Knowl Inf Syst (2006)
DOI 10.1007/s10115-006-0054-2

Knowledge and Information Systems

REGULAR PAPER

Martin Hepp · Joerg Leukel · Volker Schmitz

A quantitative analysis of product categorization standards: content, coverage, and maintenance of eCl@ss, UNSPSC, eOTD, and the RosettaNet Technical Dictionary

Received: 20 October 2005 / Revised: 15 March 2006 / Accepted: 18 August 2006

Abstract Many e-business scenarios require the integration of product-related data into target applications or target documents at the recipient's side. Such tasks can be automated much better if the textual descriptions are augmented by a machine-feasible representation of the product semantics. For this purpose, categorization standards for products and services, like UNSPSC, eCl@ss, the ECCMA Open Technical Dictionary (eOTD), or the RosettaNet Technical Dictionary (RNTD) are available, but they vary in terms of structural properties and content. In this paper, we present metrics for assessing the content quality and maturity of such standards and apply these metrics to eCl@ss, UNSPSC, eOTD, and RNTD. Our analysis shows that (1) the amount of content is very unevenly spread over top-level categories, which contradicts the promise of a broad scope implicitly made by the existence of a large number of top-level categories, and that (2) more expressive structural features exist only for parts of these standards. Additionally, we (3) measure the amount of maintenance in the various top-level categories, which helps identify the activelymaintained subject areas as compared

A preliminary and shorter version of this paper was presented at the IEEE International Conference on e-Business Engineering, ICEBE 2005.

M. Hepp (✉)
Digital Enterprise Research Institute, Semantics in Business Information Systems Group, University of Innsbruck, Technikerstrasse 21a, 6020 Innsbruck, Austria
E-mail: mhepp@computer.org; URI http://www.heppnetz.de
Florida Gulf Coast University, Fort Myers, FL, USA

J. Leukel
University of Hohenheim, Stuttgart, Germany
E-mail: joerg.leukel@uni-hohenheim.de

V. Schmitz
University of Duisburg-Essen, Essen, Germany
E-mail: volker.schmitz@uni-essen.de

to those which ones are rather dead branches. Finally, we show how our approach can be used (4) by enterprises for selecting an appropriate standard, and (5) by standards bodies for monitoring the maintenance of a standard as a whole.

Keywords Products and services classification · Metrics · UNSPSC · eCl@ss · RosettaNet · Ontologies · Electronic commerce · Electronic catalogs

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 [23]. Two very common examples are the integration of product descriptions from multiple suppliers into one consistent, multi-vendor catalog or 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.[1] These vocabularies are usually built around a hierarchy of 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 several 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 not possible due to a lack of formal semantics and differences in granularity and focus, companies face 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 characteristics 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 product) 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

[1] http://www.unspsc.org

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, and if so, whether this takes place in the sections most relevant for the standards user.

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

In short, individual e-business participants and value chains as a whole 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 methods or tools 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 current and multiple past releases of the three most prominent horizontal (i.e. cross-industry) standards UNSPSC, eCl@ss, and eOTD, and one vertical (i.e. industry-specific) standard, namely the RosettaNet Technical Dictionary (RNTD). For those standards that are partitioned in top-level categories spanning well-defined scopes, we also do a sectoral analysis that makes visible the differences between top-level categories with regard to these measures.

1.1 Categorization standards for products and services

There are countless approaches for the classification 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 U.S. Census Bureau [24]), to expressive descriptive languages for products and services, like eCl@ss, eOTD, or the 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.

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 relationship 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.* So it regards products as similar that are (1) bought from the same range of suppliers, (2) needed by the same consumers inside the organization, (3) billed using the same cost accounting categories, (4) being subject of the same type of procurement process, or any combination of those. A typical consequence of this perspective is that related classes (e.g. service or maintenance) are often subclasses of the general good category. The category "TV Set Maintenance" will thus be regularly a subclass of "TV Sets", and "Oil for Sewing Machines" will be a subclass of "Sewing Machinery".

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

Enumerated Property Values: For properties where a standard datatype 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 (as in eCl@ss).

Keywords: Sets of keywords and relations between such words and classes 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 often 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 [20], with specific focus on catalog data integration [3, 5, 6, 21], and as "product classification standards" (PCS) [17, 22], or "product schema" [25] 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. [22] or [2].

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, UNSPSC, eOTD, or the RNTD as an externally given solution to the non-trivial requirement of sufficient coverage and detail. Except for our earlier works [9, 10, 12, 13], we do not know of any in-depth analysis of the content quality of PSCS. The empirical study by *Fairchild* and *de Vuyst* analyzes the concepts of standardized PSCS, but describes only some characteristics of UNSPSC on a high level of abstraction [4]. Li [18] has analyzed XML grammars for electronic commerce regarding their content, but only on the basis of very simple metrics.

Similar work to ours can be found in the ontology community in [19]; 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, in the course of deriving an OWL ontology from UNSPSC and eCl@ss, there have been findings on the actual semantics of the taxonomic relationships in these standards [11, 15].

Also, there is a natural proximity to the discipline of software metrics, e.g. [7], but the common approaches have to our knowledge not been applied to business vocabularies as some form of software.

In a broader context, our work is related to [8], which is a survey of the state of the art in agent-based electronic commerce, and to [1], which introduces an agent-based electronic market architecture that uses ontologies for representing the various views on products.

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 and taxonomic standards in other domains (e.g. medical informatics or library science) and (2) apply these metrics to the current and multiple past releases of eCl@ss, UNSPSC, eOTD, and the 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. Also, we present a sectoral analysis that breaks down the aggregated results to the respective top-level categories.

The structure of this paper is as follows. In Sect. 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 the RNTD, and demonstrates how our metrics can additionally support decision-making in typical use case scenarios. In Sect. 4, we discuss the findings. In Sect. 5, we summarize the work and highlight its implications for both e-business participants and standards bodies.

2 Methodology and metrics

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 bodies in between release intervals?
(5) Is maintenance taking place in all branches of the standard, or are there just a few actively maintained islands of coverage in an otherwise not maintained collection?

In this section, we identify relevant dimensions of the measuring problem and propose suitable metrics.

2.1 Relevant dimensions

The proposed metrics target four aspects of the respective PSCS:

(1) The absolute amount of content, i.e. the number of categories, properties, and enumerative values for properties,
(2) the degree of balance along the hierarchy (especially the distribution of content over top-level categories) and the content focus of the standards,
(3) the specificity of property assignment in class-wise property lists, and
(4) the absolute growth and the amount of maintenance over time, both for the overall standard and per top-level category.

Amount of Content: The metrics in this section reflect the absolute number of elements in the standards. For pure taxonomies, these are just the categories. For those standards that include a library of predefined properties, which can be used to describe a product or service in more detail, also the size of this property library and the number of properties with enumerative data typing can be determined. The latter is a good indicator for the specificity of such properties that cannot be properly represented using standard datatypes, since the more generic alternative of using unrestricted string values often hampers automated processing.

Metrics for Hierarchical Order and Balance of Scope: 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 (see below), 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.

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 category. This component is often referred to as "property lists", though they are in fact property sets. In this paper, we use both terms as synonyms. 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 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.

Growth and Maintenance: The metrics in this section reflect the 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 number of new or modified elements to the amount of time passed between two release dates. Those metrics show the evolution of the number of common concepts in the standard, i.e. those that reflect some degree of domain consensus. 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. Such analysis is useful both for the overall standard and per top-level hierarchy, and it reveals very interesting details about which industry segments are the most actively maintained ones.

2.2 Proposed metrics

In this section, we define the individual metrics for the relevant dimensions as identified in Sect. 2.1.

2.2.1 Number of classes, properties, and enumerative values

For this dimension, we just count the absolute number of elements. These metrics are very basic and often easily available. They are also often used by standards bodies to promote their work, but it can be shown later that they are useful only in combination with other metrics.

2.2.1.1 Number of products and services classes

Definition of the metric: 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 in our counting.

Rationale: This metric reflects the vocabulary size, i.e. the number of generic products and services concepts in the respective PSCS.

2.2.1.2 Number of properties

Definition of the metric: For each release of a given PSCS, we count the total number of properties in the property library.

Rationale: The size of the property library reflects the number 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.

2.2.1.3 Number of enumerative data types Product properties (e.g. "disk diameter") can either refer to a standard data type (e.g. integer, float, ...), often in combination with a unit of measurement (e.g. "inches"), or to a set of symbols reflecting valid concepts. The second form of data typing is usually referred to as enumerative data types, because the lexical space is an explicit set or list of a limited amount of items.

Definition of the metric: 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 number of properties.

Rationale: 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 datatypes. 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, since it introduces the ambiguity of natural languages into the representation of such values.

2.2.2 Metrics for hierarchical order and balance of scope

All of the standards in our analysis (except for the RNTD) are horizontal standards that aim at covering categories across industries. However, the breadth of top-level categories may falsely indicate an equal coverage and quality in all of those segments. The metrics defined in this section help to reveal whether a given standard is equally well developed across its hierarchy, or whether the breadth of scope found at the top-level is just an unfulfilled promise.

2.2.2.1 Number of classes per top-level category

Definition of the metric: For each release of a given PSCS, we determine the total number of classes per each top-level, i.e. all descendents plus the top-level category itself. The results can be summarized in a bar chart listing all top-level categories ordered by descending number of classes in this category.

Rationale: Many PSCS were created by merging existing standards from specific domains (eCl@ss: sourcing needs of the chemical industry; eOTD: NATO procurement). The mere numbers of categories often used for standards marketing obscure the true coverage in the various sections, because a few highly populated sections, resulting from the bulk import of sometimes very specific concepts, often contribute to a large amount of the total number of concepts.

2.2.2.2 Services vs. products Categorization standards can contain concepts for products, for services, or both. The mere existence of services categories, however, does not reveal the actual number of services categories as compared to products.

Definition of the metric: We count the total number of services concepts (on all levels) based on the description of the first level of the hierarchy and relate them to the total number of concepts (on all levels). This approach does not take into account services that are hidden in a deeper level of the hierarchy, but the latter can only be found by manually counting each single entry, which is unfeasible.

Rationale: 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 (the aggregation of invoicing data for cost accounting reasons). It thus makes sense to analyze the percentage of services categories in the total amount of categories.

2.2.2.3 Distribution properties of the number of classes per top-level category

Definition of the metric: We determine the distribution parameters for the data gained in Sect. 2.2.2.1, i.e. the minimal value, maximal value, mean, median, first quartile, third quartile, interquartile range, standard deviation, and the coefficient of variation.

Rationale: These metrics show how the distribution of classes along the categories 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 that have a different mean, it is a good indicator for the comparison of multiple PSCS.

2.2.2.4 Percentage of content in the most populated top-level categories
In many PSCS, we find a few huge top-level categories, often reflecting bulk contributions from previous industry-specific efforts. The following metric sheds light on this issue.

Definition of the metric: For the current release of a given PSCS and based on the data gained in Sect. 2.2.2.1, we determine the percentage of concepts contained in (1) the most populated and (2) in the three most populated top-level categories.

Rationale: 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 top-level categories, but focused quite vertically at the more detailed level. A true horizontal standard requires not only the *existence of top-level categories* for a broad range of concepts but also *actual content* in the deeper branches.

2.2.2.5 Size of the most populated category vs. median of all top-level categories It is not only worthwhile to know the percentage of concepts in the biggest category, but also the order of magnitude of the imbalance between the biggest category and the median.

Definition of the metric: For the most recent version of a PSCS, we divide the number of elements in the most populated top-level category by the median of all categories.

Rationale: This metric reveals the order of magnitude of the number of concepts in the most populated top-level category as compared to the median. The bigger this ratio, the more is the content of the standard dominated by one single category.

2.2.2.6 Number of descendents per superordinate category Besides the distribution of categories among the top-level sections, it is useful to see how the degree of detail varies among the various levels of the hierarchy.

Definition of the metric: For each level of the hierarchy individually, we count the number of *direct* descendents per superordinate node, and determine the minimal value, maximal value, mean, median, standard deviation, and the coefficient of variation for the resulting data.

Rationale: This metric reveals how the degree of detail (i.e., the number subconcepts) varies among the levels of the hierarchy.

2.2.3 Quality of class-specific property sets

The pure existence of properties as an additional feature of a PSCS is of little use unless the standard includes a mechanism that helps maintain consensus among standards users about which properties are to be used in combination with which category. Otherwise, individual participants will fail to use compatible sets of properties for the same classes, which would make processing the property values very difficult. The common mechanism for this is providing a list of property recommendations per class.

While this structural element looks very attractive from a theoretical point of view, it is very hard to implement and maintain for standards bodies, since it requires industry consensus on a very detailed level. Also, product properties are often the distinguishing aspect of competing offerings; thus, manufacturers, if involved in the standardization process, are very keen on seeing properties included that reflect the advantageous dimensions of their products, while they have little incentive to agree to the addition of a property that only helps their competitors to describe advantageous dimensions of their own products in more detail. Also, the need for maintenance on the level of properties can be assumed to be higher than on the level of categories, since they are more closely coupled to technical innovation (e.g. the category "TV sets" has been needed for many decades now while the property "support for digital TV" is a recent requirement).

The metrics in this section aim at revealing the actual number of specific property sets for products and services classes and at quantifying the degree of specificity. They can also be applied on the level of top-level categories in order to show the distribution of specific property sets over the scope of a given standard.

2.2.3.1 Specific property lists ratio A certain amount of the properties from the property library is usually assigned to either all or the vast majority of all classes. Since these properties add little specificity to the description at the class level, we need a metric that removes such mechanically generated property sets from the total number of property sets.

Definition of the metric: We count all products and services categories that contain at least one specific property in their property sets. 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 concepts. A property is regarded as a specific property if it is included in no more than 75% of all property sets, and a property set is regarded as specific as soon as it contains at least one specific (e.g. non-generic) property. For the justification, please see Sect. 2.1.

Rationale: Only the number of specific property assignments indicates the amount of progress in the creation of fully-fledged products and services concepts.

2.2.3.2 Distribution of specific property sets per top-level categories While the overall percentage of categories with specific property sets is already a good indicator for applications that use the entire scope of the standard, there are many situations where the amount of specific property sets broken down to top-level categories is interesting.

Definition of the metric: We apply the metric 2.2.3.1. individually to each top-level category. The decision whether a property is generic or specific is being made on the basis of the whole standard. If the standard contains a reliable tag for specific property sets, this may be used as an estimate.

Rationale: This metric helps identify those top-level categories that actually contain a high amount of specific property lists.

2.2.3.3 Property usage in property lists If users want to assess the amount of labor for creating product descriptions and data conversion etc., they should know the number of properties per category. The following metric reflects this aspect.

Definition of the metric: For each concept 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.

Rationale: Property lists 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. Additionally, a high variation in the number of properties per each category indicates only partial progress in the creation of property assignments.

2.2.3.4 Semantic Weight and Semantic Value

The two metrics Semantic Weight and Semantic Value have been developed in our earlier works [10, 12, 13] 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 lists ratio" (Sect. 2.2.3.1), a property list is either considered specific, as soon as it contains a single property that is used in not more than 75% of all property lists, or generic, if it contains only properties assigned to at least 75% of the classes or no properties. One should note that none of the popular standards uses inheritance for properties; therefore, we can simply count the explicit property assignments in the standard.

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

$$P_\mathrm{i} \text{ with } i = 1, \ldots, \text{number of properties}$$

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):

$$\mathrm{SW}_i = \frac{1}{\text{Number of property lists containing } P_i}$$

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 big share of the classes influence the absolute semantic weights. A base property will have a semantic weight of

$$\frac{1}{\alpha * \text{number of property lists}}$$

with $1 \geq \alpha \times 0.75$.

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 in Sect. 2.1. 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 $j = 1, \ldots,$ number of classes

$$\mathrm{SV}_j = \sum \mathrm{SW}_i | P_i \in S_j$$

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 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.

While Semantic Weight and Semantic Value have been explicitly developed for measuring the distribution of specificity of property sets in PSCS, they can likely be applied to other domains, e.g. the usage of properties for describing instances in OWL ontologies (since OWL properties have a global scope) or weighing the specificity of references in scholarly work. These additional usages are, however, outside the scope of this paper.

2.2.4 Growth and maintenance

Due to the ongoing innovation in the products and services domain, standards bodies have to keep on creating new categories for new types of goods. The actual amount of new categories is constrained by at least two limitations: The amount of input received from the market side and the speed of processing such input. The amount of innovation dynamics likely varies across industries (e.g. IT components will have more new concepts per time than office supplies), but it can be assumed that some innovation dynamics exists in all domains covered by popular PSCS. This assumption is confirmed by simulation experiments described in [9].

In short, the number of new classes and modified classes is a relevant indicator for the seriousness of maintenance, and it is also of importance for users of the standards in order to implement a suitable strategy for coping with release changes. This is because modified elements in the standard often require manual checking whether the class assignments and properties for existing data are still valid, and many enterprise resource planning (ERP) systems store only one single category per item, which limits the amount of time available for migrating from one PSCS version to the next.

Such analysis may reveal very interesting details about which industry segments are the most actively maintained ones.

2.2.4.1 Number of new classes per month

Definition of the metric: For each release change of a given PSCS, we determine the number of (1) new and (2) modified classes (if there is a hierarchical order: on any intermediate level). Additionally, we divide it by the number of months passed since the two release dates.

Rationale: 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.

2.2.4.2 Number of new and modified classes per top-level category

Definition of the metric: For each top-level category of a given PSCS, we determine the number of (1) new and (2) modified classes (if there is a hierarchical order: on any intermediate level) between two releases. Additionally, we divide it by the number of months passed since the two release dates.

Rationale: While the metric 2.2.4.1. can identify standards that undergo no serious maintenance, it cannot show whether the overall amount of maintenance in a given standard is taking place across all subject areas as defined by the top-level categories, or whether it is limited to a few segments with a very active user community, while the rest of the standard is more of a dead collection.

3 Application to eCl@ss, UNSPSC, eOTD, and the RosettaNet Technical Dictionary

In a comprehensive analysis, we determined the metrics defined above for the most recent releases of the four standards eCl@ss, UNSPSC, eOTD, and RNTD. This section describes our experiences and presents the resulting data.

3.1 Data extraction and applicability

As a first step, we tried to download the most recent release and previous releases of all four standards. While the RNTD is available freely, all others require registration and, in the case of eOTD and UNSPSC, even a membership fee. The membership fee and registration is only necessary if someone wants to download the complete standard. Browser-based search is not subject to these limitations.

The next step was a typical combination of data integration challenges. First, all four standards come in different formats. eCl@ss is delivered as a set of CSV files that are well documented, but required time-consuming manual import into our RDBMS. eOTD comes as a set of MS Access files, UNSPSC as one MS Excel file per release, and RNTD is packed into a proprietary XML syntax. The situation is even more complicated as the individual schemas used for the storage of the standards have often changed over time. We also observed multiple inconsistencies, e.g. that explicit flags are not always set correctly. Whenever possible, we corrected those inconsistencies. The result was one huge set of relations in one RDBMS, which we then analyzed using SQL queries.

Not all metrics could be determined for all four standards. All property-related metrics are not applicable to UNSPSC since the official release does not contain properties. Those metrics based on the hierarchical structure cannot be used for RNTD, as there is no hierarchy in this standard; they can also be applied to eOTD only in part, as the classifying identifier EGIC is not a fully-fledged hierarchical order. The eOTD files summarize all releases in one huge audit file, so that we had to derive the valid concepts at a given point in time based on the addition and deletion dates in that file. When official release dates were not easily available, we used file dates or date entries in the data bases as an approximation.

We chose to compute all metrics based on the full population data instead of samples. The main reason for this was that a properly gained sample would have required at least similar effort. As a consequence of this approach, respective statistical parameters (i.e. the standard deviation and the coefficient of variation) must be determined as defined for *population* data and not as defined for a sample. This especially affects the computation of the variance and standard deviation.

All in all we gained an enormous set of interesting observations, of which we present the most interesting ones.

3.2 Results

Our analysis shows surprising results in all four types of metrics that go beyond the total number of elements, i.e. with regard to

- the hierarchical order and balance of content,
- the property library,
- the quality of class-specific property sets, and
- the growth and amount of maintenance work.

It is well justified to assume that these findings would have affected the decision about the adoption of a particular standard by an e-business participant.

3.2.1 Absolute size

Table 1 shows the total number of categories, properties, and values as the most obvious metrics for all four PSCS. These numbers, especially the total number of classes, are often used by standards bodies when promoting their standards. All values given in here are based on the actual data in the respective standards and not simply taken from marketing materials.

Table 1 Total number of categories, properties, and values

	Version	Release date	Total number of classes	Total number of properties	Total number of enumerative data values
eCl@ss	5.1de	09-28-2004	25,658	5,525	4,544
UNSPSC	7.0901	09-01-2004	20,789	None	None
eOTD	08-01-2004	08-01-2004	58,970	21,129	16,006
RNTD	4.0	09-15-2004	789	3,623	(497)

One can see that eOTD has the biggest number of categories, followed by eCl@ss and UNSPSC. The fact that RNTD is substantially smaller is no surprise, since the scope of it is limited to a narrow segment. It must also be said that the total number of enumerative data values in RNTD cannot be directly compared to the other standards, since the respective structural element in RNTD is more of a “Term Definition” which holds not only data values but also other entries for useful concepts. Thus it is given in parentheses in the table.

3.2.2 Hierarchical order and balance of content

For horizontal products and services standards, the hierarchical order and the balance of content reveals whether the standard is a true horizontal approach or horizontal just with regard to the existence of segments, but in reality focused quite vertically at the more detailed level. A true horizontal standard requires quite naturally not only the existence of segments for a broad range of concepts but also actual entries in the deeper branches of all segments. Table 2 shows the percentage of concepts contained in the largest and the three largest top-level sections, and relates the size of the largest category to the median. This metric reveals the order of magnitude of the number of concepts in the most populated segment as

Table 2 Dominance of most-populated top-level categories

	Release	% of classes in largest category	% of classes in three 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	5,255
	11-01-2003	24	40	5,254
	03-01-2004	24	40	5,255
	06-01-2004	24	40	5,255
	08-01-2004	24	40	5,255
UNSPSC	6.0315	12	30	1,128
	6.0501	12	29	1,134
	6.0801	12	30	1,134
	6.1101	12	30	1,108
	7.0401	12	30	1,107
	7.0901	12	30	1,107

compared to the median (i.e. the size of the top-level category in the middle of the distribution). The bigger this ratio, the more is the content of the standard dominated by one single segment.

One can clearly see that in all three horizontal standards, the biggest share of categories stems from a very few branches. In eCl@ss, eOTD, and UNSPSC at least 30% of the total number of categories is in the three largest top-level sections. Especially when compared to the median size of all categories, the degree of imbalance is obvious. The largest segment in eOTD is more than 52 times as big as the median, which can be traced back to the bulk import of classes from past standards. In eCl@ss, the largest top-level category is still seven times as big as the median, and in UNSPSC it is eleven times as big.

One can see that the coefficient of variation (i.e. the standard deviation divided by the mean) 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 of 58,970 categories).

The dominance of a few subject areas in all three standards becomes most obvious when the size of the individual top-level categories is visualized in a bar chart, see Fig. 1 (UNSPSC 7.0901), Fig. 2 (eOTD August 1, 2004), and Fig. 3 (eCl@ss 5.1de).

The three biggest top-level categories in the three standards are as follows:

UNSPSC:
1) Medical equipment, accessories, and supplies
2) Manufacturing components and supplies
3) Drugs and pharmaceutical products

eOTD:
1) Medical, dental, and veterinary equipment and supplies
2) Services
3) Miscellaneous

eCl@ss:
1) Organic chemicals
2) Automation, electrical engineering, and PLT[2]
3) Office supplies, furniture, equipment, and papeterie[3]

The degree of imbalance is consistent over time. Apart from the eCl@ss release change from version 4.1 to 5.0, where the ratio of the largest category vs. the median was reduced from 814% down to 731%, there is no significant change in the dominance of a few classes measurable.

This sectoral breakdown shows quite clearly that the simple size metrics defined in Sect. 2.2.1. are insufficient indicators when comparing alternative standards. For example, eCl@ss contains 1,992 categories in the top-level category "Office supplies, furniture, equipment, and papeterie" (segment 24), while UNSPSC offers only 576 categories in the respective subject area (353 in segment 44,

2 This category was renamed to "Electric engineering, automation, process control engineering" in version 5.0.

3 This category was renamed to "Office products, facilities and technics, papeterie" in version 5.0.

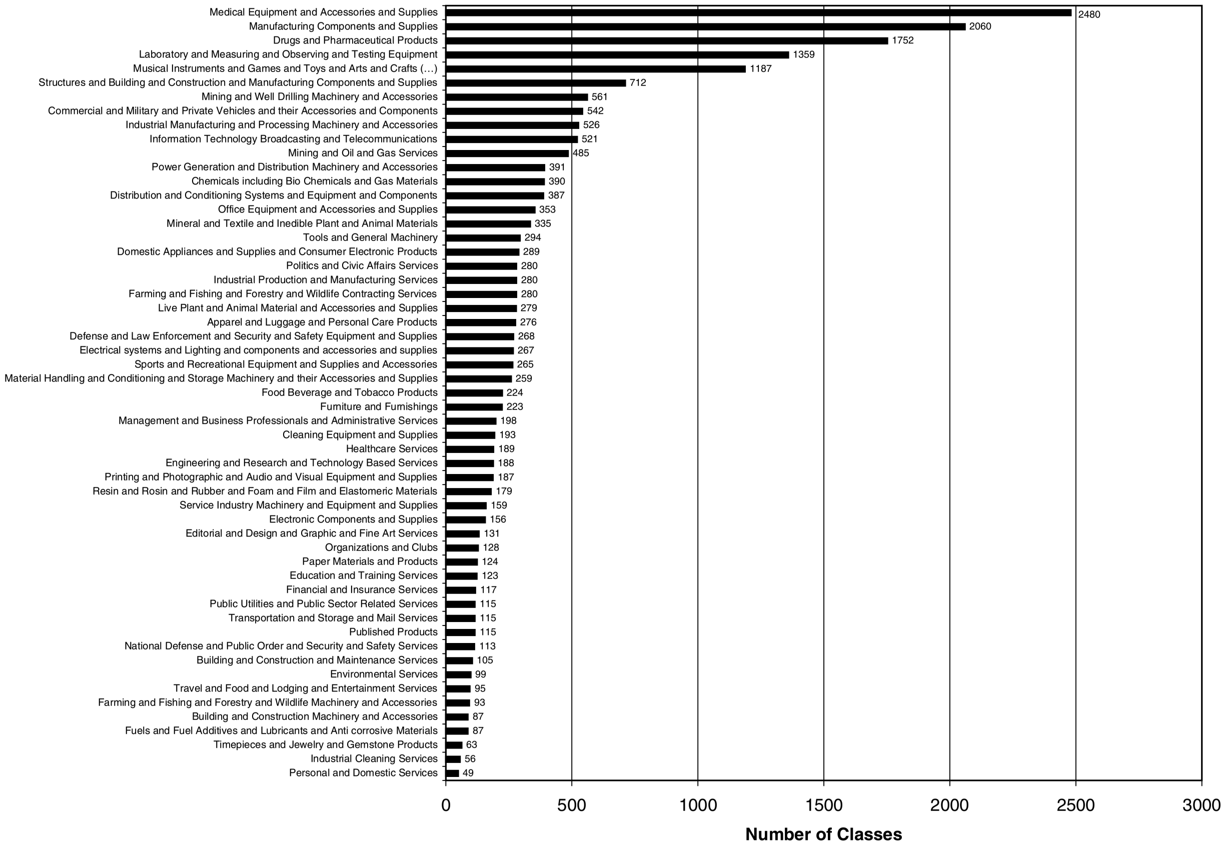

Fig. 1 Uneven population of segments in UNSPSC V. 7.0901

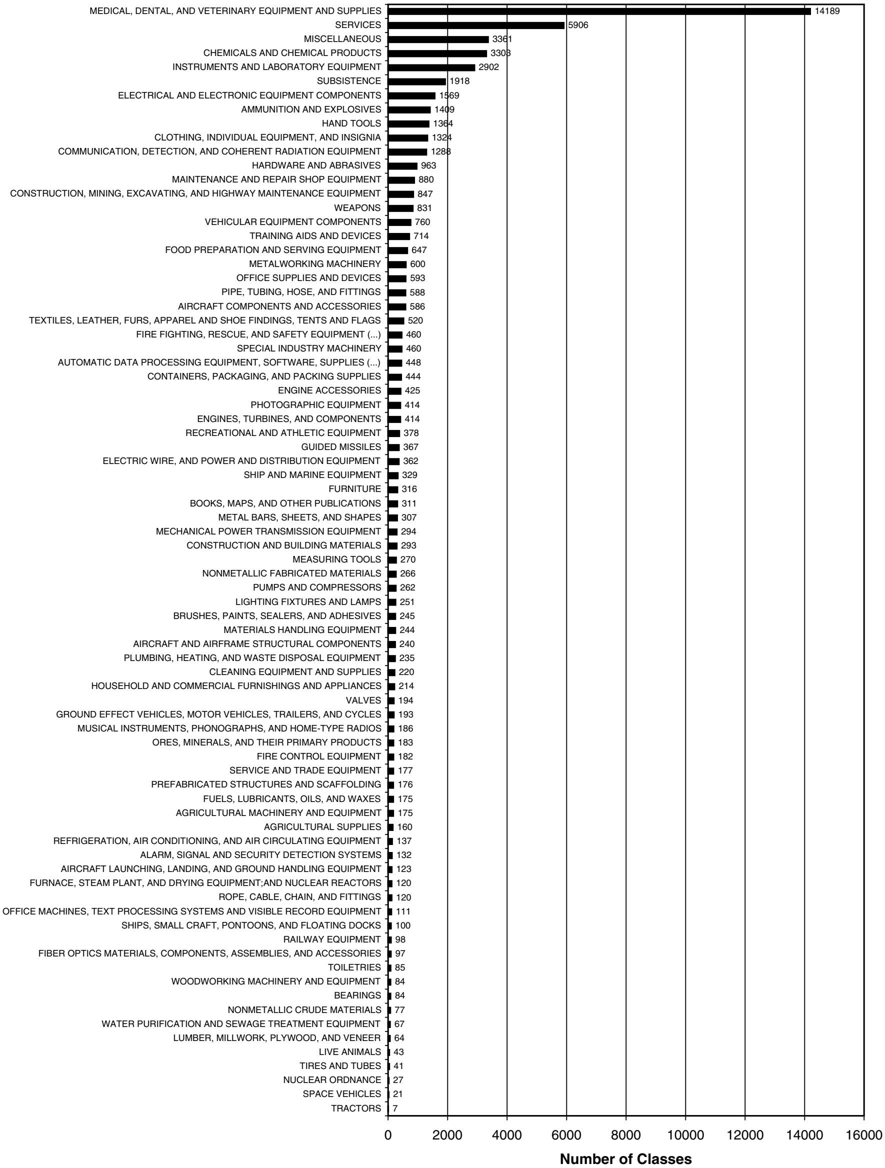

Fig. 2 Uneven population of segments in eOTD (Version dated August 1, 2004)

“Office Equipment and Accessories and Supplies”, and 223 in segment 56, “Furniture and Furnishings”), and eOTD only 1020 (593 in “Office Supplies and Devices”, 316 in “Furniture”, and 111 in “Office Machines, Text Processing Systems, and Visible Record Equipment”). The scope covered by the respective categories will likely not be completely the same, but since the order of magnitude is so substantial, we can assume that the coverage of this topic is much better in eCl@ss than in UNSPSC and eOTD, despite the fact that eOTD contains more than twice

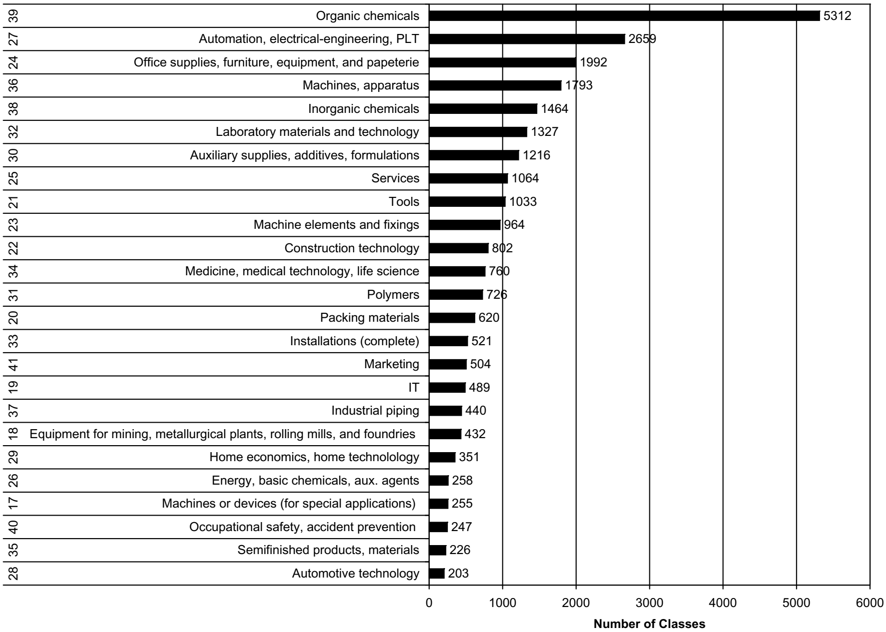

Fig. 3 Uneven population of segments in eCl@ss 5.1de

as many categories in total. This is an important finding, because exactly this range of products is of paramount relevance for e-procurement, since respective projects often target the sourcing of everyday office supplies. Unfortunately, the size of a top-level category is not immediately visible for a decision maker, since it requires importing the standard into an RDBMS and the subsequent execution of nested SQL queries.

The coefficient of variation is a statistical parameter that can be used to compare distributions with a different mean; it is thus a good indicator for the comparison of multiple PSCS with regard to the degree of balance. Table 3 shows the coefficient of variation and other distribution parameters of the number of classes per top-level category for eCl@ss, eOTD, and UNSPSC. One can see that eCl@ss and UNSPSC vary only half as much with regard to the degree of balance than eOTD. This matches what is obvious from the numbers and diagrams; eOTD has extremely small segments (e.g. “Tractors” with just seven categories) and extremely big ones, like the ones names above, with up to 14,189 categories.

Table 3 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	5,312	726.0	432.0	1216.0	1064.6	104
eOTD	08-01-2004	08-01-2004	7	14,189	270.0	148.5	590.5	1764.3	236
UNSPSC	7.0901	09-01-2004	49	2,480	224.0	120.0	370.0	483.3	128

Table 4 Direct descendents per superordinate node: Top-level → 2nd level

	Release	Top-Level → 2nd Level Min	Max	Mean	Median	STD	Coefficient of Variation (%)
eCl@ss	5.1de	4	45	19.2	18	11.8	61
eOTD	08-01-2004	1	8,650	357.0	129	999.6	280
UNSPSC	7.0901	1	29	6.4	5	4.8	74

Table 3 is to be read as following: Q1 (first quartile) indicates the size of that top-level category which is between the 25% smallest categories and the remaining bigger ones. The median is the size of a category that is exactly in the middle of the population, i.e. the one which has the same number of smaller categories below itself than bigger ones above. Q3 (third quartile) reflects the size of that top-level category which separates the 75% smallest categories from the remaining bigger ones. (When there is no single element that holds exactly the position defined for the quartiles and the median, the respective value is the mean of the two adjacent values; this explains the values ending ".5").

Tables 4 through 6 show how the branching downward the hierarchy is distributed, i.e. how many descendents exist per parent node. This reflects whether a specific PSCS is evenly developed at all levels, or whether only some branches are completed down to the leaf level, while others end at a higher level.

Table 7 shows the number and percentage of *services* categories in the three standards having a hierarchical order. As per the definition of the metric in Sect. 2.2.2.2., this does not include services hidden in the deeper levels of the hierarchy. Services differ from 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.

Table 5 Direct descendents per superordinate node: 2nd level → 3rd level

	Release	2nd Level → 3rd Level Min	Max	Mean	Median	STD	Coefficient of Variation (%)
eCl@ss	5.1de	1	83	8.5	6	9.0	106
eOTD	08-01-2004	Not applicable					
UNSPSC	7.0901	1	54	5.8	4	6.3	110

Table 6 Direct descendents per superordinate node: 3rd level → 4th level

	Release	3rd Level → 4th Level Min	Max	Mean	Median	STD	Coefficient of Variation (%)
eCl@ss	5.1de	1	85	5.2	2	8.1	156
eOTD	08-01-2004	Not applicable					
UNSPSC	7.0901	1	92	9.0	6	9.0	100

Table 7 Total number and share of categories reflecting services

	Release	% of services concepts	# of services concepts
eCl@ss	5.1de	4	1,064
eOTD	08-01-2004	10	5,906
UNSPSC	7.0901	21	4,313

Table 8 Size of property library and amount of enumerative data typing

	Release	# of properties (including unused)	# of properties with enumerative data type	% of properties with enumerative data type
eCl@ss	5.1de	5,525	1,064	19
eOTD	08-01-2004	21,129 (28,025)	555	3
RNTD	4.0	3,623	714	20

3.2.3 Property library

eCl@ss, eOTD, and RNTD include a library of properties that can be used to describe product or services in more detail, while UNSPSC does not. The net size of this library indicates 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, e.g. due to different terminological conventions. In addition, enumerative data types for such properties that cannot be unambiguously represented using standard data types are highly desirable. However, we can often observe that such property definitions are incomplete (e.g. defined as any alphanumeric sequence of less than 30 characters), which impedes the automated interpretation of property values.

Table 8 shows the size of the property libraries in eCl@ss, eOTD, and RNTD and the amount of properties with enumerative data typing. The latter is a good indicator for the specificity of properties, since it requires a lot of domain knowledge and community consensus to define enumerative data types for product properties. One can see that eOTD has the biggest amount of properties but the lowest absolute number of enumerative data types. Since nearly a quarter of the 28,025 properties in eOTD is not included in any property set, we counted only the 21,129 properties that are assigned to at least one category.

3.2.4 Quality of class-specific property sets

Property lists tell a standards user which properties should be used to describe a product or service 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 proper-

Table 9 Number of classes with specific property lists

	Release	Total number of properties	# of classes with specific property assignment	% of classes with specific property assignment
eCl@ss	5.1de	5,525	10,930	43
eOTD	08-01-2004	21,120	20,456	35
RNTD	4.0	3,623	789	100

Table 10 Number of properties per property list

		Number of properties in specific property lists					
	Release	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

ties 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 9 shows the number 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 10 illustrates how the number of properties per class varies between the various PSCS. The median and coefficient of variation is surprisingly consistent. In general, a high variation in the number of properties indicates only partial progress in the development of property assignments. As an extension, this metric can be applied to each segment in order to identify those segments that actually contain a high amount of specific property lists.

Very interesting is a detailed analysis of which top-level categories contain specific property lists. Since eOTD does not contain a fully-fledged hierarchy, RNTD contains no hierarchy, and UNSPSC does not include properties at all, we did this analysis for eCl@ss 5.1de only.

For reasons of simplicity, we used the flag “mkbsa” in eCl@ss, which indicates the existence of a specific property set. There is a very small difference to the actual number (our analysis of the actual data base returns 10,930 specific property sets, while the flag is set for 10,933). This, however, can be neglected. The results of this analysis are shown in Table 11. We computed the percentage both based on the total number of categories and based on the leaf level, since eCl@ss does currently not attach properties to any category on the 1st, 2nd, and 3rd level of the hierarchy. One can see that only six of 25 top-level categories have specific property sets for at least 40% of their entries. 14 top-level categories have such for less than 5% of their entries, which means that in the respective subject areas, eCl@ss does not really offer any competitive advantage over UNSPSC, which

Table 11 Percentage of categories with specific property lists in eCl@ss 5.1de, by top-level categories

Top-level	Name	Total number of classes	Number of 4th level classes	# of classes with mkbsa=2	% of classes with mkbsa=2 (of all classes)	% of classes with mkbsa=2 (of leaves only)
17	Machines or devices (for special applications)	255	121	0	0	0
18	Equipment for mining, metallurgical plants, rolling mills, and foundries	432	286	2	1	1
19	IT	489	403	111	23	28
20	Packing materials	620	445	1	0	0
21	Tools	1,033	832	273	26	33
22	Construction technology	802	589	131	16	22
23	Machine elements and fixings	964	739	298	31	40
24	Office supplies, furniture, equipment, and papeterie	1,992	1,669	1,319	66	79
25	Services	1,064	850	827	78	97
26	Energy, basic chemicals, aux. agents	258	214	0	0	0
27	Automation, electrical-engineering, PLT	2,659	2,181	844	32	39
28	Automotive technology	203	131	1	1	1
29	Home economics, home technolology	351	253	16	5	6
30	Auxiliary supplies, additives, formulations	1,216	1,024	0	0	0
31	Polymers	726	499	0	0	0
32	Laboratory materials and technology	1,327	1,186	660	50	56
33	Installations (complete)	521	318	0	0	0
34	Medicine, medical technology, life science	760	580	0	0	0
35	Semifinished products, materials	226	175	94	42	54
36	Machines, apparatus	1,793	1,317	1	0	0
37	Industrial piping	440	329	0	0	0
38	Inorganic chemicals	1,464	1,315	1,315	90	100
39	Organic chemicals	5,312	5,038	5,038	95	100
40	Occupational safety, accident prevention	247	189	2	1	1
41	Marketing	504	417	0	0	0
	Total	**25,658**	**21,100**	**10,933**	**43**	**52**

lacks properties at all. Also, the property assignments in top-level categories 38 and 39 (inorganic and organic chemicals) are specific only in the sense that they include properties not used in other top-level categories, but are otherwise almost identical for all 6353 entries.

As explained in Sect. 2.2.3.4., the novel metrics of Semantic Weight and Semantic Value can help measure the specificity of property sets. We applied this metric to eCl@ss, eOTD, and RNTD. Tables 12 and 13 show the resulting distribution properties of the Semantic Weights and Semantic Values. The median of 0.5 for the Semantic Weights of eCl@ss and eOTD indicates 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 no property set at all. The coefficient of variation in Table 12 shows how much the frequency of inclusion of properties in property sets varies across the property library, and the same column in Table 13 shows how much the specificity of property sets varies over the whole standard. For a detailed discussion of the implications of these values see Sect. 4.

Table 12 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	08-01-2004	0.0000170	1	0.52430	0.5000	0.3908	75
RNTD	4.0	0.0012658	1	0.63834	1.0000	0.3846	60

Table 13 Semantic Values

	Release	Semantic Value					
		Min	Max	Mean	Median	STD	Coefficient of Variation (%)
eCl@ss	5.1de	1.84E-08	9.10E-03	4.74E-05	6.16E-07	2.48E-04	523
eOTD	08-01-2004	2.41E-09	2.21E-03	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

3.2.5 Growth and maintenance

We have already pointed out that the products and services domain is subject to substantial conceptual dynamics due to product innovation (see also [9]), and this on the level of categories, properties, and enumerated data values. This alone makes the amount of maintenance work in a given standard an important indicator. Since we could also show with the previous metrics that current PSCS are far from complete, the amount of additions becomes additionally interesting, since it might help us extrapolate whether there is reason to believe that subject areas that are currently incomplete but important for a given business application will improve in the foreseeable future. In order to achieve a good coverage of concepts needed in the domain, any PSCS requires (1) timely and complete feedback about missing entries from the user community, and (2) a streamlined standardization process that adds respective new elements in a timely manner.

The application of our metrics 2.2.4.1. and 2.2.4.2. returns significant important findings for decision makers in business and standards bodies. We can show that RNTD and eOTD are rather dead collections than actively maintained standards, while eCl@ss and UNSPSC are continuously updated. A breakdown by top-level categories, however, shows that even in eCl@ss and UNSPSC this is true for only a very limited part of their scope. Also, it must be noted that the number of new elements per month has decreased in UNSPSC.

Table 14 shows the total number of classes for a series of releases of eCl@ss, eOTD, UNSPSC, and the RNTD. Table 15 relates the number of new and modified classes for the most recent releases to the amount of time passed in between releases. We can observe that eOTD as the largest set of concepts (58,970 classes in the latest 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. It must be stressed, though, that the amount of update work found in UNSPSC per month has

Table 14 Release intervals and change of the total number of classes over time

PSCS	Version	Previous version	Release date	Days since last release	Total number of classes
eCl@ss	4.1	(3.0)	09-09-2002	n/a	15,315
	5.0	4.1	12-14-2003	461	24,814
	5.0SP1	5.0	03-26-2004	103	24,919
	5.1beta	5.0SP1	08-25-2004	152	25,585
	5.1de	5.1beta	09-28-2004	34	25,658
eOTD	01-17-2003	n/a	01-17-2003	n/a	58,973
	10-01-2003	01-17-2003	10-01-2003	257	58,898
	11-01-2003	10-01-2003	11-01-2003	31	58,901
	03-01-2004	11-01-2003	03-01-2004	121	58,975
	06-01-2004	03-01-2004	06-01-2004	92	58,970
	08-01-2004	06-01-2004	08-01-2004	61	58,970
RNTD	1.2	1.1	07-06-2001	n/a	741
	1.3	1.2	09-21-2001	77	741
	1.4	1.3	11-29-2001	69	741
	2.0	1.4	04-15-2002	137	744
	3.0	2.0	03-03-2003	322	943
	3.1	3.0	12-10-2003	282	770
	3.2	3.1	02-25-2004	77	770
	4.0	3.2	09-15-2004	203	789
UNSPSC	6.0315	5.1001	03-15-2003	165	19,778
	6.0501	6.0315	05-01-2003	47	20,212
	6.0801	6.0501	08-01-2003	92	20,498
	6.1101	6.0801	11-01-2003	92	20,683
	7.0401	6.1101	04-01-2004	152	20,739
	7.0901	7.0401	09-01-2004	153	20,789

decreased substantially in the most recent releases. The comparatively high mean value can be traced back mainly to the releases 6.0315 and 6.0501.

As said, however, even the amount of maintenance work in eCl@ss and UNSPSC is not evenly distributed over the respective standards, but instead restricted to a limited number of top-level categories. In order to analyze this in detail, we measured the amount of change in size per each top-level category of eCl@ss and UNSPSC. Since eOTD underwent almost no maintenance and RNTD has no hierarchical order, we did not perform this analysis for these two standards.

Table 16 shows the change in size between the most recent releases of eCl@ss. We started our analysis with version 4.0, so the values for this first version are not available. One can see that there is maintenance work in almost all top-level categories, including consolidation work that reduced the size of a category. There are minimal discrepancies between the total amount of change between Tables 14 and 16, since we used a slightly different approach for counting new and modified entries; however, the order of magnitude of this is negligible. Table 17 shows the same data as a *relative percentage over the size of this category in the previous release*. Categories 17, 18, and 19 were newly introduced in version 5.0 and have thus no respective value in Table 17.

One can see that especially the major release 5.0 included a substantial revision. The segments “Office supplies, furniture, equipment, and papeterie” (24 with 1493 new classes), “Auxiliary supplies, additives, formulations” (30 with 1,061 new classes) and “Organic Chemicals” (39 with 1709 new classes) accounted for

Table 15 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	279.6	157.4	1271.4
	5.0SP1	5.0	47.8		10.2	
	5.1beta	5.0SP1	131.6		4918.0	
	5.1de	5.1beta	74.1		0	
eOTD	10-01-2003	01-17-2003	6.1	6.2	0	0
	11-01-2003	10-01-2003	4.8		0	
	03-01-2004	11-01-2003	18.3		0	
	06-01-2004	03-01-2004	1.6		0	
	08-01-2004	06-01-2004	0		0	
RNTD	2.0	1.4	0.7	1.3	6.4	1.5
	3.0	2.0	2.4		1.0	
	3.1	3.0	0		0.1	
	3.2	3.1	0		0	
	4.0	3.2	3.4		0	
UNSPSC	6.0315	5.1001	907.8	233.9	135.6	47.5
	6.0501	6.0315	304.5		53.0	
	6.0801	6.0501	97.5		15.0	
	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 16 Breakdown of change in the number of classes per top-level category in eCl@ss

Top-level	Name	4.0en	4.1en	5.0en	5.1en
17	Machines or devices (for special applications)	n/a	0	222	33
18	Equipment for mining, metallurgical plants, rolling mills, and foundries	n/a	0	222	210
19	IT	n/a	0	486	3
20	Packing materials	n/a	45	-40	0
21	Tools	n/a	35	322	127
22	Construction technology	n/a	81	313	37
23	Machine elements and fixings	n/a	36	514	19
24	Office supplies, furniture, equipment, and papeterie	n/a	28	1,493	9
25	Services	n/a	11	700	35
26	Energy, basic chemicals, aux. agents	n/a	15	-26	22
27	Automation, electrical-engineering, PLT	n/a	118	756	127
28	Automotive technology	n/a	46	-79	0
29	Home economics, home technology	n/a	47	40	3
30	Auxiliary supplies, additives, formulations	n/a	52	1,061	0
31	Polymers	n/a	101	45	1
32	Laboratory materials and technology	n/a	18	79	7
33	Installations (complete)	n/a	59	258	59
34	Medicine, medical technology, life science	n/a	55	591	0
35	Semi-finished products, materials	n/a	7	-3	53
36	Machines, apparatus	n/a	138	957	59
37	Industrial piping	n/a	30	-281	30
38	Inorganic chemicals	n/a	315	46	0
39	Organic chemicals	n/a	1,144	1,709	9
40	Occupational safety, accident prevention	n/a	19	20	0
41	Marketing	n/a	0	94	1
			2,400	**9,499**	**844**

Table 17 Relative growth of top-level-categories between release changes in eCl@ss

Top-level	Name	4.0en	4.1en	5.0en	5.1en
17	Machines or devices (for special applications)	n/a	n/a	n/a	14.9%
18	Equipment for mining, metallurgical plants, rolling mills, and foundries	n/a	n/a	n/a	94.6%
19	IT	n/a	n/a	n/a	0.6%
20	Packing materials	n/a	7.3%	-6.1%	0.0%
21	Tools	n/a	6.4%	55.1%	14.0%
22	Construction technology	n/a	21.8%	69.2%	4.8%
23	Machine elements and fixings	n/a	9.1%	119.3%	2.0%
24	Office supplies, furniture, equipment, and papeterie	n/a	6.1%	304.7%	0.5%
25	Services	n/a	3.5%	212.8%	3.4%
26	Energy, basic chemicals, aux. agents	n/a	6.1%	-9.9%	9.3%
27	Automation, electrical-engineering, PLT	n/a	7.1%	42.6%	5.0%
28	Automotive technology	n/a	19.5%	-28.0%	0.0%
29	Home economics, home technology	n/a	18.0%	13.0%	0.9%
30	Auxiliary supplies, additives, formulations	n/a	50.5%	684.5%	0.0%
31	Polymers	n/a	17.4%	6.6%	0.1%
32	Laboratory materials and technology	n/a	1.5%	6.4%	0.5%
33	Installations (complete)	n/a	40.7%	126.5%	12.8%
34	Medicine, medical technology, life science	n/a	48.2%	349.7%	0.0%
35	Semi-finished products, materials	n/a	4.1%	-1.7%	30.6%
36	Machines, apparatus	n/a	21.6%	123.2%	3.4%
37	Industrial piping	n/a	4.5%	-40.7%	7.3%
38	Inorganic chemicals	n/a	28.6%	3.2%	0.0%
39	Organic chemicals	n/a	46.7%	47.6%	0.2%
40	Occupational safety, accident prevention	n/a	9.1%	8.8%	0.0%
41	Marketing	n/a	0.0%	23.0%	0.2%

almost half (44%) of the additions in this release change. However, one has to say that eCl@ss is being maintained in almost all subject areas.

The situation in UNSPSC is quite different, as Table 18 shows: There has been one major release change (6.0315) in which almost all top-level categories were extended. This was the version in which the two competing branches of UNSPSC, one maintained by the UNDP[4] and one maintained by ECCMA[5] were consolidated and thus much of the user input from ECCMA's online system for user feedback was incorporated. Apart from that one-time event, substantial additions have occurred only in three segments: 41 ("Laboratory and Measuring and Observing and Testing Equipment"), 42 ("Medical Equipment and Accessories and Supplies"), and 51 ("Drugs and Pharmaceutical Products"). Segment 60 ("Musical Instruments and Games and Toys and Arts and Crafts and Educational Equipment and Materials and Accessories and Supplies") grew by as many as 1187 categories in that 6.0315 release change but is without any observable maintenance ever since.

Both the substantial change in the 6.0315 update and the isolated additions in the three other segments can be observed very well in the visualization shown in Fig. 4.

A similar breakdown by top-level categories can be applied to the specific property lists. Table 19 shows how the share of specific (see above for a definition) property lists has developed in eCl@ss per each top-level category. One can see that the top-level categories 17, 18, 20, 26, 28. 30, 31, 33, 34, 36, 37, 40, and 41 have no or almost no specific property lists in the most current version. Only

[4] http://www.undp.org and http://www.unspsc.org
[5] http://www.eccma.org/unspsc/

Table 18 Breakdown of change in the number of classes per top-level category in UNSPSC

Segment	Segment Title	7.0901	7.0401	6.1101	6.0801	6.0501	6.0315	5.1001	5.0915	5.0701	5.0415	5.0301	5,0201	5,0115
10000000	Live Plant and Animal Material and Accessories and Supplies	0	0	0	1	-1	39	0	0	0	0	0	0	0
11000000	Mineral and Textile and Inedible Plant and Animal Materials	0	34	0	0	-1	51	0	0	0	0	0	0	0
12000000	Chemicals including Bio Chemicals and Gas Materials	0	-2	1	0	-1	-155	0	0	0	0	0	0	0
13000000	Resin and Rosin and Rubber and Foam and Film and Elastomeric Materials	0	0	0	0	-1	0	0	1	0	0	0	0	0
14000000	Paper Materials and Products	1	0	0	0	-1	49	0	0	0	0	0	0	0
15000000	Fuels and Fuel Additives and Lubricants and Anti corrosive Materials	0	3	1	0	0	30	0	0	0	0	0	0	0
20000000	Mining and Well Drilling Machinery and Accessories	0	3	3	0	-1	55	0	0	0	0	0	0	0
21000000	Farming and Fishing and Forestry and Wildlife Machinery and Accessories	0	0	0	0	-1	3	0	0	0	0	0	0	0
22000000	Building and Construction Machinery and Accessories	0	0	0	0	-1	57	0	0	0	0	0	0	0
23000000	Industrial Manufacturing and Processing Machinery and Accessories	3	0	0	6	-1	206	0	0	0	0	0	0	0
24000000	Material Handling and Conditioning and Storage Machinery and their Accessories and Supplies	0	0	1	0	-1	70	0	0	0	0	0	0	0
25000000	Commercial and Military and Private Vehicles and their Accessories and Components	0	0	0	0	-1	62	0	0	0	0	2	0	0
26000000	Power Generation and Distribution Machinery and Accessories	6	0	1	2	1	64	0	37	1	0	0	-5	0
27000000	Tools and General Machinery	0	0	0	1	-1	96	0	0	0	-1	2	0	0
30000000	Structures and Building and Construction and Manufacturing Components and Supplies	0	0	0	0	0	91	0	0	0	0	0	0	0
31000000	Manufacturing Components and Supplies	0	1	1	3	-1	38	0	21	-3	0	3	0	0
32000000	Electronic Components and Supplies	0	1	0	2	-1	36	0	0	0	0	0	0	0
39000000	Electrical systems and Lighting and components and accessories and supplies	12	1	0	0	0	191	0	0	0	0	1	0	0
40000000	Distribution and Conditioning Systems and Equipment and Components	0	10	1	3	0	56	0	-1	0	-2	0	-10	0
41000000	Laboratory and Measuring and Observing and Testing Equipment	-1	2	3	0	-4	252	0	0	0	**475**	0	0	0
42000000	**Medical Equipment and Accessories and Supplies**	**0**	**-3**	**33**	**0**	**2**	**641**	**5**	**2**	**0**	**1**	**1,275**	**0**	**0**
43000000	Information Technology Broadcasting and Telecommunications	20	1	2	1	3	150	0	45	0	0	0	1	0
44000000	Office Equipment and Accessories and Supplies	2	0	2	1	-1	150	0	0	0	0	0	0	0
45000000	Printing and Photographic and Audio and Visual Equipment and	-1	1	0	0	-1	71	0	0	0	0	0	0	0

categories 24, 25, 38, and 39 have specific property lists for more than 50% of their entries. The table indicates, however, that there is active development of specific property lists in 17 of 25 top-level categories. To our surprise, specific property lists in the top-level category "Industrial piping" (37) have been removed in release 5.1. This is in line with consolidation work on the class level in this subject area. We suppose that eCl@ss is restructuring this area completely.

3.3 Application to use case scenarios

While the application of our metrics as described above has already returned important results that document the content quality of all relevant PSCS, the metrics

Table 18 Continued

	Supplies													
46000000	Defense and Law Enforcement and Security and Safety Equipment and Supplies	0	1	1	0	-1	72	0	1	-1	0	0	0	0
47000000	Cleaning Equipment and Supplies	0	0	0	1	-1	50	0	0	0	0	0	0	0
48000000	Service Industry Machinery and Equipment and Supplies	0	0	0	0	0	11	0	0	0	0	0	0	0
49000000	Sports and Recreational Equipment and Supplies and Accessories	0	0	0	0	-1	8	0	0	0	0	0	0	0
50000000	Food Beverage and Tobacco Products	0	0	17	0	-1	24	0	0	0	0	0	0	0
51000000	**Drugs and Pharmaceutical Products**	**0**	**2**	**96**	**245**	**420**	**461**	**0**	**0**	**0**	**3**	**0**	**0**	**0**
52000000	Domestic Appliances and Supplies and Consumer Electronic Products	2	0	0	4	-1	97	0	0	0	0	0	0	0
53000000	Apparel and Luggage and Personal Care Products	0	0	1	5	-1	80	0	0	0	-1	-4	0	0
54000000	Timepieces and Jewelry and Gemstone Products	0	0	0	6	-1	13	0	0	0	0	0	0	0
55000000	Published Products	0	0	0	2	-1	42	0	0	-1	0	0	0	0
56000000	Furniture and Furnishings	7	0	0	0	-1	117	0	0	0	0	0	0	0
60000000	Musical Instruments and Games and Toys and Arts and Crafts and Educational Equipment and Materials and Accessories and Supplies	0	0	0	0	-1	1,187	0	0	0	0	0	0	0
70000000	Farming and Fishing and Forestry and Wildlife Contracting Services	0	0	0	0	-1	-1	0	0	0	0	0	0	0
71000000	Mining and Oil and Gas Services	0	0	0	0	-1	-1	0	0	0	0	0	0	0
72000000	Building and Construction and Maintenance Services	0	0	3	1	-1	3	0	1	5	0	11	0	0
73000000	Industrial Production and Manufacturing Services	0	0	0	0	-1	0	0	0	1	0	0	0	0
76000000	Industrial Cleaning Services	1	1	0	0	-1	2	0	0	0	0	0	0	0
77000000	Environmental Services	0	0	0	0	-1	0	0	0	0	0	0	0	0
78000000	Transportation and Storage and Mail Services	0	0	0	0	-1	9	0	0	0	0	1	1	0
80000000	Management and Business Professionals and Administrative Services	-2	0	6	1	-1	51	0	0	2	1	1	0	0
81000000	Engineering and Research and Technology Based Services	0	0	0	0	-1	7	0	1	0	0	0	1	0
82000000	Editorial and Design and Graphic and Fine Art Services	0	0	6	0	-1	5	0	0	0	0	0	0	0
83000000	Public Utilities and Public Sector Related Services	0	0	0	0	0	24	0	14	0	0	0	0	0
84000000	Financial and Insurance Services	0	0	2	1	0	16	0	0	0	0	0	0	0
85000000	Healthcare Services	0	0	0	0	-1	3	0	0	0	0	0	0	0
86000000	Education and Training Services	0	0	1	0	-1	0	0	0	0	0	0	0	0
90000000	Travel and Food and Lodging and Entertainment Services	0	0	3	0	-1	5	0	0	0	0	0	0	0
91000000	Personal and Domestic Services	0	0	0	0	-1	0	0	0	0	0	0	0	0
92000000	National Defense and Public Order and Security and Safety Services	0	0	0	0	-1	0	0	0	0	0	0	0	0
93000000	Politics and Civic Affairs Services	0	0	0	0	-1	0	0	0	0	0	0	0	0
94000000	Organizations and Clubs	0	0	0	0	-1	0	0	0	0	0	0	0	0

can also be used by individual market participants for decision making in more specific use cases. The following are examples of such scenarios.

3.3.1 Scenario 1: Maintenance strategy and maintenance efforts

The data models of most ERP packages support the storage of categories and sometimes also respective property values for products and services. However, this is often a 1:1 relationship between a product or purchased item and a respective description based on a PSCS. In other words, it is not possible to handle entries for multiple releases of the same standard. This leads to a situation where the

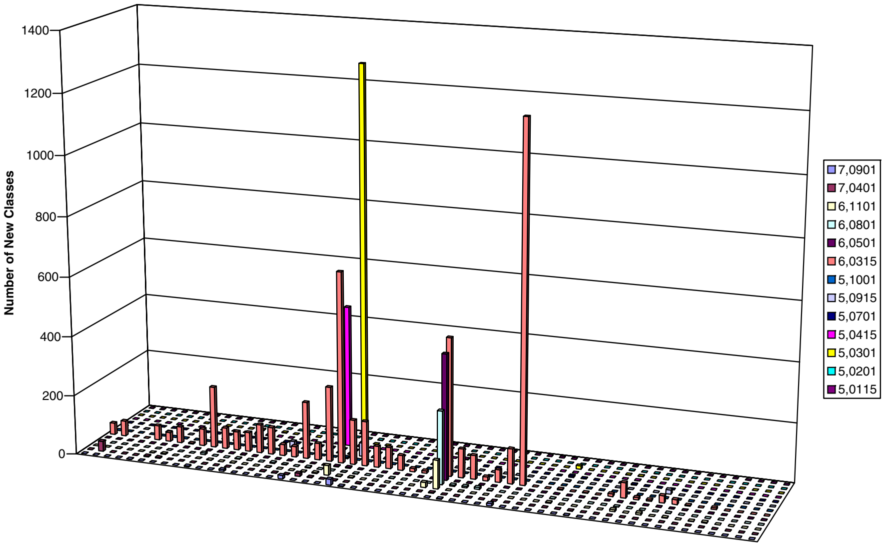

Fig. 4 Visualization of the growth per top-level category from version 5.0115 through version 7.0901 of UNSPSC

migration from one PSCS release to another should take place in one turn, since the master data is inconsistent in the meantime. The metrics that quantify the amount of change in categories and property sets can be used to assess the amount of product entries that require manual updates. In combination with estimates for the amount of time needed per item to be checked, both the need for staffing and the duration of the update can be approximated. Especially the breakdown per top-level categories can be used to check whether the most relevant subject areas are currently actively maintained.

3.3.2 Scenario 2: Steering the development of a standard

The organizations coordinating the maintenance of such standards have a strong need to monitor the total content quality, the amount of progress, strengths and weaknesses in the covered subject areas, for they must distribute limited resources for the maintenance in the most effective way. Also, they may want to identify weak parts of their standard in order to launch extra activities for their improvement, e.g. motivating lead market participants to submit proposals for new content.

In addition to that, the identified weaknesses can be used to assess the amount of labor and other resources necessary to add the content necessary to fill the gaps, and probably to quantify the need for public funding, given that the overall economic benefit of such a lingua franca of e-business can be communicated.

Some PSCS delegate the maintenance of selected top-level categories to independent groups of market participants. In this case, our metrics can be used to monitor the structural consistency of the work done by the respective group (e.g. number of properties per property set, etc.).

Table 19 Comparison of the amount of specific property lists per top-level category in various versions of eCl@ss (based on all classes, not just leaves)

Top-level category	4.0en	4.1en	5.0en	5.1en
17	0%	0%	0%	0%
18	0%	0%	0%	**0.5%**
19	0.0%	0%	**1.6%**	**22.7%**
20	0.0%	0%	**0.2%**	**0.2%**
21	**51.4%**	**51.9%**	**29.2%**	**26.4%**
22	0%	0%	**1%**	**16.3%**
23	0%	**38.5%**	**15.3%**	**30.9%**
24	**15.2%**	**19.0%**	**0.7%**	**66.2%**
25	0%	0%	0%	**77.7%**
26	0%	0%	0%	0%
27	**8.1%**	0%	**10%**	**31.7%**
28	0%	**0.7%**	**0.5%**	**0.5%**
29	0%	0%	0%	**4.6%**
30	0%	0%	0%	0%
31	0%	0%	0%	0%
32	**19%**	**46.7%**	**49.6%**	**49.7%**
33	0%	0%	0%	0%
34	0%	0%	0%	0%
35	**55.6%**	**53.4%**	**54.3%**	**41.6%**
36	0%	**0.3%**	**0.1%**	**0.1%**
37	**44.5%**	**44.9%**	**30.7%**	0%
38	0%	**89.6%**	**89.8%**	**89.8%**
39	0%	**94.2%**	**94.8%**	**94.8%**
40	0%	**0.4%**	**0.4%**	**0.8%**
41	0%	0%	0%	0%

3.3.3 Scenario 3: Comparing the specificity of competing standards

We can assume that a vendor of products wants to select the most specific standard prior to beginning to augment existing product data with PSCS-based descriptions. This often makes sense, since the more specific the annotation is, the more likely it becomes that future content operations can achieve a high degree of automation. An example is the automated generation of catalogs for special interest groups based on the execution of rules over properties, e.g. the creation of a catalog of camping-related products that is to contain all products that operate on 12 V and weigh less than 5 kg. If there are two competing standards available, the decision can be made using our metrics as follows:

(1) Take a random sample of products from the actual range of products (e.g. $n = 10$).
(2) Identify the most suitable product category for each of these products in the PSCS and retrieve their property sets.
(3) Compute the Semantic Value for the property lists for each of these products.
(4) Repeat steps 2 and 3 for any alternative standard.

This allows comparing (1) the percentage of products for which a suitable category exists plus (2) the number of properties per class and (3) the degree of specificity. Note that the mean of a sample (e.g. the percentage of products in the sample that has a suitable category) is a very reliable estimate for mean of the full population, even if the sample size is comparatively small in comparison to the size of the population. If necessary, the sample size can be chosen so that the likelihood that the mean of the sample deviates not more than a given amount from the mean of the population is guaranteed to be below a given probability.

In actual business scenarios, the choice of the internal representation is often additionally constrained by decisions of other parties in the value chain.

4 Discussion

In this section, we compare our results with the underlying research questions as defined in Sect. 2.

4.1 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 most entries (24%) are in one single branch ("Medical, dental, and veterinary equipment and supplies"). Compared to the mean of all segments, this branch is 52 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-→2nd level) to 156% (3rd level → 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.

4.2 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, all PSCS with properties contain 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 assignment using the Semantic Value, one can see that both eCl@ss and eOTD show an enormous spread. The coefficient of variation is as high as 523% (eCl@ss) and 432% (eOTD), while RNTD shows only 155%. Also, RNTD has a mean about 33 (eCl@ss) to 100 times (eOTD) of the size of the others. Those orders of magnitude are 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 a thousand 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 other 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).

4.3 Maintenance

Both eCl@ss and UNSPSC undergo sustained improvement with an average of more than 200 new classes per month, even though the number of new and modified classes in UNSPSC is declining significantly over time. 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 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.

Especially the breakdown by top-level categories reveals very relevant details: UNSPSC has received substantial additions in only three segments: 41 ("Laboratory and Measuring and Observing and Testing Equipment"), 42 ("Medical Equipment and Accessories and Supplies"), and 51 ("Drugs and Pharmaceutical Products"). The rest of the standard has not been actively maintained since March 15, 2003 until September 2004. This is surprising, since UNSPSC is otherwise more balanced in its content than eCl@ss and eOTD.

On the other hand, eCl@ss sees additions in most of its top-level categories. It seems to be that user involvement in eCl@ss is organized in a more effective way. From talks with this standards body, we know that several major enterprises dedicate substantial staff resources to improving the content quality. The fact that there are also classes removed from time to time seems to indicate that also significant consolidation work is taking place.

With regard to the properties assigned to categories, one can see that the ambitious goal of eCl@ss is by far not fulfilled, since a majority of classes does not yet have specific property sets. However, one can see from Table 19 that there is progress in most top-level categories.

One thing seems obvious for us: The current way of submitting feedback is too tedious and time-consuming and the standards bodies add new content only with significant delay. The maintenance and evolution of the standards seems to be too much detached from its users. In previous work, we proposed (already in 2003) that intelligent Web front-ends are necessary which can be employed by standards users to report missing categories or submit change requests easily, and that do pre-classify such requests using simple rule-based mechanisms [9]. UNSPSC underwent the most substantial amount of additions when the content from the ECCMA variant, which included a Web-based system for suggesting new entries, was integrated into the official version. This makes us believe that lowering the barriers for users' participation in the evolution of such standards is crucial. In another stream of work [14, 16], we have recently shown that Wikipedia communities are able to yield a vast amount of consensual concepts without central coordination, just based on clever community control, which also supports this assumption.

5 Conclusions

In this paper, we have provided a wealth of quantitative data about one of the core application domains of representation in business information systems: Products and services. Apart from our direct findings, we hope that many other researchers can use our data for additional work, e.g. cost/benefit analysis or various types of correlation analysis. The direct implications of our work are as follows.

5.1 Theoretical implications

We have presented a comprehensive framework of metrics for analyzing the content quality of all structural elements of popular categorization schemes. This includes a novel pair of metrics called "Semantic Weight" and "Semantic Value" that allow quantifying the specificity of property assignment in such scenarios.

While the metrics were designed specially for the purpose of the analysis presented in this paper, we hope that they can be applied to measuring problems in other domains with similar structural characteristics.

5.2 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.

5.3 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 number 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.

Acknowledgements The work presented in this paper is partly funded by the European Commission under the project DIP (FP6-507483), the TransIT Entwicklungs- und Transfercenter at the University of Innsbruck, and Florida Gulf Coast University.

References

1. Albers M, Jonker CM, Karami M, et al. (2004) Agent models and different user ontologies for an electronic market place. Knowl Inf Syst 6:1–41
2. Beneventano D, Guerra F, Magnani S, et al. (2004) A Web Service based framework for the semantic mapping amongst product classification. J Electron Comm Res 5:114–127
3. Corcho O, Gómez-Pérez A (2001) Solving integration problems of e-commerce standards and initiatives through ontological mappings. In: CEUR workshop proceedings 47, Workshop on e-business and intelligent web, pp 131–140
4. Fairchild AM, de Vuyst B (2002) Coding standards benefiting product and service information in e-commerce. In: Proceedings of the 35th annual Hawaii international conference on system sciences (HICSS-35). IEEE Computer Society, Los Alamitos, CA
5. Fensel D, Ding Y, Omelayenko B, et al. (2001) Product data integration in B2B e-commerce. IEEE Intell Syst 16:54–59
6. Fensel D, McGuinness DL, Schulten E, et al. (2001) Ontologies and electronic commerce. IEEE Int Syst 16:8–14
7. Fenton NE, Pfleeger SL (1996) Software metrics: a rigorous and practical approach. International Thomson Computer Press, London

8. He M, Leung HF (2002) Agents in e-commerce: state of the art. Knowl Inf Syst 4:257–282
9. Hepp M (2003) Güterklassifikation als semantisches Standardisierungsproblem. Deutscher Universitäts-Verlag, Wiesbaden
10. Hepp M (2004) Measuring the Quality of descriptive languages for products and services. In: Dorloff FD, Leukel J, Schmitz V (eds) E-business—Standardisierung und Integration. pp 157–168. Cuvillier, Göttingen
11. Hepp M (2005) Representing the hierarchy of industrial taxonomies in OWL: the gen/tax approach. In: CEUR workshop proceedings 155, Workshop semantic web case studies and best practices for eBusiness (SWCASE05), pp 49–56
12. Hepp M, Leukel J, Schmitz V (2005a) Content metrics for products and services categorization standards. In: Proceedings of the IEEE international conference on e-technology, e-commerce and e-service (EEE-05), pp 740–745. IEEE Computer Society, Los Alamitos, CA
13. Hepp M, Leukel J, Schmitz S (2005b) A quantitative analysis of eCl@ss, UNSPSC, eOTD, and RNTD content, coverage, and maintenance. In: Proceedings of the IEEE international conference on e-business engineering (ICEBE 2005), pp 572–581. IEEE Computer Society, Los Alamitos, CA
14. Hepp M, Bachlechner D, Siorpaes K (2005c) OntoWiki: community-driven ontology engineering and ontology usage based on Wikis. In: Proceedings of the 2005 international symposium on Wikis (WikiSym 2005)
15. Hepp M (2006) Products and services ontologies: a methodology for deriving OWL ontologies from industrial categorization standards. Int J Semantic Web Inf Syst (IJSWIS) 2:72–99
16. Hepp M, Bachlechner D, Siorpaes K (2006) Harvesting Wiki consensus—using Wikipedia entries as ontology elements. In: CEUR workshop proceedings 206: SemWiki2006—from Wiki to Semantics
17. Leukel J, Schmitz V, Dorloff FD (2002) A modeling approach for product classification systems. In: Proceedings of the 13th international workshop on database and expert systems applications (DEXA'02), pp 868–874. ACM Press, New York, NY
18. Li H (2000) XML and industrial standards for electronic commerce. Knowl Inf Syst 2:487–497
19. Magkanaraki A, Alexaki S, Christophides V, et al. (2002) Benchmarking RDF schemas for the semantic web. In: LNCS Vol. 2342, Proceedings of the first international semantic web conference (ISWC2002), pp 132–146. Springer, London
20. Obrst L, Wray RE, Liu H (2001) Ontological engineering for B2B e-commerce. In: Proceedings of the international conference on formal ontology in information systems (FOIS'01), pp 117–126. ACM Press, New York, NY
21. Omelayenko B (2001) Ontology integration tasks in business-to-business e-commerce. In: LNCS Vol. 2070, Proceedings of the 14th international conference on industrial & engineering applications of artificial intelligence & expert systems, pp 119–124. Springer, London
22. Schulten E, Akkermans H, Botquin G et al. (2001) The e-commerce product classification challenge. IEEE Intell Syst 16:86–89
23. Stonebraker M, Hellerstein JM (2002) Content integration for e-business. In: Proceedings of the 2001 ACM SIGMOD international conference on management of data, pp 552–560. ACM Press, New York, NY
24. U.S. Census Bureau (2006) North American industry classification system (NAICS). Available at http://www.census.gov/epcd/www/naics.html. Retrieved March 15, 2006
25. Yan G, Ng WK, Lim EP (2002) Product schema integration for electronic commerce—a synonym comparison approach. IEEE Trans Knowl Data Eng 4:583–598

Author Biographies

Martin Hepp is a Senior Researcher at the Digital Enterprise Research Institute (DERI) at the University of Innsbruck in Innsbruck, Austria, where he heads the research group "Semantics in Business Information Systems". He created eClassOWL, the first industry-strength ontology for products and services and is currently working on using Semantic Web services technology for Business Process Management. Before joining DERI, he was an Assistant Professor of Computer Information Systems at Florida Gulf Coast University, Fort Myers (FL) and a Visiting Scientist with the e-Business Solutions Group at the IBM Zurich Research Laboratory. Martin holds a Master's degree in Business Management and Business Information Systems and a Ph.D. in Business Information Systems from the University of Wuerzburg (Germany).

Joerg Leukel is a Senior Researcher and Lecturer in the Department of Information Systems II at the University of Hohenheim in Stuttgart, Germany, where he leads the Logistics Research Group. Prior to this, he was with the University of Duisburg-Essen, where he earned a M.Sc. and a Ph.D. degree in Business Information Systems. His research interests are in the areas of inter-organizational information systems, supply chain management, product lifecycle management, semantic interoperability, business ontologies, and standardization.

Volker Schmitz is a Senior Researcher and Lecturer at the Institute for Computer Science and Business Information Systems at the University of Duisburg-Essen, Germany. He holds a M.Sc. in Computer Science from RWTH Aachen University and is a member of several national and international standardization initiatives. Since 1999, he has been responsible for the development of the e-business standards BMEcat and openTRANS. His research and teaching interests include e-procurement, e-markets, catalog and product data management, product classifications, business ontologies, and standardization.