Crosswalking from Terminology to Terminology: Leveraging Semantic Information across Communities of Practice

Sue Ellen Wright and Dave Summers
Kent State University Institute for Applied Linguistics
109 Satterfield Hall, Kent, Ohio 44242 USA
E-mail: sellenwright@gmail.com, dgsummers@gmail.com

Abstract

Both subject language terminologies (controlled vocabularies) as represented by the Simple Knowledge Organization Scheme of the W3C (SKOS) and discourse-related terminologies as elaborated by ISO TC 37 (TBX) contain terms, definitions, and semantic relations, which could under ideal circumstances be leveraged across applications and platforms in order to provide semantic anchors for conceptual references when creating knowledge resources in web environments. This paper proposes a cross-mapping between comparable elements in the two environments, taking into account both close similarities and significant differences in the semantic content of important elements on either side of the equation. The final evaluation proposes that the richer concept ordering environment of TC 37 concept systems requires the use of OWL-Lite in order to accommodate all levels of granularity.

1. Terminologies and communities of practice

Since the Terminology and Knowledge Engineering (TKE) conference in 2005, a working group that grew out of the SALT project (SALT, 2001; Budin and Melby, 2000) has envisioned a crosswalk between the Simple Knowledge Organization System (SKOS) (SKOS Reference and SKOS Primer, 2008)[1] and the ISO 12620:1999 / LISA TBX[2] (ISO DIS 32042:2008) standards. In the emerging environment of the Semantic and Pragmatic Web, there is keen interest in options for leveraging existing knowledge ordering schemes from a variety of knowledge representation resources (KRRs). Differences in communities of practice going back deep into the previous century have resulted in incompatible approaches to the representation of terms, definitions, and relationships between concepts, often based on different intentional aims. Although these differences may be known on a superficial level, it is not always obvious where specific discontinuities lie. As a consequence, despite similarities in surface-level vocabulary usage and apparent goals, cross-leveraging of linguistic and semantic information among different types of resources that lay claim to the "terminological" approach is not trivial. Hence the apparent tautology in the title for this paper references the incompatibility of usage between different communities that use the term *terminology*. We will distinguish between *subject-language terminologies* (SLTs; Svenonius, 2000), which reflect practice in the SKOS community, and *discourse-oriented terminologies*, also called *language-purposed vocabulary* (Tudhope et al., 2006, 26), or terminological databases (termbases, TDBs), which are the province of the TBX environment.

[1] Given the time frame involved with the submission of this abstract, it does not reflect an in-depth study of the latest SKOS draft specifications listed in the references.
[2] Check *References* for missing acronym full forms.

ISO Technical Committee (TC) 37, *Terminology and other content and language resources* defines *terminology* in ISO 1087-1:2000 as a "set of designations belonging to a special language", whereby these designations reflect the concepts used in that special language and are represented by *terms*: "verbal designations of general concepts in a specific subject field." Although not specifically stated here, *terminologies* in the sense of TC 37 document these designations for use in discourse, primarily in written texts. The range of approaches commonly covered by such terminologies comprises a continuum of resources reflecting increasing degrees of control and prescriptivity. On the far left side of such a cline, truly *uncontrolled vocabulary* native to a domain is used for both oral and written communication. Although this usage is specialized, it is also a subset of natural language and is embedded in general language. Still within the realm of natural language, but gradually moving further away from general language, there are standardized terminologies created as pre-negotiated, consensus-based resources for vocabulary used in official documents of various kinds, as well as rigorously constrained terminology intended for use in controlled English or other controlled languages (O'Brien, 2003). Here the prescription of grammatical usage, syntax and style moves discourse-oriented terminology further away from general language, but without crossing totally over into the status of artificial language.

The scope statement for SKOS has undergone a number of iterations, whereby the latest definition states, "SKOS – Simple Knowledge Organization System – provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other types of controlled vocabulary" (SKOS Primer, 2008). Earlier references to terminologies and terminological concept systems have wisely been removed from the definition. Nevertheless, when we examine the various standards and literature of the documentation community, *terms* and *terminology* appear in conjunction with controlled vocabularies. The formal definition of *term* in NISO Z39.19 (2005) is deceivingly similar to that in ISO

1087: "One or more words designating a concept." Here the official designation for a set of these particular terms is *controlled vocabulary*: *A list of terms that have been enumerated explicitly.* As an almost obvious consequence, such collections of terms are also widely called *terminologies,* even in the standard. This standard, as well as most resources in the field, does not distinguish between terminologies that represent the language of relatively uncontrolled discourse in special subject fields on the one hand and the notion of controlled vocabularies, such as thesauri, on the other. SKOS-type resources are primarily used to label and retrieve information or resources and are made available in the form of *terminology services* (Tudhope et al., 2007).

In an effort to distinguish between terminologies and natural language lexicons, Svenonius attempts to sort out this ambiguity by designating controlled vocabularies as *subject languages*, which she contrasts to natural language. "Subject languages are artificial languages, designed for the special purpose of retrieving information. As such they differ in certain essential respects from natural languages… In a natural language the extension, or extensional meaning, of a word is the class of entities denoted by that word, such as the class consisting of all butterflies. In a subject language the extension of a term is the class of all documents about what the term denotes, such as all documents about butterflies." (Svenonius, 2000, 129-132.)

Svenonius also stresses the need for univocality (one concept, one term) in subject language terminologies, a requirement that echoes the Z39.19 requirement that terms in controlled vocabularies be clarified as *preferred*, with synonyms mapped to preferred terms for search and retrieval purposes. This distinction contrasts practice in discourse-oriented terminology management, where one has the option of assigning a preferred attribute to a term, especially for prescriptive or controlled resources, but where many synonyms are admitted, and even deprecated terms are fully documented. The assumption here is that such non-preferred terms are not just pointers to descriptor terms, but actually occur in texts, depending on register, usage community, and text purpose.

In this context, Tudhope et al. (2006) cite *lexical databases* used for linguistic purposes as resources that are compatible with controlled vocabularies, but it should be noted that they cite *Wordnet*, with its "synset" notations, as their primary example of such a resource, which significantly narrows the understanding of what is meant by *lexical database*.

2. Fundamental distinctions

Despite Svenonius insightful distinction, she does not account for the fact that in addition to subject language "terminologies" (our SLTs) and general language lexical resources, there is also a class of resources commonly called *terminologies* that:

- are concept-based;
- document special languages;
- are only in some cases prescriptive by establishing univocality (one term, one concept);
- frequently document levels of usage and register, thus designating multiple preferred terms subject to multiple pragmatic constraints;
- are designed for use by authors and translators creating text, but not necessarily for information, document, or other types of object retrieval.

Distinctions between TDBs and SLTs are critical for efforts to map TBX data categories to SKOS elements because the *terms* included in the two systems do not necessarily represent the same conceptual content, although valuable referential relations between the two approaches exist.

TDBs run the gamut from multilingual, strongly text-oriented bilingual glossaries created on the fly for project-related purposes in the localization industry to extensive enterprise-wide or government-sponsored term banks. Some feature little explicit systematic content in the form of potentially ontological information, while others elaborate concept systems and semantic networks. In attempting to categorize knowledge representation resources with respect to relative systematicity, Wright (2007, 159) identifies TDBs as a continuum including both systematic and non-systematic approaches. Insofar as some terminologies elaborate concept fields and networks, they are akin in some degree to other concept schemes used for knowledge representation. As shown in Table 1, systematic TDBs explicate a variety of relations, most prominently generic hierarchies, meronymy and metonymy, sequentiality, and a variety of other freer types. In addition to the fairly constrained set cited in ISO 12620:1999, Nuopponen documents over 40 terminal nodes in her classification of ontological concept relations used in TDBs (Nuopponen, 2005, 213).

Although some smaller terminological collections may be ordered into a single cohesive concept system, many large dynamic systems are more likely to identify shallow concept fields, or simply designate parent concepts or simple relations. Indeed, it is a feature of so-called "ad hoc terminology management," which is strongly text or corpus-based, that elaborators of TDBs frequently have difficulty recognizing precise concept relations until a certain critical mass of concept entries has been collected for a given subject field. Current thinking postulates the possibility of moving the generation of concept systems outside the core structure of terminology management systems, using persistent identifiers to reference individual concept entries within the system in order to provide authoritative definitions and other information for conceptual references in relational or knowledge resources (ISO WD 24618, 2008).

3. What might be leveraged

Tudhope et al. document current efforts to create terminology web services that are in some degree interactive. Given the challenges involved in solving accessibility issues for different controlled vocabulary networks, any effort to link TDBs and SLTs needs to define potential mappable contact points between the two approaches. As illustrated in Table 1, there are enough differences between SKOS and TDB elements to require a conscious, potentially complex mapping strategy. Although it is not the purpose of this paper to outline an implementation of RDF representation for TDBs, some sort of RDF representation must be the first step towards realizing the general availability of such content within the terminology services environment. Before moving to this step, however, those data elements (data categories) that are of mutual interest must be identified, and methodologies for reliable, persistent access must be resolved. Table 1 represents an initial approach to this sort of mapping. Rather than present elements in the familiar alphabetical order found in the two standards, we have chosen to group similar elements together in order to facilitate discussion. We do not, however, propose any doctrinaire attachment to this presentation as an ordering scheme per se.

The core data categories contained in most terminological resources, and the ones that are most critical for any kind of semantic cross-usability, are *terms* and *definitions*. The issue arises as to whether terms map to terms (labels) and definitions map to definitions, which will be examined below. It should also be noted in this context that metadata registries (MDRs) also contain rigorous definitions prepared by subject-field experts which are also made available in web environments as anchors for the disambiguation of references in semantic resources (see Windhouwer et al., 2008), but this consideration goes beyond the scope of the current paper.

For its part, TBX "is designed to support the analysis, representation, dissemination, and exchange of information from terminological databases (termbases)" (ISO DIS 30045:2008). The data categories listed in the TBX standard are taken in part from ISO 12620:1999 and have been modified to some extent for purposes of simplification. In future they will reside as a defined data category specification (DCS) in the TC 37 **ISOcat** metadata registry. It is not our intention to replicate functionalities across communities of practice, but rather to leverage specific items of information or to reference authoritative documentation residing in diverse resources. Data categories can be grouped as follows:

- *Specific references to controlled vocabulary*

Whereas ISO 12620:1999 contained a list of items related to controlled vocabularies (*thesaurus name, thesaurus descriptor, top term, broader term, narrower term*), the current TBX specification includes only *thesaurus descriptor*. The elimination of other items reflects in part an effort to avoid replicating the functionality of a controlled vocabulary, something that is seldom done and is actually seldom desirable. A simple cross-reference to a label that matches the term or concept in question is a more efficient approach in that it leaves positional specification within the controlled vocabulary environment. In our view, however, the elimination of *thesaurus name* (for which we would suggest the substitution of the SKOS element *inScheme*) is unfortunate. The reason for this proposal is that concepts and terms documented in TDBs can easily be associated with multiple concept schemes expressed in multiple controlled vocabularies. It would be highly useful, even for diverse concept systems within the local environment of a TDB, to be able to indicate the specific scheme referenced by a relation, although it can be assumed that for external references, the URI used to link to the targeted item will provide unequivocal identification for the resource involved.

Within the framework of the discussion surrounding TDBs and SKOS, one suggestion that has been made is to incorporate SKOS elements into TBX. Another, which has been reflected in earlier stages of our own work, is to identify missing elements and selectively add them to TBX. At this juncture, we would propose instead that the entire SKOS element set be added to the **ISOcat** registry, where it would be available for all TC 37 language resources. There is a precedent for this kind of insertion in that other standards, such as Dublin Core (2008) and the Language Codes (ISO 639 family of standards) have also been incorporated into the DCR.

It is also not highly likely that the fairly direct link between the thesaurus-related elements cited ISO 12620:1999 and the SKOS environment will be particularly fruitful as an avenue for mapping termbase elements to SKOS elements because very few TDBs utilize this option, with perhaps the possible exception of some national term banks that may map to authoritative national thesauri or classification schemes. Thus we have dispensed with the old *documentation language* elements from 12620:1999 and focus on those data categories that are more widely used and that consequently may be present for use in any future mapping exercise.

- *Labels and terms*

The fundamental difference between SLTs and TDBs is nowhere more critical than with respect to the relationship between SKOS labels and termbase terms. Although overtly prescriptive termbase environments specify the term status attribute as a required value, many do not, which can mean, for instance, that concept entries that only contain one term in a language will not be differentiated as *preferred*. Furthermore, TBX offers a variety of ways to distinguish synonyms in a concept entry, for instance by indicating sortable subset information (i.e., *customer, project, product, application, environment, businessUnit, security* (A.10.3)), any of which could be used instead of a status attribute for purposes of indicating the preferential status of a term in a particular context. Furthermore, and perhaps more importantly, the status attribute or any of these sorting attributes, can be used in TBX terminologies to indicate appropriate usage in discourse, whereas the *preferred, alternate,* and *hidden* elements in SKOS reflect knowledge retrieval strategies rather than discourse usage. Well-structured label sets in the SKOS environment should ideally reflect synonym sets such as those presented in concept-oriented termbase entries, but this has not always been the case in the past, where subordinate concepts or closely related concepts are grouped together with respect to a label in order to increase the concision of controlled vocabularies.

Another factor that plays a role here is that TBX and **ISOcat** attributes provide a much finer granularity for distinguishing term types than is afforded by SKOS labels, reflecting the degree of specificity required in text-oriented situations. Hence, terms in TBX are further specified according to twenty some term types (e.g., abbreviation, full form, and symbol, to name a few). Another factor, of course, is that the assignment of the *preferred* or *admitted* attributes in either case is a function of usage or system design, so data integrity in exchanging information between the two system types is compromised by the need to distinguish two properties (*termType* and *status*) on the TBX side of the equation for each of the individual elements on the SKOS side. One missing element on the TBX side, interestingly, is the inclusion of misspellings in the SKOS *hidden label* element. Although this could conceivably be documented in TBX as a *deprecated term* that is also a *variant*, perhaps the addition of *misspelling* as an optional value of *termType* might be interesting even outside the interoperability environment. Given the considerations cited here, it seems most expedient to ignore the option of identifying the status of a label from the TBX side and only try to map to a SKOS label (of whatever class) that matches the TBX term at the string + subject field level.

- *Definitions*

Although SKOS classifies definitions as one of several notes, definitions play a central role in TDBs, so we have pulled them out for special discussion. We have rather arbitrarily left *scopeNote* for the time being together with *notes* in Table 1, but it is impossible to discuss rigorous terminological definitions without attempting to distinguish them from (or indeed, equate them to) scope notes. Unfortunately, the various recommendations and examples provided in the literature do not make this a simple task. *Definition* in SKOS is defined as: *A statement or formal explanation of the meaning of a concept*, and *ScopeNote* is: *A note that helps to clarify the meaning of a concept*. (Miles and Brickley, 2005). This distinction supports our tentative mapping of *skos:scopeNote* to TBX *example*. The examples provided in the context of SKOS Core, however, project a confusing image:

```
<skos:Concept
rdf:about="http://my.example.org/GCL/702#s
copeNote">
  <skos:prefLabel xml:lang="en">
Competitiveness</skos:prefLabel>
  <skos:scopeNote xml:lang="en">The ability
of businesses to compete in local, national
or international markets.</skos:scopeNote>
</skos:Concept>

<skos:Concept
rdf:about="http://www.example.com/concepts
#banana">
  <skos:prefLabel xml:lang="en">
banana</skos:prefLabel>
  <skos:definition xml:lang="en">A long
curved fruit with a yellow skin and soft,
sweet white flesh inside.</skos:definition>
</skos:Concept>
```

In terms of definition theory, for instance as specified in ISO 704 (2000), both the definition and the scope note shown here are more or less rigorous definitions that define the discourse-oriented concepts represented by *competitiveness* and *banana*, respectively. Further efforts to clarify this distinction lead to O'Reilly's xml.com webpage, where the hope for an authoritative explanation remains, alas, unfulfilled:

To clarify the difference between skos:definition and skos:scopeNote, a definition should be an attempt to completely explain the meaning of a concept, whereas a scope note may consist of partial information about what is or is not included within the meaning (or scope) of a concept.

```
<skos:definition>A feature type category
for places such as the Erie Canal
</skos:definition>

<skos:scopeNote>Manmade waterway used by
watercraft or for drainage, irrigation,
mining, or water power</skos:scopeNote>
```

(Mikhalenko 2005)

Interestingly, from the standpoint of ISO 704, this second *skos:scopeNote* is actually the more appropriate definition with respect to discourse-oriented TDB approach, while the *skos:definition* provides an SLT-style definition of the feature as a label in a controlled vocabulary. This distinction looks very useful to us, but it would lead us to map the TBX *definition* to a SKOS *scopeNote*. Needless to say, this is an issue that needs to be resolved at some point, particularly if it reflects differences in practice within the SKOS community.

- *Note*

As indicated in Table 1, SKOS specifies a variety of notes, most of which can be expressed in the TBX framework by positioning the *note* element inside various containers in the TBX metamodel. (We do not have the space to illustrate this feature in the context of this paper.) *Example* in TBX matches well to the *skos:example*, but is considered to be a core descriptive element used to delimit the concept treated by a terminological entry, so it is not classified as *note* material in **ISOcat**. This distinction is not problematic, however, with respect to mapping from one system to the other.

- *ConceptScheme / Concept System*

All TBX *concept systems* (defined as: *The structured set of concepts established according to the relations between them, each concept being determined by its position in the set.* (ISO 12620:1999)) can be classified as *skos:concept Schemes* (*A set of concepts, optionally including statements about semantic relationships between those concepts.* (SKOS Core)), but the *optional* component of the SKOS definition means that not all concept schemes are concept systems in the sense of ISO 12620 or ISO 1087-1. Since each concept in a termbase is treated in its own entry, the non-mnemonic entry identifier, which could ideally be configured as a persistent identifier, serves as the concept identifier and the means whereby the entry is accessed either internally within the native termbase structure or from external resources via some sort of URI-type link. As noted above, the *inScheme* element provided by SKOS has, in our view, tremendous potential for use in TBX quite apart from any crosswalk, because the concept entries in a TDB generally exist independently of any given concept system and can participate in multiple schemes even within their native environments, which renders position notations within an unnamed scheme or simple indications of parent-child relationships potentially ambiguous.

- *Subject / SubjectField*

SKOS subject-related elements are designed to account for point of view, pointing both from a concept entry to a subject marker with which it is associated or enabling the assembly of a set of entries associated with a given subject. Here, however, the distinction between SLTs and TDBs resurfaces as a significant issue: in SKOS, where the label represents a resource, the *subject* is the subject or one of the subjects treated by that resource. In TBX, the *subject Field* is one or more of the specialized disciplines with which the concept treated in an entry is associated. This distinction is critical to the specification of the concept via the definition and other descriptive elements, because these items are only considered valid within the framework of the subject or subjects declared for a given concept entry. Although some TDBs allow for multiple subject field assignments in cases where term-concept pairs are shared by more than one declared subject field, it is not common practice in TDB management to differentiate primary or preferred subject fields. Despite these differences, the use of *subject/subjectField* attributes in determining search or mapping criteria is likely to be useful for purposes of semantic retrieval. Nevertheless, this is a problematic area even when matching controlled vocabulary to controlled vocabulary or termbase to termbase because of the frequent ambiguities and discontinuities involved in the specification of subject field categories across application boundaries. The most valuable environment for such matching would require a shared subject classification model.

- *SemanticRelation / Concept relations*

On the one hand, concept relations would appear to be a critical focus for this discussion, which is essentially centered around leveraging semantic information across methodological boundaries. But by the same token, the subtle differences between the systems pose the real risk that the equivalences proposed in Table 1 may be highly deceiving. At the most apparent level, the equation of *has TopTerm* with *broaderConceptGeneric* is potentially dangerous because a broader concept is not necessarily a top term in a scheme or system: *A concept two or more levels of abstraction higher than subject concept in a hierarchical concept system* (ISO 12620:1999). This designation is frequently used when formulating rigorous definitions if the immediate superordinate concept is deemed to be too specific or unfamiliar for use as a transparent genus element. Broader and narrower concepts map fairly safely to superordinate and subordinate generic concepts (*isA* relations), as long as one bears in mind the distinctions already cited between the conceptual extensions of labels and terms.

- *Collections, collectable properties, and members*

Currently there is a serious discussion in the SKOS community attempting to differentiate collections and concept schemes, a concern which to a certain extent alarms the authors, because we had thought we understood the distinction and felt it fit comfortably into TBX structures. In our way of viewing things, a *skos:collection* is any set of coordinate concepts in a generic system assembled to represent a set of concepts that comprise siblings dependent on a superordinate node. Typically termbases do not necessarily explicate such sets, but they are nevertheless often generatable at the user interface level based on the specification of the parent node as a search criterion for the data category *superordinateConcept Generic*, as evidenced by our mapping in Table 1. Any such subordinate concept is automatically a member of such a set. ISO 12620/TBX has no facility for creating ordered collections, i.e., for imposing ordering rules on such a dynamically assembled set, except in those instances where a *conceptPosition* (A.7.2) number is specified in the data model, in which case this number can be used to impose order in the set. This kind of ordering, however, can be very difficult to achieve in large, dynamically changing TDBs because it would probably entail human intervention, together with an overview of the various siblings involved..

- *Related concepts*

Although simple *skos:related* would appear to map comfortably to *relatedConcept,* from this point on in Table 1 it becomes apparent that TBX and termbase solutions on the whole provide a richer set of relation types than is necessary for *simple* knowledge organization within the framework of controlled vocabularies. Currently we visualize using OWL-Lite in order to facilitate the expression of more diverse relation types. Indeed, we propose *relationType* as an addition to the DCR, although one can also make an argument for moving more complex relations outside the core of a termbase, in which case relation resources using OWL technology could be built external to a termbase, using persistent identifiers to anchor RDF relation specifications via the concept entry IDs embedded in terminological entries. This scenario has also been proposed for linking relation resources to metadata registries (Wittenberg, 2007).

4. Who would want to do this?

The idea of information interchange has always inspired a certain population of nay-sayers. In the 1980s, it was not unusual to encounter resistance to the elaboration and deployment of data interchange standards. Today we see a wide variety of such standards in an equally broad number of application areas: DITA, XLIFF, LMF, TBX, TMX, SBVR. The advantage of these kinds of prenegotiated markup formats is that they facilitate automatic processing and the ability to leverage information from diverse resources residing in distributed networks and on different platforms. Nevertheless, the very variety of formats presents issues with respect to interoperability and exchangeability, particularly in dynamic environments. XML provides the ability to reference material expressed using one XML application inside resources that are formatted in some other XML application. This capability suggests the advisability of utilizing existing formats for specified functionalities rather than building those functionalities into other formats.

Although this powerful option enables leveraging to a certain extent, it does not necessarily ensure true interoperability, given the fact that even very closely related environments can nonetheless employ dramatically different approaches to conceptual ordering systems. Within communities of practice, mapping, crosswalk, and even mashup strategies are designed to attain more or less lossless interchange – either on a dynamic or a snapshot basis. This paper is proposing an approach that takes the principle of interchange to a higher meta-level by suggesting crosswalks between different interchange and interoperability formats that were not necessarily designed with the same metamodels and the same functional applications in mind. The goal here would be to be able to retrieve and utilize some, but not necessarily all, of the information stored in diverse resources. Potential implementations might include:

NLP researchers developing yet another new toy, which is interesting on an intellectual level, but which has little practical application in pragmatic organization environments – This "application" reflects a criticism that "no one" in industry is doing this now, so it is not really essential to think about doing it in the future.

Publically available resources, such as metadata registries and standardized terminologies (e.g., a planned ISO concepts directory) contain parsably accessible terms, labels, and rigorous definitions that can be harvested for use in other similar resources and as semantic anchors for values used in ontological resources of various kinds. This is a significant factor because one of the issues involved in ontology design is that ontology developers frequently are not themselves subject area specialists, coupled with the fact that the latter are not necessarily knowledge engineers. Especially as the creators of standardized resources expand their skills in developing non-ambiguous knowledge resources, the presence of authoritative, persistent, accessible semantic information can contribute to the Trust Layer in the framework of the Semantic Web.

Private enterprises, particularly those with large distributed infrastructures, are relying on formal ISO-level standards as well as markup languages, exchange formats, and other consensus-based approaches that are being developed by business-to-business consortia such as OMG, OASIS, LISA, etc. These organizations might use this approach to access and exploit the information that resides within in-house corpora and knowledge bases for internal use. Despite concerns for protecting proprietary information, some companies (Microsoft, IBM) maintain extensive public information resources and have or are launching ambitious "community computing" services that are base on or will be building linguistic resources.

Governmental and other public service entities frequently possess at least theoretically extensive resources that contain terminological and semantic content, but that have been developed in widely diverse environments over time using approaches and formats that may or may not be mutually compatible. One approach to this is to harvest so-called "snapshots" of resources, convert such snapshots into a common annotated format, and make them available as combined content accessed through a common interface. The concern here is that many of these resources may continue to evolve in their native environments, with the result that the combined resource runs the risk of being obsoleted if there is no on-going, potentially costly, maintenance strategy. Procedures for establishing interoperability in real time for dynamically evolving resources (coupled with the implementation of viable persistent identifiers) might make it possible for real-time interaction of aggregated knowledge representation resources, even if they have been originally configured for use by different communities of practice.

5. Theoretical approaches and task definitions

Obviously, this paper represents only a statement of the problem at hand and an initial sketch of potential mapping paths between the systems involved. A number of individual items require further clarification in dialogue between the communities of practice, and some are subject to on-going decision making on both sides of the question. One future task involves the further refinement of the mapping tables, but it is nonetheless important to note that the premise upon which the mapping exercise rests is the assumption that semantic mapping can or should take place between resources developed in environments that are playing by significantly different rules. Any decisions along these lines, as noted above, must be accompanied by the generation of RDF notation for TDB data. The question arises whether existing methodologies, at least viewed from the perspective of discourse-oriented terminology management, make sense in an environment where tools exist for creating terminological concept systems external to, but linked to, concept-oriented TDBs.

6. References & Abbreviations

ANSI/NISO – American National Standards Institute/ National Information Standards Organization
___. Z39.19-2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies. http://www.mt-archive.info/CLT-2003-Obrien.pdf.
Budin, G. and Melby, A. (2000) Accessibility of Multilingual Terminological Resources – Current Problems and Prospects for the Future. http://www.ttt.org/salt/.
DCR – Data Category Registry.
DCS – Data Category Selection.
DIS – Draft International Standard.
DITA – Darwin Information Typing Architecture.
Dublin Core Metadata Element Set, Version 1.1 2008 .http://dublincore.org/documents/dces/.
ISO – International Organization for Standardization
ISO standards: Geneva, International Organization for Standardization:
ISO 639-1:2002. Codes for the representation of names of languages – Part 1: Alpha-2 Code.
ISO 639-2:1998. Code for the representation of languages – Part 2: Alpha-3 Code.
ISO 639-3:2008. Codes for the representation of languages – Part 3: Alpha-3 Code for comprehensive coverage of languages. http://www.sil.org/iso639-3/default.asp
ISO 704:2000. Terminology work – Principles and methods.
ISO 1087-1:2000. Terminology work – Vocabulary – Part 1: Theory and applications.
ISO 12620:1999. Computer applications in Terminology — Data categories
ISO 12620 DIS 2007. Terminology and other language and content resources – Data Categories – Specifica-

tion of data categories and management of a Data Category Registry for language resources

ISO 16642:2003. Computer applications in terminology – TMF (Terminological Markup Framework)

ISO DIS 32042:2008. TermBase eXchange (TBX) Format Specification.

ISO WD 24618. Citation of Electronic Resources.

ISO/IEC 11179 :2007. Information Technology – Metadata registries (MDR) : ISO/IEC JTC1 SC32 WG2 Development/Maintenance, http://metadata-standards.org/11179/.

ISOcat. (2008). Data Category Registry: Defining widely accepted linguistic concepts. http://www.isocat.org/.

KRR – Knowledge Representation Resource

LISA – Localisation Industry Strandards Association.

LMF – Lexical Markup Framework.

Microsoft. (2008). Microsoft Terminology Live Community. Windows Live (German). (Periodic accessibility) https://members.microsoft.com/wincg/de-de/mtcf_home.aspx?s=1&langid=1210.

MDR – Metadata Registry.

Mikhalenko, P. 2005 Introducing SKOS. O'Reilly xml.com. http://www.xml.com/pub/a/2005/06/22/skos.html

Nuopponen, A. (2005). Concept Relations v2. An update of a concept relation classification. In *Terminology and Content Development. Proceedings of the 7th International Conference on Terminology and Knowledge Engineering*, B. Nistrop-Madsen and Thomsen, H.E. (eds.), Copenhagen, Litera, pp. 128-138.

OASIS – Organization for the Advancement of Structured Information Standards.

O'Brien, S. (2003). Controlling Controlled English: An Analysis of Several Controlled English Rule Sets. http://www.eamt.org/archive/dublin/OBRIEN.PDF.

OWL Web Ontology Language Oveview. http://www.w3.org/TR/owl-features/.

OMG – Object Management Group.

RDF – Resource Description Framework.

SALT: (2001). Standards-based Access service to multilingual Lexicons and Terminologies http://www.ttt.org/salt/

SBVR – Semantics of Business Vocabulary and Rules.

SKOS Simple Knowledge Organization System Primer: W3C Working Draft 21 February 2008. Antoine Isaac, A. and Summers, E., Eds. http://www.w3.org/TR/2008/WD-skos-primer-20080221/

SKOS Simple Knowledge Organization System Reference: W3C Working Draft 25 January 2008. Miles, A. and Bechhofer, S., Eds. http://www.w3.org/TR/skos-reference/

SKOS Core Vocabulary Specification 2005. Miles, A. and Brickly, D., Eds. http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/.

Svenonius, E. (2000). The Intellectual Foundation of Information Organization: Digital Libraries and Electronic Publishing. Cambridge, Mass. MIT Press.

TBX – Termbase eXchange, *see* ISO 32042.

TDB – Terminology Database

SLT – Subject Language Terminology

TKE – Terminology and Knowledge Engineering

Tudhope, D., Koch, T., Heery, R. (2006). Terminology Services and Technology: JISC state of the art review. http://www.ukoln.ac.uk/terminology/TSreview-jisc-final-Sept.html

URI – Uniform Resource Identifier.

W3C – Worldwide Web Consortium.

WD – Working Draft.

Windhouwer, M.; Kemp-Snijders, M.; Wittenberg, P., and Wright, S.E. (2008). ISOcat: Coralling Data Categories in the Wild. Poster article for LREC 2008.

Wittenberg, 2007. ISO position paper on persistent identifiers.

WordNet: A lexical database for the English language. http://wordnet.princeton.edu/. Accessed 2008-03-02.

Wright. S.E. (2007). Coping with Indeterminacy: Knowledge Organization Systems in Digital Environments. In. Indeterminacy in Terminology and LSP: Studies in honour of Heribert Picht. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 137-179.

XLIFF – XML Localization Interchange File Format.

Table 1 : SKOS-TBX Cross-mapping

SKOS classes and properties	12620 data category → SKOS	12620 data category ≠ Not in SKOS
Labels and Terms		
label	N/A	term (A.1)[1]
prefLabel	term (A.1), status = preferred term	
altLabel	variant, abbreviation, full form, etc.	term, status = admitted term (A.2.9.1)
hiddenLabel	term termType = deprecated term (A.2.9.1)	term, status = deprecated term (not recommended) (A.2.9.1)
	Misspelling	(not included in TBX)
symbol	term termType = symbol (A.2.1)	
altSymbol	Ditto above, status = admitted term (A.2.9.1)	
prefSymbol	Ditto above, status = preferred term (A.2.9.1)	
Definitions		
definition [2]	definition (A.5.1)	(See discussion above)
Notes		
note	note (A.8)	
changeNote	note inside transacGrp, type="modification"	
editorialNote	note (A.8) within terminology management transactions (A.10.1)	
example	example (A.5.4)	
historyNote	history note	(not included in TBX)
scopeNote	explanation (A.5.2)	(See discussion above)
Concept System/Scheme		
ConceptScheme	concept system (A.7.1)	
Concept	entry identifier (A.10.15)	
inScheme	[pointer to A.7.1]	(suggested for TBX)
Subject references		subject field (A.4.5)
subject		(not broken down in TBX; see discussion above)
primarySubject		
subjectIndicator		
isSubjectOf		
isPrimarySubjectOf		
Relations		
semanticRelation	N/A	Concept relations
hasTopConcept	Potentially: broader concept generic (A.7.2.1)	
broader	superordinate concept generic (A.7.2.2)	
narrower	subordinate concept generic (A.7.2.3)	
CollectableProperty	N/A	[Embedded as genus element in rigorous definitions]
Collection		[Any superordinate concept that could become the subject of a collection]

[1] Non-mnemonic TBX data category ID number. There is no way in SKOS to deal with multiple preferred terms classified by other categories, although this could be finessed by treating these categories as scheme identifiers, such as inScheme=clientSet1, businessUnit1, etc.
[2] In SKOS, definitions appear under note, whereas they are fundamental elements in termbases. See also *scopeNote.*

SKOS classes and properties	12620 data category → SKOS	12620 data category ≠ Not in SKOS
member	coordinate concept (A.7.2.4)	[Any subordinate concept that could become a member of a collection]
OrderedCollection	Ordered thesaurus (A.9.6) ??	
memberList	ordered coordinate concept (A.7.2.4.1)	
related	related concept (A.7.2.5)	
		broader concept partitive
		superordinate concept partitive
		subordinate concept partitive
		coordinate concept partitive
		temporally related concept
		spatially related concept
		associated concept
		(Nuopponen categories)
		antonym (is disjoint with) (10.18.6.1)