A Quantitative Analysis of eCl@ss, UNSPSC, eOTD, and RNTD: Content, Coverage, and Maintenance

Martin Hepp[1,2], Joerg Leukel[3], Volker Schmitz[3]
[1]Digital Enterprise Research Institute (DERI), Innsbruck, Austria
[2]Florida Gulf Coast University, Fort Myers, FL, USA
[3]University of Duisburg-Essen, Essen, Germany
mhepp@computer.org, {joerg.leukel, volker.schmitz}@uni-essen.de

Abstract

A significant part of data and content management in e-business scenarios deals with exchanging product-related data between business entities, and integrating them into target applications (e.g. ERP systems) or target documents (e.g. e-catalogs) at the recipient's side. Content integration tasks can be much better automated if the textual descriptions are augmented by a machine-readable representation of the semantics. For this purpose, categorization standards for products and services, like UNSPSC, eCl@ss, eOTD, or the Rosettanet Technical Dictionary (RNTD) are widely in use. Existing research, however, has focused on the architecture and structure of such standards, and did not investigate their actual content. In this paper, we present a framework of metrics for the quality and maturity of categorization standards, and apply these metrics to eCl@ss, UNSPSC, eOTD, and RNTD. The results clearly show weaknesses which hamper the use in many application domains. Also, we can reveal that only some of these standards are actually maintained and updated, while others are rather inactive, dead collections.

1. Introduction

Data and content management in an e-business environment consists to a significant extent of content integration tasks, where content integration is, following the definition by *Stonebraker* and *Hellerstein*, the "integration of operational information across enterprises", which is highly volatile, and large in data volume and number of transactions [1]. Two very common examples are the integration of product descriptions from multiple suppliers into one consistent, multi-vendor catalog, and the aggregation of itemized invoicing data into a financial target hierarchy for analytical purposes like spend analysis. The mere number of such tasks on one hand and the limited amount of time available on the other hand make a high degree of mechanization of any such tasks highly desirable. As mechanized integration solely based on natural language analysis of unstructured data has so far not achieved a sufficient level of precision, the common approach is tagging individual data sets with references to entries in a standardized vocabulary of products and services terminology, such as UNSPSC. These vocabularies are usually centered around a hierarchy of goods categories, e.g. "office supplies" with "pencils" and "rulers" as subclasses. Within this paper, we refer to such standardized vocabularies for products and services terminology as *Products and Services Categorization Standards (PSCS)*. For some years now, multiple standards bodies have been developing and providing such standards, and businesses have tried to make use of them for the mechanization of product-related data processing. However, the current situation is unsatisfying for the following reasons:

(1) The initial enrichment of unstructured data with such machine-readable semantics like UNSPSC codes is a labor-intensive task, which should be done only once. Since automated mapping between multiple such standards is extremely difficult due to differences in granularity and focus, companies have the problem of selecting the most suitable standard and cannot easily correct this decision at a later point in time.

(2) While the structure and properties of the standards are known in advance and can be used for the comparison of alternatives, the actual coverage and level of detail provided in a given category of products is not obvious. This leads to a situation where the decision for a standard is based mainly on its skeleton (e.g. whether it in general provides properties for a more detailed description of a category) and not on the degree to which such properties are actually defined for the product range of interest.

(3) Products and services categories undergo continuous change due to innovation. This creates the need for new categories or additional properties for existing categories. Without maintenance activities, any standard outdates quickly and its coverage of representational needs decreases. It is thus crucial to

know whether a given standard is being actively maintained and supported by a user community.

(4) The actual content quality of a categorization standard cannot be derived from very obvious figures, like the total number of classes or properties for products. This is because such numbers are positively affected by activities like the bulk import of very specific, but not widely used categories from other standards (e.g. military sourcing categories), or by redundancy in the set of supported product properties.

In short, e-business participants have a strong need for measuring the actual content quality of products and services categorization standards, because they must select the most suitable standard prior to investing in the annotation of unstructured data, but have currently no metrics at hand that can be used for this purpose.

In this paper, we describe a comprehensive set of quantitative metrics that allow evaluating the maturity, specificity, and coverage of products and services categorization standards, and apply them to the most current and multiple past releases of the three most prominent horizontal (i.e. cross-industry) standards UNSPSC, eCl@ss, and the ECCMA Open Technical Dictionary (eOTD), and one vertical (i.e. industry-specific) standard, namely the Rosettanet Technical Dictionary (RNTD).

1.1 Categorization Standards

There are countless approaches for the categorization of goods, ranging from rather coarse taxonomies, created for customs purposes and statistics of economic activities, like the North American Industry Classification System (NAICS) and its predecessor SIC (see [17]), to expressive descriptive languages for products and services, like eCl@ss, eOTD, or RNTD. The UNSPSC, widely cited as an example of a product ontology, is in the middle between those two extremes, providing an industry-neutral taxonomy of products and services categories, but no standardized properties for the detailed description of products. It is out of the scope of this paper to list and compare all available standards in this area, but one can say that UNSPSC, eCl@ss, and eOTD are currently the most important horizontal standards (i.e. covering a broad range of industries), and RNTD should be included in the analysis because of its high degree of detail, albeit limited to a narrow segment of products (electronic and IT components).

All of those standards reflect a varying combination of the following components:

Product Classes: All PSCS are based on a set of product categories that aim at grouping similar products. This grouping is often influenced by the purpose of the PSCS. For example, the categories can try to collect products by the *nature* of the products or by their intended *usage*. This can create confusion, as there is an N:M mapping between the nature of a product and product usages. The meanings of the product classes are usually captured in a rather informal way, ranging from just very short class names to quite precise natural language definitions, sometimes available in multiple languages.

Hierarchy of Classes: Most PSCS arrange the classes in hierarchical order. It is crucial to understand that this hierarchy is directly connected to the intended usage of the PSCS. For example, eCl@ss was designed with the idea of grouping products from the perspective of a *buying organization* or a *purchasing manager*.

Dictionary of Properties: More sophisticated PSCS include a dictionary of standardized properties that can be used to describe product instances or product models in more detail and allow parametric search. Usually, those property dictionaries contain a quite rich definition of the contained properties, including not only sophisticated data typing, but also references to international standards for the units of measurement.

Enumerated Property Values: For properties where an arbitrary string is not sufficient to capture the value in a semantically unambiguous way, some PSCS maintain a list of supported values in a separate collection. The mapping between recommended values and such properties is usually kept in a separate relation.

Class-Property Relation: Most PSCS with a dictionary of properties include a mapping between classes and recommended properties, i.e. property sets per each class, sometimes referred to as "attribute lists" or "class-specific property sets". The semantics of this assignment varies between different standards. It can range from very loose recommendations (as in eOTD) to a strict definition of those properties necessary and sufficient to completely describe an instance of the respective class.

Keywords: Sets of keywords and relations between such words and categories or properties support manual searches for proper entries.

Due to the continuous innovation in the product and services domain, all PSCS are a work in progress with multiple releases per year.

1.2 Related Work

Collections of consensual concepts for the communication about products and services have been subject to much research in diverse research communities, e.g. under the label "ontologies" in the knowledge representation and data management field [2], with specific focus on catalog-data integration [3-6], and as "product classification standards" (PCS) [7, 8] or "product schema" [9] in the e-commerce arena. Also, "descriptive languages for products and services" has been proposed as an alternative term [10]. Many researchers have worked on the task of integrating two standards by finding similar concepts and establishing mappings between them, e.g. [8] or [11].

Very surprising is that the vast majority of previous work takes the existence of such categorization standards for granted and treats the most prominent approaches eCl@ss, eOTD, UNSPSC, or RNTD as an externally given solution to the non-trivial requirement of sufficient coverage and detail. Except for our earlier works as presented in [10, 12, 15], we do not know of any in-depth analysis of the content quality of PSCS. The empirical study by Fairchild and de Vuyst [13] analyzes the concepts of standardized PSCS, but describes only some characteristics of UNSPSC on a high level of abstraction. Similar work to ours can be found in the ontology community in [14]; they propose metrics for the structural properties of RDF-S schemas for the Semantic Web, but include only one product-related schema in their analysis of 28 schemas.

Also, there is a natural proximity to the discipline of software metrics (see e.g. [18]), but the common approaches have to our knowledge not been applied to business vocabularies as some form of software.

1.3 Our Contribution

In this paper, we (1) present a framework of metrics that can be used to assess the quality and maturity of products and services categorization standards, (2) apply these metrics to the most current and multiple past releases of eCl@ss, UNSPSC, and RNTD. Based on this, we (3) reveal that most of those standards, though advertised as industry-neutral undertakings, are fully developed in only a few selected branches. Also, we can (4) clearly show which of these standards are actually maintained and updated, and which others are rather inactive, dead collections.

The structure of this paper is as follows. In section 2, we define a set of metrics that reflect the dimensions of content, domain coverage, and maintenance.

Section 3 describes our experiences and the resulting data of applying our metrics to multiple releases of eCl@ss, UNSPSC, eOTD, and RNTD. In section 4 we discuss the findings and implications for both e-business participants and standards bodies, and summarize the work.

2. Methodology

In this section, we define a set of metrics that aim at reflecting dimensions relevant for judging the content quality, domain coverage, and amount of maintenance of products and services categorization standards. The general approach is to determine the amount of structural elements and relationships between such elements. We want to answer the following questions:

(1) To which degree do the elements supported by the skeleton of any given PSCS actually exist (e.g. are there class-specific property assignments for all existing classes)?

(2) If the respective elements exist, to which degree are they specific?

(3) Is the degree of completeness and detail consistent throughout all top-level categories (often called segments), or are there rather "islands" of mature content in an otherwise very incomplete skeleton?

(4) How much maintenance work and updating is actually done by the standards body between release intervals?

2.1 Relevant Dimensions

The proposed metrics target four aspects of the respective PSCS:

(1) The size, growth, and maintenance volume,

(2) the degree of balance among segments, hierarchical order, and the breadth of coverage,

(3) the size and expressiveness of the property library, and

(4) the specificity of property assignment in class-wise property lists.

Size, growth, and maintenance: The metrics in this section reflect the size and pace of growth of a given PSCS by comparing multiple releases of the same PSCS with regard to the number of products and services classes, and relating the amount of new or modified elements to the amount of time passed between two release dates. Those metrics show the amount of common concepts in the standard, i.e. those that reflect some degree of domain consensus. The

metrics do not take into account the coverage of concepts of a specific application domain. Measuring the growth and the maintenance work for a given PSCS per period of time indicates the amount of feedback received from the application domain and the "bandwidth" and delay of the standardization process, whichever is the limiting factor. This is an indicator for the seriousness of maintenance, and it is also of importance for users of the standards in order to implement a suitable versioning scheme.

Metrics for hierarchical order and balanced content: Most PSCS include a hierarchy of all products and services classes. This can be used to partition the total number of classes into the respective top-level sections (segments) and draw conclusions about the distribution along the hierarchy. We can also use this approach for the analysis of how the distribution of classes develops over time, in order to see whether a given PSCS is getting more balanced or whether the degree of imbalance increases, and in which areas the content is actually being improved. The resulting data is interesting, because it

(1) reveals the degree of balance among the different categories and

(2) shows the most populated categories and thus the true domain focus of a given PSCS.

Obviously, those metrics cannot be applied to standards that do not contain at least some form of hierarchical order.

Property library: Many PSCS include a library of predefined properties that can be used to describe a product or service model in more detail. The metrics in this section reflect the total size of this property library and its development over time, and the amount of properties with enumerative data typing. The latter is a good indicator for the specificity of such properties that cannot be properly represented using numerical data types, since the more generic alternative of using unrestricted string values hampers automated processing.

Quality of class-specific property sets: Many PSCS contain a property-class relation that assigns necessary or recommended properties from the property library to individual products and services classes. This tells a standards user the suitable properties for the description of an item of the respective products or services category. Unfortunately, the quality and specificity of those property-class assignments varies significantly. On one hand, there is usually a small set of very generic properties assigned to any (or almost any) class. Property lists containing just such generic properties (e.g. commercial properties such as EAN, product ID, or manufacturer name) add little to the description of a category. On the other hand, it happens that property lists hold a huge number of arbitrarily selected and often redundant properties.

A first approach to measure the quality of and progress in class-property assignment is to count the number of class-specific property lists. In the context of this paper, a property is considered a generic property when it is contained in more than 75 % of the property lists, and a property list is considered specific as soon as it holds one single specific (i.e. not generic) property. The cut-off point of 75 % was selected because it is compatible with the implicit design decisions of all common PSCS and also does not count inconsistencies in the assignment of generic properties to the favor of the respective standards, which would have been the case if a cut-off point near 100 % was chosen.

2.2 Metrics

Table 1 summarizes the metrics and their definition. For a more elaborated discussion of the rationales underlying the metrics, see [15].

The two metrics Semantic Weight and Semantic Value have been developed in our earlier works [10, 15] and are motivated and defined as following. The motivation is to take into account the degree of specificity of the property lists, based on the fundamental idea that a property being used very frequently is generally less specific than a property assigned to only a few categories.

In the simpler metric "Specific Property Ratio", a property list is either considered specific, as soon as it contains a single property that is used in less than 75 % of all property lists, or generic, if it contains only properties assigned to at least 75 % of the classes or no properties.

The extended approach Semantic Weight/Semantic Value consists of two steps: First, the *Semantic Weight* for each property in the property library is determined. In a second step, the *Semantic Value* for each single property list is computed by adding the Semantic Weights of all properties contained. The Semantic Value for classes without a property list is by definition equal to zero.

Semantic Weight of properties: For each property

P_i with $i = 1, \dots ,$ *Number of Properties*

in the property library, we count the number of entries in the class-property relation. This yields the number of occurrences of property P_i. Then, each property P_i in the property library receives a Semantic Weight SW_i that is equal to the reciprocal value of its usage frequency in a given release of the PSCS (this idea

resembles basic concepts in information and communication theory):

$$SW_i = \frac{1}{Number\ Of\ Property\ Lists\ Containing\ P_i}$$

It is important to note that this is not a characteristic of the respective property alone, but reflects its usage in a given PSCS. The uneven distribution of classes and the fact that node specific property lists do not yet exist for a huge portion of the classes influence the absolute semantic weights.

A base property will have a semantic weight of

$$\frac{1}{\alpha * Number\ Of\ Property\ Lists}$$

with $1 \geq \alpha \geq 0.75$

The value α reflects the percentage of property lists that actually contain this base property. Its range results from the definition of a non-specific property as above. A very specific property used only in one single property list has a Semantic Weight of 1. Properties in the property library that are not used in any property list should be simply ignored, because no meaningful value can be determined.

Semantic Value of property lists: Now, for each product or service class C_j in the PSCS having a property set S_j, we sum up the Semantic Weights of all contained properties. This yields the Semantic Value SV_j for each Class C_j with j=1, … , *Number of Classes*

$$SV_j = \sum SW_i \mid P_i \in S_j$$

The fundamental rationale is that more properties mean a higher semantic specificity of the property list for the class, but very frequently used properties add less semantics than specific properties. SV_j is an indicator for the semantic specificity of the class C_j. The higher SV_j, the more distinct is the respective property list from that of any other class.

It is important to note that the Semantic Value is not an absolute measurement, because it is influenced by the size and structure of the property library. For

Table 1: Proposed metrics for PSCS content

	Metric	Description	Definition
Size and Growth	Size	Total number of classes; amount of new and modified classes	For each release of a specific PSCS, we count the overall number of products and services classes. For hierarchically organized standards, we include intermediate nodes on all levels of the hierarchy. Then, we determine (1) the number of new and (2) the number of modified classes, i.e. such concepts that existed in the previous release but have now a new version number due to changes in the definition of the concept.
	Speed of Growth	Number of new classes and number of modified classes per month	For each release change of a given PSCS, we determine the amount of (1) new and (2) modified classes (if there is a hierarchical order: on any intermediate level) and divide it by the number of months passed since the two release dates.
Hierarchical Order and Balance of Content	Size of Segments	Number of classes per segment	For each release of a given PSCS, we determine the number of classes per segment, i.e. all descendents plus the top-level class itself. For the most recent version of the respective PSCS, the results can be summarized in a bar chart listing all segments ordered by descending number of classes.
	Services Ratio	Number of services classes divided by the total number of classes	We count the total number of services classes (on all levels) based on the description of the first level of the hierarchy and relate them to the total number of classes (on all levels). This approach does not take into account services that are hidden in a deeper level of the hierarchy, but the later can only be found by manually counting each single entry, which is unfeasible.
	Variability of Segments Size	Distribution parameters for the size of segments	We determine the distribution parameters for the data gained in section "Size of Segments", i.e. the minimal value, maximal value, mean, median, first quartile, third quartile, interquartile range, standard deviation, and the coefficient of variation.
	Segments Domination	Size ratio between the most populated segments in comparison with all other segments	For the current release of a given PSCS and based on the data gained in section "Size of Segments", we determine the percentage of classes contained in (1) the most populated and (2) in the three most populated segments. For the most recent version of a PSCS, we divide the number of classes in the most populated segment by the median of all segments.
	Balance of Branching	Distribution parameters for the number of direct descendents on each level of the hierarchy	For each level of the hierarchy individually, we count the number of direct descendents per superordinate class, and determine the minimal value, maximal value, mean, median, standard deviation, and the coefficient of variation for the resulting data.
Property Library	Size	Total number of properties	For each release of a given PSCS, we count the total number of properties in the property library.
	Enumeration Properties Ratio	Share of properties with enumerative data typing	We count all properties in the property library that are assigned at least one enumerative data value and relate the number of those properties to the total amount of properties.
Quality of Class-specific Property Sets	Specific Property Lists Ratio	Share of classes with specific properties assigned	We count all classes that contain at least one specific property in their property list. Even if a given PSCS assigns properties only at the leaf level (i.e. no properties are assigned to intermediate nodes in the hierarchy), we compute, for reasons of comparability, the percentage based on the total number of classes.
	Property Usage Statistics	Distribution parameters for the number of specific properties assigned per class	For each class that has a specific property list, we count the number of properties in this list and determine the minimal value, maximal value, mean, median, standard deviation, and coefficient of variation.
	Semantic Weight	Quantified significance of each property	See text
	Semantic Value	Quantified specifity of property assignment per class	See text

example, a badly structured property library with duplicate entries for identical properties will increase the Semantic Values. The major gain is not the value itself, but its *distribution properties* with regard to the PSCS as a whole.

As an attempt to take into account the size of the property library and penalize overly big property collections with lots of redundant entries, the raw value SV_j should be divided by the number of properties.

3. Application to Selected Standards

In a comprehensive analysis, we determined the metrics defined above for all available releases of the four standards eCl@ss, eOTD, RNTD, and UNSPSC. This section describes our experiences and presents the resulting data. All in all we gained an enormous set of interesting observations, of which we can present only a selection in this paper.

3.1 Size and Growth

These metrics reflect the vocabulary size, i.e. the number of generic products and services concepts in the respective PSCS, and how this changes over time. It also shows the degree of change dynamics between any two subsequent releases, which is important for standards users, as it helps determine a suitable strategy to cope with release changes. Modified classes often require manual checking whether the existing class assignments are still valid.

Table 2. Total number of classes in various PSCS

	Version	Total number of classes
eCl@ss	4.1	15315
	5.0	24814
	5.0SP1	24919
	5.1beta	25585
	5.1de	25658
eOTD	01-17-2003	58973
	10-01-2003	58898
	11-01-2003	58901
	03-01-2004	58975
	06-01-2004	58970
	08-01-2004	58970
RNTD	2.0	744
	3.0	943
	3.1	770
	3.2	770
	4.0	789
UNSPSC	6,0315	19778
	6,0501	20212
	6,0801	20498
	6,1101	20683
	7,0401	20739
	7,0901	20789

Table 2 shows the absolute size of the four PSCS, which is the most easily available metric and one frequently used by standards bodies for promoting their standards. eOTD has almost three times as many classes as UNSPSC and more than twice as many as eCl@ss in its most recent release. This basic metric, however, does not reveal the actual amount of maintenance and can also be very much biased by a large amount of categories in very specific areas. Many PSCS were created by merging existing standards from specific domains (eCl@ss: sourcing needs of the chemical industry; eOTD: NATO procurement).

Table 3. New and modified classes per month

	Release	Previous release	New classes per 30 days	Mean	Modified classes per 30 days	Mean
eCl@ss	5.0	4.1	865.0		157.4	
	5.0SP1	5.0	47.8	279.6	10.2	1271.4
	5.1beta	5.0SP1	131.6		4918.0	
	5.1de	5.1beta	74.1		0.0	
eOTD	10-01-2003	01-17-2003	6.1		0.0	
	11-01-2003	10-01-2003	4.8		0.0	
	03-01-2004	11-01-2003	18.3	6.2	0.0	0.0
	06-01-2004	03-01-2004	1.6		0.0	
	08-01-2004	06-01-2004	0.0		0.0	
RNTD	2.0	1.4	0.7		6.4	
	3.0	2.0	2.4		1.0	
	3.1	3.0	0.0	1.3	0.1	1.5
	3.2	3.1	0.0		0.0	
	4.0	3.2	3.4		0.0	
UNSPSC	6,0315	5,1001	907.8		135.6	
	6,0501	6,0315	304.5		53.0	
	6,0801	6,0501	97.5	233.9	15.0	47.5
	6,1101	6,0801	69.1		50.2	
	7,0401	6,1101	13.8		29.4	
	7,0901	7,0401	10.8		2.0	

Table 3 shows the amount of new and modified classes for the most recent releases. We can observe that eOTD as the largest set of concepts (58,970 classes in the most recent release) and RNTD as a comparatively small set (789 classes) have both almost no growth with regard to their content. In contrast, eCl@ss has been, on average, growing by as much as 280 and UNSPSC by about 230 new classes per 30 days, and both also show significant maintenance of existing entries.

For a good coverage of concepts needed in the domain, any PSCS requires timely and complete feedback about missing entries from the user community, and a streamlined standardization process that makes respective new elements available in a timely manner.

3.2 Hierarchy and Balance of Content

These metrics show how the distribution of classes along the segments developed over time, in order to see whether a given PSCS is getting more balanced or whether the degree of imbalance increases. Also, since the coefficient of variation can be used to compare distributions with a different mean, it is a good indicator for the comparison of multiple PSCS.

For horizontal products and services standards, this reveals whether the standard is a true horizontal approach or horizontal just with regard to the existence of segments, but focused quite vertically at the more detailed level. A true horizontal standard requires not only the existence of segments for a broad range of concepts but also actual entries in the deeper branches of all segments.

Table 4 compares the amount of concepts contained in the largest and the three largest top-level sections. Especially interesting is the order of magnitude between the largest section and the median, shown in the rightmost column. This metric reveals the order of magnitude of the number of concepts in the most populated segment as compared to the median. The bigger this ratio, the more is the content of the standard dominated by one single segment.

Table 4. Most populated top-level sections

	Release	% of classes in largest category	% of classes in 3 largest categories	Largest category / median of the category size
eCl@ss	4.1	23%	44%	814%
	5.0	21%	40%	731%
	5.0SP1	21%	40%	731%
	5.1beta	21%	39%	732%
	5.1de	21%	39%	732%
eOTD	10-01-2003	24%	40%	5255%
	11-01-2003	24%	40%	5254%
	03-01-2004	24%	40%	5255%
	06-01-2004	24%	40%	5255%
	08-01-2004	24%	40%	5255%
UNSPSC	6,0315	12%	30%	1128%
	6,0501	12%	29%	1134%
	6,0801	12%	30%	1134%
	6.1101	12%	30%	1108%
	7,0401	12%	30%	1107%
	7,0901	12%	30%	1107%

One can clearly see that in all three horizontal standards, the biggest share of categories stems from a very few branches. Especially when compared to the median size of all categories, the degree of imbalance is obvious. The largest segment in eOTD is more than 50 times as big as the median, which can be traced back to the bulk import of classes from past standards. Figure 1 illustrates the dominance of a few categories in UNSPSC 7,0901. Table 5 shows the amount and percentage of services categories in the three standards having a hierarchical order. As per the definition of the metric in section 2, this does not include services hidden in the deeper levels of the hierarchy.

Table 5. Services categories

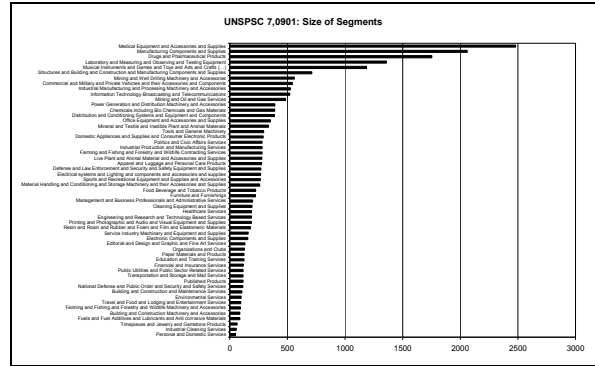|	Release	% of services concepts	# of services concepts
eCl@ss	5.1de	4%	1064
eOTD	08-01-2004	10%	5906
UNSPSC	7,0901	21%	4313

Figure 1. Uneven population of segments in UNSPSC V. 7,0901

The services domain differs from the representation of tangible products, e.g. because the fulfillment is bound to properties of the service customer, especially with regard to location and time. Also, there might be industries where, due to their high volume, services are of special interest for spend analysis. It thus makes sense to determine the percentage of services classes. Tables 11 and 12 summarize the distribution of elements along the hierarchy. This metric reveals how the number of sub-concepts varies among the levels of the hierarchy. One can see that the coefficient of variation of the number of classes per top-level category for eOTD is about twice the value for both UNSPSC and eCl@ss, pointing to very diverse top-level sizes. The smallest ("tractors") contains just seven descendents, the biggest ("medical, dental, veterinary,…") two-thousand times as much (14,189).

3.3 Property Library

The size of the property library reflects the amount of concepts for properties in the given standard. However, it can be suspected that redundancy is a big problem with regard to properties, because the often distributed development of PSCS makes it very likely that redundant properties are created when the existence of an equivalent property is not realized due to different terminological conventions. In its current stage, this is a rather raw metric, as it does not indicate the amount of consolidation work (e.g. the deletion of redundant properties). If the number of properties has decreased whereas the number of classes has increased, we can assume that some consolidation has taken place.

It is highly desirable to have properly defined lexical spaces for all properties and thus enumerative data types for such properties that cannot be unambiguously represented using standard data types.

However, we can often observe that such property definitions are incomplete (e.g. defined as any alphanumeric sequence of less than 30 characters). This impedes automatic interpretation of property values.

Table 6 shows the size of the property libraries in eCl@ss, eOTD, and RNTD (UNSPSC does not include properties) and the amount of properties with enumerative data typing. One can see that eOTD has the biggest amount of properties but the lowest absolute number of enumerative data types.

Table 6. Enumerative data typing

	Release	# of properties (including unused)	# of properties with enumerative data type	% of properties with enumerative data type
eCl@ss	5.1de	5525	1064	19%
eOTD	08-01-2004	21129 (28025)	555	3%
RNTD	4.0	3623	714	20%

3.4 Quality of Class-specific Property Sets

Property lists tell a standards user which properties should be used to describe a product model in detail. These recommendations are part of many PSCS and should contain all necessary properties, but not a wild collection of any usable property, because this makes automated processing of product data difficult, as elements of the same type might be described using different properties. Creating and maintaining such property sets per each category is a tremendous task, because it requires consensus on a very detailed level. The provision of properties is often regarded as an important discriminator between PSCS, but has so far been just regarded on the structural level, i.e. whether the data model of the PSCS supports properties, and not whether the PSCS actually contains specific property assignments. The metrics in this section reveal the degree to which the various PSCS actually implement property sets. Also, only the amount of specific property assignments indicates the amount of progress in the creation of fully-fledged products and services concepts.

Table 7 shows the amount of specific property lists, i.e. such that do contain at least one property that is not assigned to more than 75 % of all classes. Table 8 illustrates how the amount of properties per class varies between the various PSCS. The median and coefficient of variation are surprisingly consistent. In general, a huge variation in the amount of properties indicates only partial progress in the development of property assignments. As an extension, this metric

could be applied to each segment in order to identify those segments that actually contain a high amount of specific property lists.

Table 7. Amount of specific property lists

	Release	Total number of properties	# of classes with specific property assignment	% of classes with specific property assignment
eCl@ss	5.1de	5525	10930	43%
eOTD	08-01-2004	21120	20456	35%
RNTD	4.0	3623	789	100%

Table 8. Properties per property list

	Release	Number of properties in specific property lists					
		Min	Max	Mean	Median	STD	Coefficient of Variation
eCl@ss	5.1de	1	156	32.3	44	15.2	47%
eOTD	08-01-2004	7	417	50.3	47	19.5	39%
RNTD	4.0	26	284	53.3	47	24.0	45%

Tables 9 and 10 contain the distribution properties of the Semantic Weights and Semantic Values.

Table 9. Semantic Weights

	Release	Semantic Weight					
		Min	Max	Mean	Median	STD	Coefficient of Variation
eCl@ss	5.1de	0.0000503	1	0.57469	0.5000	0.3903	68%
eOTD	1/8/2004	0.0000170	1	0.52430	0.5000	0.3908	75%
RNTD	4.0	0.0012658	1	0.63834	1.0000	0.3846	60%

The median of 0.5 for the Semantic Weights of eCl@ss and eOTD indicate that half of the properties are used in no more than two property sets, and the respective value of 1 for RNTD says that half of the properties are used in only one or none property set. The coefficient of variation in Table 10 reflects how the specificity of property sets varies across the segments of the standard.

Table 10. Semantic Values

	Release	Semantic Value					
		Min	Max	Mean	Median	STD	Coefficient of Variation
eCl@ss	5.1de	1.84E-08	9.10E-05	4.74E-05	6.16E-07	2.48E-04	523%
eOTD	1/8/2004	2.41E-09	2.21E-06	1.70E-05	6.52E-09	7.32E-05	432%
RNTD	4.0	1.45E-05	2.45E-02	1.15E-03	6.41E-04	1.79E-03	155%

4. Discussion and Conclusion

In this section, we compare our results with the underlying research questions as defined in section 2.
Degree of Completeness and Balance of Content: All three horizontal standards contain an impressing number of categories for products and services, but the categories are quite unevenly distributed among the various top-level segments. The labels and number of

top-level categories promise a very broad, industry-neutral scope, which is an unfulfilled claim in the current stage of the standards. Especially the impressive number of categories in eOTD (58,970) obscures that a very significant share of all entries (24%) are in one single branch ("Medical, dental, and veterinary equipment and supplies"). Compared to the mean of all segments, this branch is 50 times as big. UNSPSC and eCl@ss are much more evenly populated, but still have 7 times (eCl@ss) respectively 11 times (UNSPSC) as many entries in their biggest category. All three have more than 30% of all entries in three large sections and thus only a small partition of their 25 (eCl@ss), 55 (UNSPSC), or 79 (eOTD) top-level categories.

When looking at the number of direct descendents per superordinate node, one can see that the degree of completeness decreases in eCl@ss from top to down; the coefficient of variation increases from 61% (Top-Level$\rightarrow$2nd level) to 156 % (3^{rd} level$\rightarrow$4th level), whereas it is much more consistent in UNSPSC (74% compared to 100 %). In both standards, however, the population at the leaf level varies greatly, with a minimum of just one leaf and a maximum of 85 (eCl@ss) or 92 (UNSPSC). This metric cannot be determined for eOTD because it lacks a fully-fledged hierarchy.

Of course, one cannot assume that all branches need the very same amount of entries, but this objection does not justify the order or magnitude found in current PSCS. As a summary, the total number of classes obscure that many of the branches are still very much incomplete, and potential users are advised to check the coverage of entries in their domain prior to adopting a PSCS.

Specificity: The degree of specificity can be evaluated best by looking at class-specific property assignments.

RNTD has specific property lists for all of its classes, as compared to only 43% (eCl@ss) and 35% (eOTD). In other words, more than 2/3 of all eOTD classes and more than half of all eCl@ss classes are currently without specific property lists. On the other hand contain all PSCS many properties that are used with only one or two classes. This can point either to redundancy, to the "arbitrary" creation of property lists on demand, or a combination of both.

When measuring the semantic specificity of property assignments using the Semantic Value, one can see clearly that both eCl@ss and eOTD show an enormous spread. The coefficient of variation is as high as 523% (eCl@ss) and 432% (eOTD). RNTD has a mean about 33 (eCl@ss) to 100 times (eOTD) of the size. Those orders of magnitude are very much compatible with our manual observations. In absolute values, RNTD has the highest median of 6.41E-04 as compared to 6.16E-07 (eCl@ss) and 6.52E-09 (eOTD). In other words, the property lists (in the middle of the population) of RNTD are 1,000 times more specific than the property lists in eCl@ss, and the property assignment in eOTD is hundred times less specific as compared to eCl@ss. This very well reflects our observation that both have huge differences in the quality of the property assignment. The big difference between RNTD and the two others can be traced back to the fact that RNTD is a very narrow, specific PSCS and can thus achieve coherence much easier, but the gap between eCl@ss and eOTD seems for us mainly a matter of performance. It also correlates with the maintenance activities (see below).

Maintenance: Both eCl@ss and UNSPSC undergo continuous improvement with an average of more than 200 new classes per month. On the other hand, eOTD had less than one new class per month in 2004, and this despite its wide coverage. It is hardly possible that

Table 11. Distribution properties of the number of classes per top-level category

	Release	Release date	min	max	median	Q1	Q3	STD	Coefficient of Variation
eCl@ss	5.1de	09-28-2004	203	5312	726.0	432.0	1216.0	1064.6	104%
eOTD	08-01-2004	08-01-2004	7	14189	270.0	148.5	590.5	1764.3	236%
UNSPSC	7,0901	09-01-2004	49	2480	224.0	120.0	370.0	483.3	128%

Table 12. Direct descendents per superordinate node

	Release	Top-Level->2nd Level						2nd Level->3rd Level						3rd Level->4th Level					
		Min	Max	Mean	Median	STD	Coefficient of Variation	Min	Max	Mean	Median	STD	Coefficient of Variation	Min	Max	Mean	Median	STD	Coefficient of Variation
eCl@ss	5.1de	4	45	19.2	18	11.8	61%	1	83	8.5	6	9.0	106%	1	85	5.2	2	8.1	156%
eOTD	08-01-2004	1	8650	357.0	129	999.6	280%	not applicable						not applicable					
UNSPSC	7,0901	1	29	6.4	5	4.8	74%	1	54	5.8	4	6.3	110%	1	92	9.0	6	9.0	100%

there is no need for new classes in 79 segments. RNTD has also received only minimal additions with a mean of 1.3 new classes per month for the last five releases. For us this points to either lack of user feedback, lack of users, insufficient maintenance procedures, or any combination of these. Any of the three causes are very disadvantageous for a business user of the respective standards, for he or she cannot hope for a timely addition of missing categories.

Implications for Standards Users: Our results show that neither the provided structural components and the data model of a given PSCS nor the general scope as indicated by the top-level categories should be the major criteria for the evaluation of a PSCS. Instead, the amount of actual entries in the branches of interest, the specificity of property assignment, the consistency and specificity of the property library, and the seriousness of maintenance activities should be closely regarded. The metrics we presented in this paper can easily be applied to only a segment of interest in order to evaluate the content quality in the branches relevant to the decision maker.

Corporations can for example compare the semantic values and the amount of maintenance work for their industry segments among multiple PSCS. This will prevent investment into such standards that neither cover existing representational needs nor show convincing efforts of improvement.

Implications for Standards Bodies: Our metrics revealed weaknesses and shortcomings in all of the four PSCS. The metrics indicate quite clearly the type of action needed and also point to the weak branches of hierarchical categorization standards. They can also be used to monitor the development of content quality from a management perspective if the maintenance of branches is organized in a distributed manner.

References

[1]   M. Stonebraker and J. M. Hellerstein, "Content Integration for E-Business", Proc. of the ACM SIGMOD 2001, Santa Barbara, California, USA, 2001, pp. 552-560.

[2]   L. Obrst, R. E. Wray, and H. Liu, "Ontological Engineering for B2B E-Commerce", Proc. of the International Conference on Formal Ontology in Information Systems (FOIS'01), Ogunquit, Maine, USA, 2001, pp. 117-126.

[3]   O. Corcho and A. Gómez-Pérez, "Solving Integration Problems of E-commerce Standards and Initiatives through Ontological Mappings", Proc. of the Workshop on E-Business and Intelligent Web at IJCAI-2001, Seattle, USA, 2001, pp. 131-140.

[4]   D. Fensel, Y. Ding, B. Omelayenko, E. Schulten, G. Botquin, M. Brown, and A. Flett, "Product Data Integration in B2B E-Commerce", *IEEE Intelligent Systems,* vol. 16, 2001, pp. 54-59.

[5]   D. Fensel, D. L. McGuinness, E. Schulten, W. K. Ng, E.-P. Lim, and G. Yan, "Ontologies and Electronic Commerce", *IEEE Intelligent Systems,* vol. 16, 2001, pp. 8-14.

[6]   B. Omelayenko, "Ontology Integration Tasks in Business-to-Business E-commerce", Proc. of the 14th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, Budapest, Hungary, 2001, pp. 119-124.

[7]   J. Leukel, V. Schmitz, and F.-D. Dorloff, "A Modeling Approach for Product Classification Systems", Proc. of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02), Aix-en-Provence, France, 2002, pp. 868-874.

[8]   E. Schulten, H. Akkermans, G. Botquin, M. Dörr, N. Guarino, N. Lopes, and N. Sadeh, "The E-Commerce Product Classification Challenge", *IEEE Intelligent Systems,* vol. 16, 2001, pp. 86-89.

[9]   G. Yan, W. K. Ng, and E.-P. Lim, "Product Schema Integration for Electronic Commerce - A Synonym Comparison Approach", *IEEE Transactions on Knowledge and Data Engineering,* vol. 4, 2002, pp. 583-598.

[10]   M. Hepp, "Measuring the Quality of Descriptive Languages for Products and Services", Proc. of the Multikonferenz Wirtschaftsinformatik 2004, Essen, Germany, 2004, pp. 157-168.

[11]   D. Beneventano, F. Guerra, S. Magnani, and M. Vincini, "A Web Service based framework for the semantic mapping amongst product classification", *Journal of Electronic Commerce Research,* vol. 5, 2004, pp. 114-127.

[12]   M. Hepp, *Güterklassifikation als semantisches Standardisierungsproblem*, Deutscher Universitäts-Verlag, 2003.

[13]   A. M. Fairchild and B. de Vuyst, "Coding Standards Benefiting Product and Service Information in E-Commerce", Proc. of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35), Big Island, Hawai, USA, 2002, pp. 258b.

[14]   A. Magkanaraki, S. Alexaki, V. Christophides, and D. Plexousakis, "Benchmarking RDF Schemas for the Semantic Web", Proc. of the 1st International Semantic Web Conference (ISWC2002), Sardinia, Italy, 2002, pp. 132-146.

[15]   M. Hepp, J. Leukel, and V. Schmitz, "Content Metrics for Products and Services Categorization Standards", Proc. of the IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE-05), Hong Kong, 2005, pp. 740-745.

[16]   M. Hepp, "A Methodology for Deriving OWL Ontologies from Products and Services Categorization Standards", Proc. of the 13th European Conference on Information Systems (ECIS2005), Regensburg, Germany, 2005, pp. 1-12.

[17]   U.S. Census Bureau, "North American Industry Classification System (NAICS) ", available at: http://www.census.gov/epcd/www/naics.html
 (retrieved April 5, 2005).

[18]   N. E. Fenton and S.L. Pfleeger, *Software Metrics: A Rigorous Approach*, 2nd edition, International Thomson Press, London, 1996.