

Subject Fields in a Controlled Natural Language: How the Evolution of the ASD-STE100 Specification Led to a Proposal for a Global Structured Review of Term Categories (Short Paper)

Daniela Zambrini ¹, Orlando Chiarello ²

¹ ASD Simplified Technical English Maintenance Group, Rome, Italy

² ASD Simplified Technical English Maintenance Group, Varese, Italy

Abstract

This paper illustrates the state of the art of an ongoing pilot project for the review of subject fields in the current Issue 8 of the ASD-STE100 Specification. The idea for the project was driven by two key factors: to ensure compliance of the Specification with Terminology Standards and to meet the increasing needs of its intended users (technical authors, translators, and end-users of the technical documentation).

We will discuss the history of the legacy terminology which has been used to define subject fields and domains in STE since 1986, the challenges of harmonizing such terminology with current methods and official standards, and the reason for selecting the FAIR methodology and available tools for the related teamwork.

Keywords

Simplified Technical English, ASD-STE100, subject fields, domains, Lench, EuroVoc, FAIR

1. Introduction

Within the framework of Controlled Natural Languages (CNL) - subsets of natural languages with restricted grammar and vocabulary that reduce or eliminate ambiguity and complexity - Simplified Technical English (also known as STE) is one of the most widely used in technical documentation. It is regulated by the ASD-STE100 specification [1].

STE was created as a controlled language in the early 1980s (originally AECMA Simplified English) to help non-native English speakers understand technical manuals written in English. Initially, STE was only applied to maintenance documentation for civil aircraft. It later became a requirement for defense programs. Maintenance and technical manuals are now written in STE, in a variety of other sectors and industries.

The ASD-STE100 Simplified Technical English specification consists of two Sections:

- The Writing Rules
- The Dictionary

The writing rules cover aspects of grammar and style. They differentiate between two types of texts: procedural writing and descriptive writing.

The dictionary contains the approved general words to write in STE and some unapproved words. As a core vocabulary, these approved general words are sufficient to write technical texts. The approved words were chosen for their ease of recognition, flexibility, frequency, and simplicity.

In addition to the general STE vocabulary that is listed in the dictionary, the ASD-STE100 Specification gives rules for using *technical names* and *technical verbs* that are not included in the dictionary when authors need such words to write technical information or procedures.

Section 2 of this paper will illustrate the primary stages of the pilot project for a proposal to harmonize some STE terminology with ISO standards.

2nd International Conference on “Multilingual digital terminology today. Design, representation formats and management systems” (MDTT) 2023, June 29 – 30, Lisbon, Portugal

EMAIL: daniela@exel8.com (D. Zambrini); orlando.chiarelo@secondomona.com (O. Chiarello)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Section 3 will show if and how the categories of STE technical names can be renamed or tagged with Lenech classification codes.

Section 4 will illustrate the creation of some terminological records using the FAIR methodology. The records will refer to terms selected from the STE dictionary (both approved and not approved words) and to some examples taken from technical name categories.

2. The ASD-STE100 Dictionary and Technical Names

The ASD-STE100 dictionary has a very simple structure. It is divided into four columns. The first column contains the word, in bold text, followed by the part of speech in parentheses. The text is in uppercase for approved words and in lowercase for unapproved words. The second column gives the approved meaning or possible alternatives (if the word is not approved). The third column shows an approved example sentence in STE (the uppercase is for ease of reference) and the fourth column shows an example sentence that is not approved (the lowercase is for ease of reference and comparison).

Word (part of speech)	Approved meaning/ ALTERNATIVES	APPROVED EXAMPLE	Not approved example
lost (adj)	DECREASE (v)	IF THE PRESSURE DECREASES QUICKLY, THE SHEAR PINS IN THE LATCH WILL BREAK.	If the pressure is lost rapidly, the shear pins in the latch will break.
	NOT FIND	IT IS POSSIBLE THAT YOU CANNOT FIND THE RUBBER BUSHINGS.	The rubber bushings may be lost.
LOW (adj) (LOWER , LOWEST)	That is of small value	THE SWITCH SENSES LOW PRESSURE IN THE DISTRIBUTION CIRCUIT.	
LOWER (v), LOWERS , LOWERED , LOWERED	To move something down	LOWER THE ANTENNA TO GET ACCESS TO THE CABLE CONNECTOR.	

Figure 1: ASD-STE100 dictionary layout

Each approved word in STE is only permitted as a specific part of speech. For example, the word ‘test’ is only approved as a noun (‘the test’) but not as a verb (‘to test’). There are rare exceptions to the “One word, one part of speech, one meaning” principle.

In addition to the general vocabulary that is listed in the dictionary, it is necessary to use *technical names* and *technical verbs* to write technical texts. Nouns, noun clusters, scientific terms, or verbs such as ‘grease’, ‘discoloration’, ‘propeller’, ‘aural warning system’, ‘overhead panel’, ‘momentum’, ‘specific gravity’, ‘to ream’, and ‘to drill’ are not listed in the dictionary, but they qualify as approved terms according to writing rule 1.5 on *technical names* and writing rule 1.12 on *technical verbs*, if they can be allocated in the relevant category with the correct meaning. In current Issue 8 there are 20 categories of *technical names* and 5 categories of *technical verbs*.

The study described in this paper focuses on the challenges faced when reviewing some of the legacy terminology used within the Specification.

It is first and foremost necessary to preserve the spirit that led to the creation of the Specification to and its maintenance and development in the last forty years. It is therefore essential to keep the instructions in the Specification itself as simple and as clear as possible for authors and users. The Specification must refrain from excessive academic references or complex and obscure concepts. On the other hand, our project wishes to propose a series of changes. These suggestions shall not and must not disrupt or contradict the previous work of the pioneers and longstanding contributors of this CNL. They are merely intended to propose a method to harmonize and adjust some of its key vocabulary elements with international standards.

We therefore decided that an interesting starting point for the pilot project would involve the complex term *technical name*. Ongoing work of the project and future work focuses on testing a hierarchal structure for the examples in each technical name category.

2.1. What’s in a Name? From Technical Name to (Technical) Term

For a non-expert STE user, there may be a possible source of confusion with the prominent position and frequency of the word ‘name’. Originally, this came from the name of items in spare parts catalogs, technical drawings, etc. But, as the categories of technical names evolved from common nouns used in specified contexts, proper nouns for places, persons, or things, particular industry, company or domain-specific noun clusters to also include abstract concepts such as numbers and symbols, adjectives for colors, is the word ‘name’ in *technical name* really the most accurate choice, or should it be better replaced by ‘(technical) term’?

Taking a step back in time, in the pre-release Issue 0 of AECMA Simplified English (1985) [2], the original designation for words that can be used (nouns, symbols, numbers) but that are not included in the dictionary was *technical term*.

However, this was subsequently changed to *technical name* in Issue 1 (1986) [3]. Similarly, approved verbs that were not originally included in the dictionary were designated as *manufacturing processes* and only became *technical verbs* in Issue 1, Revision 2, in 2001) [4].

In-depth research in the working papers and in the AECMA Simplified English meeting minutes did not show any report on the reasons for replacing *technical term* (used in pre-release issue 0) with *technical name* (used from Issue 1 onward).

Because “a technical term is considered to be “the name” of a concept while a definition is the extended comprehensive interpretation of this name” (Superanskaya 2003) [5], the first challenge we encountered in this study was how to rename a legacy special term such as *technical name* which - in itself - from a meta-terminological point of view, might not seem to be the most accurate of designations.

ISO 1087-1:2000 defined a term as a *Verbal designation of a general concept in a specific subject field* [6]. ISO 1087:2019(E) revised the definition of a term as a *designation that represents a general concept by linguistic means* [7]. The latter is the definition that is used in the tekomp Europe e.V. (European Association for Technical Communication) [8] terminology database for technical communication [9], where *technical term* and *specialized term* are labeled as not to be used.

Therefore, if we simply focus on classical and standardized meta-terminology nomenclature, and in accordance with tekomp Europe’s approach in [9], we believe that a more accurate designation would be the short and simple word ‘term’. However, in keeping with a parallel of adjective-noun pair that is used for the verbs (*technical verbs* - TV -, a development from manufacturing processes, used until 2001 [4]) we suggest using the pair *technical term* (TT). Its two-letter abbreviation will also be easily recognizable and retrievable in a digital terminology management solution. We shall demonstrate, in section 3, that having both functions abbreviated with two letters will also be helpful in a digital search phase.

Change proposals for the ASD-STE100 specification undergo a strict procedure. A proposal must be submitted to the Simplified Technical English Maintenance Group (STEMG) using a Change Form. All Change Forms are analyzed and discussed during the STEMG meetings. If a proposal for change is accepted, it is incorporated in the next issue of the Specification.

When this pilot project is completed, a number of Change Forms shall be submitted to the STEMG for analysis and approval. The first Change Form will be considered as a proposal for a major revision because it suggests replacing *technical name* with *technical term*.

Because our project is still ongoing, within this paper we shall continue to refer to the official STE designation *technical name*.

2.2. Technical Name Categories

There are 20 technical name categories in current issue 8 of the ASD-STE100 specification. This number has remained unchanged, with few exceptions, since Issue 1 was released in 1986. Over the years, some of these categories merged and made room for new ones, because technological evolution led to the need for more categories and new examples. Furthermore, in each new Issue some of the original or previous examples have been removed, deprecated, or replaced when obsolete.

As the Specification broadened its scope outside the aerospace and defense industry, so were some of the names of the categories adjusted. For the same reason, a large part of the example sentences in the

general dictionary itself were modified and updated to better represent more industries and different types of technical publications.

The STE dictionary only includes general words (classified in eight parts of speech). Therefore, it is not a technical or terminological dictionary as originally defined in [6] (*technical dictionary* as a synonym for *Terminological Dictionary. Collection of terminological entries presenting information related to concepts or designations from one or more specific subject fields*). Nor is it a *special-language dictionary (LSP dictionary) terminology resource that is designed to be used as a reference work* as redefined in [7].

We cannot consider the contents of the technical name categories or the entire list of categories as a terminological dictionary. Additionally, even the examples for technical names lack the full details of a dictionary entry and the structure of even the simplest of terminology data records.

We can, however, easily confirm that the use of the designation “category” for technical names and technical verbs is compliant and consistent with ISO 5127:2017 *Category: A broad facet, primary division of a special classification system or of a main class of a general one* [10].

Thus, the next step of the pilot project was to tag the technical name categories with a publicly available subject field classification system [11].

3. Tagging Technical Names with a Subject Field Classification

The EuroVoc [12] system was discarded because its 21 subject fields only cover subject areas which are of importance for the activities of the European institutions, and these were therefore insufficient for STE and the pilot project.

Similarly to EuroVoc, one of the distinctive features of STE is its limitation of polyhierarchy.

We chose to proceed with the Lench [13] universal classification system because its more fine-grained classification system covers a wide range of subject fields (48 subject fields with further subclasses, labelled with a letter or digit that follows the subject field code).

However, it was also immediately evident that not all the titles of the technical name categories could be tagged with subject field codes using the Lench system. Indeed, this would defeat the purpose of leaving the tasks of final analysis, classification, and validation of a new technical name to responsibility of the author or of the company linguistic department [14].

Nevertheless, the broadness of the category titles (as renamed in subsequent editions, revisions and issues of the Specification, from 1986 onwards) confirms the applicability of STE to many more industries and types of documentation than those envisaged when it was first developed.

For the purpose of conciseness, although our ongoing project intends to verify the feasibility of renaming and classifying the subject fields for all categories of technical names and of classifying every listed example, in this paper we summarize the pilot project classification for four categories according to the Lench system (Categories 4, 7, 12, and 14). Then, we use the contents of those same categories (examples of technical names) to create term records according to the FAIR terminology principles.

According to the results of the pilot project analysis, the title of category 4 (Names of materials, consumables, and unwanted material) is too broad to fit into a single subject field. It would require too many subject-field tags. The large amount of further classification codes would therefore be confusing and make the coded information useless.

The title of Category 7 (Mathematical, scientific, engineering terms, and formulas) can easily be tagged with multiple codes (MA Mathematics, PH Physics, CH Chemistry) but again, “engineering terms” is too broad.

The pilot project suggests splitting Category 7 into multiple separate categories. However, this suggestion is not feasible according to the guiding principles of STE, nor with Note 1 in 3.14 in ISO 1087:2019(E) [7]. The note states that “*The borderlines and the granularity of a domain are determined from a purpose-related point of view. If a domain is subdivided, the result is again a domain.*”

On the contrary, Categories 12 (Parts of the body) and 14 (Medical terms) can be merged and tagged with the subject field code Medicine (ME), followed by its hierarchical structure (from ME1 to ME9 and MEA to MET) and relevant examples.

Therefore, to ensure consistency with [7] and even more so with the simplicity guiding principle of STE and to prevent any risk of ambiguity and misunderstanding, we concluded that it is more efficient

and logical to propose and classify the subject field relevant to the technical name examples and not the technical name categories themselves.

These subject field labels or tags can then also be used to generate a unique identifier in a TBX header as described in 4.2 ISO/TS 24634:2021 [15].

4. FAIRterm Records

The final phase of the project used the STE dictionary and the examples of technical name and technical verb categories to generate a set of terminological data records that complies with the FAIR Principles (Findable, Accessible, Interoperable and Reusable) [16]. These principles ensure that data can easily be found by both humans and computers with the ultimate goal of such data being reused. Subject field labels and unique identifiers [17] will therefore be essential in the first step of the “Fairification” process.

The FAIRterm Web Application [18] was chosen for multiple reasons. First, it is a free tool for the compilation of terminological records that has been designed meet the FAIR principles and implemented according to the ISO TC/37 SC/3 [19] standards. Additionally, it allowed participants in the pilot project to work in a collaborative space.

For the pilot project and for this paper, we populated a number of terminological records created with the FAIR methodology and the FAIRterm web app. These records, tagged with Lench codes, are for words selected from the STE dictionary (approved and not approved words) and for examples taken from categories 4, 7, 12 and 14 of technical names. The selection includes (but is not limited to) the following words and technical names:

- abnormality (n)
- obstruction (n)
- BLOCKAGE (n)
- Category 4: aluminum alloy
- Category 7: defect
- Category 12: skin
- Category 14: irritation, skin irritation.

5. Conclusions and Future Work

The early stages of the study confirmed that it is possible to harmonize part of the legacy terminology used in the ASD-STE100 specification with current standards [20]. However, implementation of the study results requires close collaboration with the STEMG and further work. The project will continue until all term records are compiled from the examples in the technical name categories [21].

Once this stage is completed, it will also be possible to submit the Change Forms to the STEMG, to propose *technical term* instead of *technical name* and to propose re-designing categories with the addition hierarchical sub-domains. The release for Issue 9 is scheduled for April 2024.

Upon positive feedback from the STEMG, the final stage will be to deploy a FAIR data resource with all the term records.

6. References

- [1] ASD-STE100 Simplified Technical English, International specification for the preparation of technical documentation in a controlled language, Issue 8 (2021), ASD, Brussels, Belgium
- [2] AECMA Simplified English for the preparation of aircraft maintenance documentation, AECMA Document: PSC-85-16598, Issue 0 (1985), AECMA, Brussels, Belgium
- [3] AECMA Simplified English, A guide for the preparation of aircraft maintenance documentation in the international aerospace maintenance language, AECMA Document: PSC-85-16598, Issue 1, Rev. 0 (1995), AECMA, Brussels, Belgium
- [4] AECMA Simplified English, A guide for the preparation of aircraft maintenance documentation in the international aerospace maintenance language, AECMA Document: PSC-85-16598, Issue 1, Rev. 2 (2001), AECMA, Brussels, Belgium
- [5] Superanskaya, A.V., Podolskaja, N.V., Vasiljeva, N.V. (2003). General Terminology: Theoretic problems. Moscow: URSS, 248.
- [6] ISO 1087-1:2000. Terminology work — Vocabulary — Part 1: Theory and application
- [7] ISO 1087:2019(E) Terminology work and terminology science — Vocabulary
- [8] tekomp Europe, URL: <https://www.technical-communication.org/>
- [9] termXplorer, URL: <https://tekomp.termtechnologies.com/>
- [10] ISO 5127:2017 Information and documentation — Foundation and vocabulary
- [11] Aukšoriūtė A., Belogrīvs I., Bielevičienė A. [et al.]. Towards Consolidation of European Terminology Resources: Experience and Recommendations from the EuroTermBank Project. Signe Rirdance and Andrejs Vasiljevs (eds.). Tilde, Riga, Latvia, 2006. 2.
- [12] EuroVoc handbook, URL: http://publications.europa.eu/resource/cellar/a2723a83-574f-11eb-b59f-01aa75ed71a1.0001.01/DOC_1
- [13] Lench classification system, URL: <http://www2.uibk.ac.at/translation/termlogy/lenoch.html>
- [14] Warburton, K. “Subject Fields in Termbases - Their Design, Use and Representation (Poster).” In Proceedings of the 1st International Conference on Multilingual Digital Terminology Today, Vol. 3161. CEUR Workshop Proceedings. Padua, Italy: CEUR, 2022. <https://ceur-ws.org/Vol-3161/#poster4>.
- [15] ISO/TS 24634:2021 Management of terminology resources — TBX-compliant representation of concept relations and subject fields
- [16] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). URL: <https://doi.org/10.1038/sdata.2016.18>
- [17] Vezzani, F. (2022). Terminologie numérique : conception, représentation et gestion. Lausanne, Switzerland: Peter Lang Verlag. Retrieved Jan 6, 2023, from 10.3726/b19407
- [18] FAIR terminology, URL: <https://shiny.dei.unipd.it/fairterm/>
- [19] ISO/TC 37/SC 3 Management of terminology resources
- [20] ISO 15188:2001 Project management guidelines for terminology standardization
- [21] ISO 12620-1:2022 Management of terminology resources Data categories — Part 1: Specifications