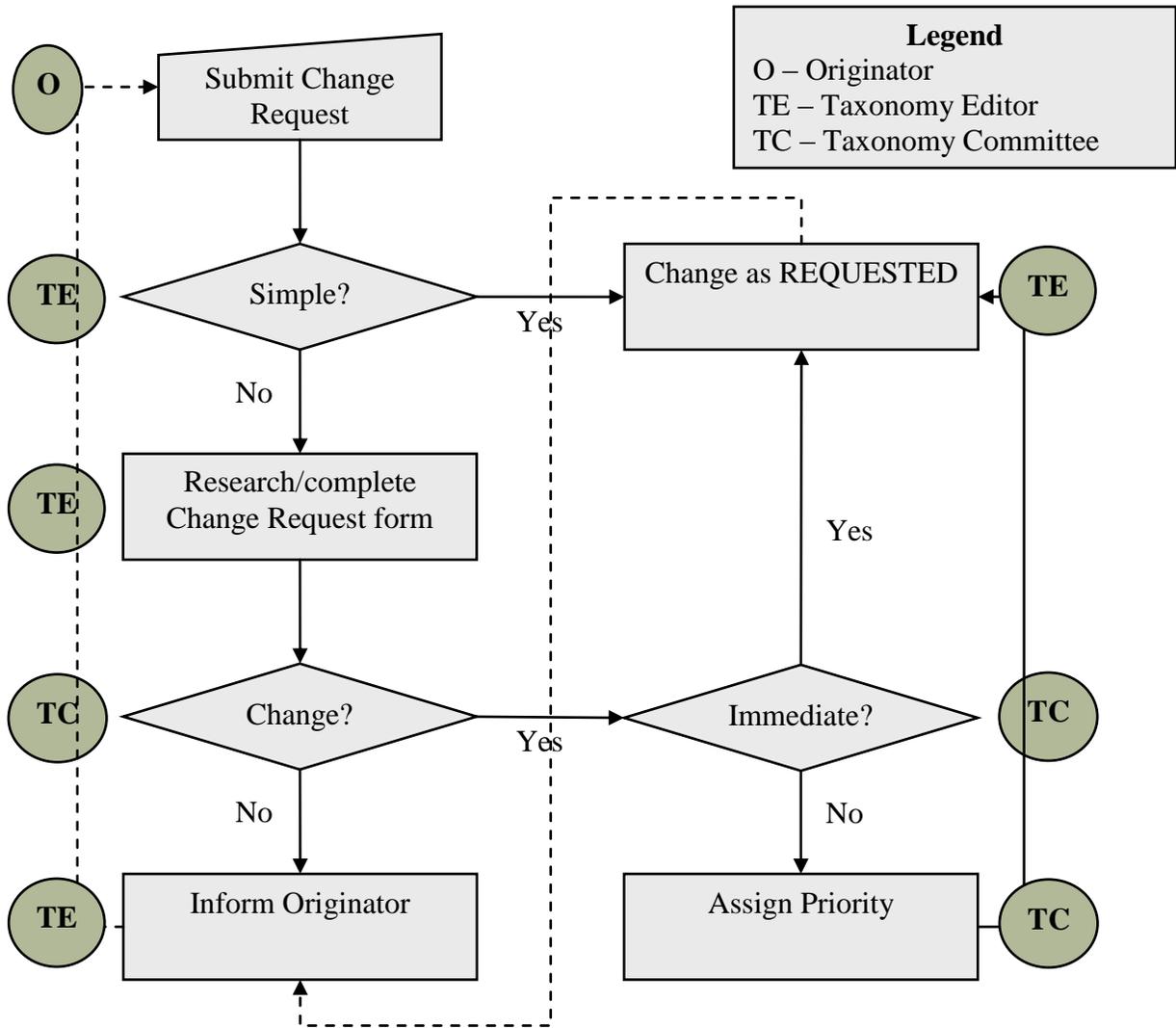
Governance Example 2: Quality Glossary Process for Vocabulary Development

Quality Glossary Term Normalization and Approval Process

Existing Quality Glossary Term

New Term Submitted

Change is Requested to a Current Term

Term is Assigned to a Data Steward based on Category

Quality Glossary + up to 4 other glossaries are searched
Data Steward may select appropriate glossaries

Data Steward searches glossaries for definitions

Other definitions found?

YES

Data Steward collects definitions and metadata in spreadsheet

NO

Data Steward submits materials to SME for review

YES

Sufficient metadata available?

NO

Further research for appropriate metadata at Data Steward's discretion

NOTE: Data Steward may choose to document the lack of metadata and submit the term anyway, or try to find the metadata

SME reviews materials based on glossary definition criteria

SME provides recommendation to Data Steward

The SME can accept a definition, accept a definition with modifications, or reject the term

Recommendation might be to reject a term and Governance Council may decide to accept it anyway; Assumption is that even term rejects must go to Governance Council for approval

Data Steward presents the recommendation to the Governance Council

Term Accepted?

YES

Term and/or definition added to the Quality Glossary

Data Steward modifies definition

YES with modifications

Governance Council accepts recommendation?

Yes

NO

Data Steward notifies term submitter with explanation for rejection

NO

Data Steward makes comments in the glossary entry as to why action was taken

DONE

Appendix F: Synaptica Term Level Metadata

Synaptica Term Fields	Definitions
System Fields[1]	
Object Class *	The vocabulary that the term resides in.
Active Status *	Identifies if the term status is active, deleted, or deactivated. This field supports the system business rules.
Sub-Element Fields[2]	
Descriptor *	Term in the vocabulary.
Def1, Def2, Def3, Def4, Def5 +	The meaning of the term. This field can also include the source of the definition in brackets following the meaning of the term, if the source is other than the vocabulary itself. Unless otherwise defined, a vocabulary will be set with 5 Def fields. They do not all need to be used.
Source Note +	Identifies the original source of the term.
Scope Note	Guidance for the use of the term. This field can also include concepts that should be included or excluded from the use of the term.
Editorial Note	This field is used as a comment field where editors can include important information about the term or the term authority.
Administrative Attributes[3]	
Approval Status *	Indicates the degree to which the term has been reviewed and accepted. The fixed values include: approval, unapproved, candidate (default), pending, and machine loaded. This field supports governance and vocabulary workflows.
Language *	Identifies the language of the term. English is the default for this field.

∗   Field required by Synaptica
+   Field recommended by Terminology Services

---

[1] System Fields are fields included in the item summary (term metadata) that apply to the business rules of the system, e.g., each term must have an Object Class.
[2] Sub-Elements Fields are the term and attributes that provide additional information (metadata) about the term such as the definition and note fields.
[3] Administrative Attributes are fixed and customized fields that support the governance and workflow for each term.

Appendix G: Relationships Types

The following table shows the Relationship Types that have been created to-date. The behavior of those terms in Synaptica result displays is described in the right-most column.

Relationship Type Code	Relationship Name	Definition	Results display (Synaptica)	Results display (Terminology Services Terms/Acronyms – Vocabulary Catalog)	Results display (Terminology Services Taxonomies – Web Taxonomy Search)
USE	Use	The relationship between a non-preferred term and a preferred term for the same concept.	Term included in default alpha display followed by (NPT). Not displayed on default hierarchical displays.	Term included in display with a reference to the preferred term.	Term included in display with a reference to the preferred term in Web Taxonomy Search only.
UF	Use For	The relationship between a preferred term and a non-preferred term for the same concept.	Term included in default alpha and hierarchical displays. Relationship shown in full record and with a special report.	Term included in display without reference to non-preferred term.	Term included in display without reference to non-preferred term.
CB	Coded By	Full Term for reciprocal coded value	Term included in default alpha and hierarchical displays. Relationship shown in full record and with a special report. Behaves like a preferred term in the UF-USE relationship.	Term included in display without reference to coded value.	Term included in display with reference to coded value in Web Taxonomy Search only.
CF	Code For	Coded value for the reciprocal full term	Term followed by (NPT). Included in default alpha displays but not hierarchical displays. Relationship shown in full record and with a special report. Behaves like the non-preferred term in a UF-USE relationship.	Term included in display without reference to full term.	Term included in display with reference to coded value in Web Taxonomy Search only.
AB	Abbreviated By	Full term associated with	Term included in default	Term included in display	No explicit display

Relationship Type Code	Relationship Name	Definition	Results display (Synaptica)	Results display (Terminology Services Terms/Acronyms – Vocabulary Catalog)	Results display (Terminology Services Taxonomies – Web Taxonomy Search)
		reciprocal abbreviation	alpha and hierarchy displays. Relationship shown in full record and with special report. Behaves like an RT relationship; neither term is non-preferred.	with reference to the acronym.	
AF	Abbreviation For	Abbreviation for reciprocal full term	Term included in default alpha and hierarchy displays. Relationship shown in full record and with special report. Behaves like an RT relationship; neither term is non-preferred.	Term included in display with full term.	No explicit display
RT	Related Term	Identifies terms with associated meanings	Term shown in all default lists. Relationship shown in full record or with a report.	No explicit display	No explicit display
NT	Narrower Term	Relationship between a term and a subordinate term	Term shown in all default lists. Relationship shown in hierarchical display.	Not displayed	Term displays in hierarchical view under Taxonomy browse.
BT	Broader Term	Relationship between a term and its superordinate term	Term shown in all default lists. Relationship shown in hierarchical display.	Not displayed	Term displays in hierarchical view under Taxonomy browse. Term displayed with breadcrumb trail in Web Taxonomy Search.
SY	Synonym	Identifies terms with identical or equivalent meaning.	Term shown in all default lists.	No explicit display. See also reference included in definitions when applicable.	No explicit display.
FNTT	Facet Name Top Term	Relationship between a facet term and its actual	Term shown in all default lists. Relationship shown in	Not displayed	Not displayed

Relationship Type Code	Relationship Name	Definition	Results display (Synaptica)	Results display (Terminology Services Terms/Acronyms – Vocabulary Catalog)	Results display (Terminology Services Taxonomies – Web Taxonomy Search)
		top term (Note: this field was created for interoperability purposes with the read-only interface)	hierarchical display.		
ATT	Actual Top Term	Relationship between a top level term in a taxonomy and its facet name term (Note: this field was created for interoperability purposes with the read-only interface)	Term shown in all default lists. Relationship shown in hierarchical display.	Not displayed	Not displayed
UL	Upper Level	Relationship between a term and another term that shows a hierarchical structure but is not a superordinate concept	Term shown in all default lists. Relationship shown in hierarchical display.	Not displayed	Not displayed
LL	Lower Level	Relationship between a term and another term that shows a hierarchical structure but is not a subordinate concept	Term shown in all default lists. Relationship shown in hierarchical display.	Not displayed	Not displayed