



# Zthes application profile

Version 1.7 October 2007

## Introduction

This document describes a specification for the exchange of vocabularies, provided as part of the Vocabulary Management Services supported by Becta.

Becta Vocabulary Studio is used for editing and managing controlled vocabularies used for UK education.

These are then published to Vocabulary Bank [<http://bank.vocman.com>]

The data model is an application profile of Zthes version 1.0 [<http://zthes.z3950.org>].

## Data Model

This data model explains the use of various parts of the scheme. It does not indicate the order of elements in the xml files, these will follow the Zthes schema.

**Zthes Label** = the name of an element allocated by Zthes and used in XML.

**Occ** = number of occurrences, the original Zthes data model values are given in square brackets if they are different.

**Name** = the name that is presented to users in the Studio user interface. The Zthes label will be used in XML. Some names are not needed as they are implicit in the way that information is presented, for example narrower and broader terms are indicated by icons and a tree structure.

Zthes Label	Occ	Name	Type	Description
termId	1	LocalId	string	An opaque string of characters which uniquely identifies the term within the thesaurus.  This can be allocated by the authority which 'owns' the term or created automatically by Studio or another service.
termName	1	Name	string	The name of the term in a form which may be displayed to a user or used as a search term in a target database.
relation	0+	(implicit – see below)	sub-record	A sub-record briefly describing a term related to this one. Each relation sub-record may carry a weight indicating how strongly the related term is related to the main term.



Zthes Label	Occ	Name	Type	Description
rmNote	0+	(see below)	various	<p>A note about the term: that is, arbitrary prose clarifying the meaning and scope of the term. Multiple notes may be included. Each note may carry a label indicating its role, e.g. scope, source. In the common case where only a scope note is required, no label need be included.</p> <p>Each note may also carry an indication of a constraint on the vocabulary of its content, e.g. indicating that the content of a specific note must be drawn from another Zthes thesaurus, or a plain list. The precise expression of this constraint is dependent on the encoding of the Zthes records: see for example the discussion of this constraint in the XML specification.</p>
termCategory	0,1 [0+]	Subject category	vocab	Identifies a term as belonging to a particular topical subset (microthesaurus).
termLanguage	0,1	Language	vocab	<p>The language of the term.</p> <p>This should be from a standard ISO controlled vocabulary (two or three letter with or without extensions).</p>
termQualifier	0,1	Qualifier	string	An additional string which, if supplied, qualifies the term name such that the combination of term and qualifier is unique within the thesaurus.
termSortkey	0,1	SortKey	string	An explicit sort key for the term, based on language- and application-specific sorting rules that may remove leading articles, parse number names, etc., to produce natural order not attainable through a strict alphanumeric sort.
termStatus	0,1	<i>Status</i>	string	<p>The status of the term, which may be active, deactivated or deleted. In general, only active terms should be used to guide searching. A deactivated term (which may have history notes explaining why it was withdrawn) can be reinstated – a user should be prevented from adding a term that is identical to a deactivated term. Deleted terms can be ignored for all purposes except reconstructing thesaurus history.</p>
termType	0,1	<i>implicit</i>	vocab	<p>An indication of the type of the term, chosen from the controlled vocabulary (PT, ND, NL).</p> <p>In Studio, NL is indicated by a check box. PT and ND indicated by relationship and lack of node label.</p>



## Relation Sub-Record

Each relation sub-record contains the following:

Zthes Label	Occ	Type	Description
relationType	1	vocab	An indication of the type of the relation, chosen from the controlled vocabulary (BT,NT,USE,UF,RT,LE)  Implicit in Studio
termId	1	string	The unique identifier of the related term within its database.
termName	1	string	The name of the related term.
relNote	0,1	vocab	An element that is used to qualify the relationship selected from a controlled vocabulary. The structure will be the same as termNote.

## Extensions

The following are proposed extensions to Zthes v1.0.

### Relationships

There are use cases where different types of relationships need to be identified. Rather than use the x-??? extension suggested by Zthes v1.0 a new element called relNote will be used to carry this information. This will be of the same structure as the extended termNote.

### Translations

If there is a term that is a non-preferred translation of another term this is indicated by linguistic equivalent LE.

There are use cases when multilingual users will need a translation of the definition or scope of a term, rather than a separate term. For this use case, a language extension will be added to termNote, relNote and thesNote. This will use the xml lang attribute with values taken from the ISO standard two- or three-letter codes with extensions.



## TermNotes

It is possible to include several values in termNote by providing a label as a qualifier. If the value comes from a controlled vocabulary then the source can be provided. TermNote is used for scope if unqualified.

The termNote information (and thesNote and relNote) consists of a value plus optional label, source and language.

The Studio will use the following list of termNote labels.

<b>termNote Label</b>	<b>Occ</b>	<b>Type</b>	<b>Description</b>
authority	1	vocab	A Studio Vocabulary for approved authorities. The Authority name from Studio will be used in the termNote. This will be created automatically for new terms based on the user's authority.
source	1	string	Identifies the resolvable global identifier of the Becta Studio reference vocabulary for this term. This will be automatically generated on import or term creation.
globallyUniqueld	1	string	An identifier that is unique within Studio made up of the authority identifier and the local identifier. This will be used for resolution using the Bank web service.
definition	0,1	string	A definition of the concept represented by the termName. Currently (and for the next release schedule) Studio only supports one definition but if multilingual termNotes are supported then more than one definition may be added. It is expected that only one definition will be provided in any given language.
scope	0,1	string	The scope of the use of the term in this vocabulary
history	0,1	string	Information about changes made to the term.
curriculumType	0,1	vocab	A label taken from a controlled vocabulary, for example objective, topic or category, used to indicate the type of term from a curriculum perspective.
displayNote	0,1	string	A note associated with a term, that may be used for display purposes.
longTerm	0,1	string	A term name that is longer than can be easily managed in Studio or user interfaces, but needs to be retained for completeness - eg parts of a curriculum document.
indexCode	0,1	string	A set of characters that are used as a prefix or suffix. E.g. 5a iii



## Import and Export

The Studio Zthes schema profile will be used to export vocabularies.

Elements in Zthes vocabularies but not in the Studio profile will not be used until there is a use case that requires the additional elements.

When importing vocabularies, Zthes elements will map to the equivalent Studio elements.

- Zthes elements not catered for in Studio will be discarded.
- termNotes will map to equivalent Studio elements with the same termNote label.
- termNotes that have no matching Studio label will be discarded.
- An termNote with no label (for example in Zthes 0.5 format) will normally map to scope.

## Thes element Descriptor

The following fields are used to describe a vocabulary or thesaurus. Other fields described in Zthes are not used.

Zthes Label	Studio Name	Occ	Type	Description
dc:identifier	LocalId	1	string	The identifier for the vocabulary allocated by the authority.
dc:title	Name	1	string	The name of the vocabulary.
dc:date	PublicationDate	1	date	The publication date  Created automatically on export to the Bank.
dc:language	Language	0,1	vocab	The language of the vocabulary.  From a standard ISO controlled vocabulary (two or three letter with extensions). Studio only supports one at present but could be in multiple languages in future.
dc:rights	Rights	0,1	string	Rights statement governing the usage of the vocabulary.
dc:description	Description	0,1	string	A description of the vocabulary.  The Studio supports only one at present, but descriptions could be provided in multiple languages in future.



## ThesNote extensions

Studio will use the following list of thesNote labels with the same structure as termNote.

thesNote Label	Occ	Type	Description
authority	1	vocab	A Studio Vocabulary for approved authorities. The Authority name from Studio will be used in the thesNote. This will be created automatically for new vocabularies based on the user's authority.
globallyUniqueid	1	string	An identifier that is unique within Studio made up of the authority identifier and the local identifier. This will be used for resolution using the Bank web service. This will be created automatically.
version	0,1	string	The version of this vocabulary.

## Appendix

The following Zthes term fields are not currently used. They are included here for completeness and can be referenced if extensions are required.

Zthes Label	Description
postings	A sub-record indicating the frequency with which the term occurs in a target database.
termApproval	<p>An indication of whether the term has been approved for inclusion in the thesaurus, has merely been proposed and awaits approval, or has been considered and rejected. Acceptable values are <b>candidate</b>, <b>approved</b> and <b>rejected</b>.</p> <p>termStatus and termApproval are orthogonal, despite their surface similarity: the three values that each can take make up nine possible combinations, and all of these are distinct.</p> <p>termApproval is held in Studio but not currently exported, as only approved terms are exported.</p>
termCreatedBy	<p>The name of the person who created the record defining the term.</p> <p>The Studio uses 'authority' to indicate the owner of a term. Only users from an appropriate authority can edit terms.</p>
termCreatedDate	<p>The date on which the record defining the term was created.</p> <p>The Studio retains information about the date when terms are created but does not import or export this information.</p>
termModifiedBy	<p>The name of the person who last modified the record defining the term.</p> <p>The Studio uses 'authority' to indicated the owner of a term. Only users from an appropriate authority can edit terms.</p>
termModifiedDate	<p>The date on which the record defining the term was last modified.</p> <p>The Studio retains information about the date that terms are modified but does not import or export this information.</p>



termUpdate	The update status of the record, which may be <b>add</b> or <b>delete</b> . This is used when a collection of Zthes terms are used to specify an update to a thesaurus. In this scenario, terms in the update set that have termUpdate=add are added to the destination thesaurus, with each new term replacing any existing term that has the same termId. Terms in the update set that have termUpdate=delete cause the terms in the destination thesaurus with the corresponding termId to be deleted. For terms that are to be deleted the records may omit elements apart from the termId and termUpdate=delete.
termVocabulary	An indication of the vocabulary the term is drawn from, if the thesaurus contains multiple vocabularies.

The following Zthes relation sub-record fields are not currently used.

Zthes Label	Description
sourceDb	If specified, details identifying a service that provides the target database in which the related term is found. If not specified, the related term is in the same database as the current one.
termLanguage	The language of the related term.
termQualifier	The qualifier of the related term.
termType	The type of the related term.
weight	A value of the strength of the relationship. Similar to equivalence.