

Committee on WIPO Standards (CWS)

Tenth Session

Geneva, November 21 to 25, 2022

PROPOSAL FOR THE REVISION OF WIPO STANDARD ST.26

Document prepared by the International Bureau

SUMMARY

1. The Sequence Listings Task Force (SEQL Task Force) is proposing revisions to the current version of WIPO Standard ST.26 to correct editorial mistakes noted by the Task Force and to include three new examples in Annex VI. At the ninth session of the Committee on WIPO Standards (CWS), the International Bureau noted the concerns of Offices regarding further revisions to WIPO ST.26 and has provided guidance on how to manage these changes.

BACKGROUND

2. At its reconvened fourth session, held in March 2016, the CWS adopted WIPO Standard ST.26 (see paragraph 52 of document CWS/4BIS/16), entitled "Recommended standard for the presentation of nucleotide and amino acid sequence listings using XML (eXtensible Markup Language)". The SEQL Task Force is responsible for amendments to WIPO ST.26, under the framework of Task No.44 whose description reads:

"Support the International Bureau by providing users' requirements and feedback on the ST.26 authoring and validation software tool; support the International Bureau in the consequential revision of the PCT Administrative Instructions; and prepare necessary revisions of WIPO Standard ST.26."

3. Revisions to WIPO ST.26 were approved by the CWS at its fifth, sixth, seventh and eighth sessions, resulting in versions 1.1, 1.2, 1.3 and 1.4 respectively (see paragraph 41 of document CWS/5/22, paragraph 112 of document CWS/6/34, paragraph 133 of document CWS/7/29 and paragraph 49 of document CWS/8/24). Most recently at its ninth session, held in November 2021, several editorial and substantial revisions were approved by the CWS and published as the most recent version of WIPO ST.26, version 1.5 (see paragraph 63 of document CWS/9/25).

This included changes to maintain alliance with the International Nucleotide Sequence Database Collection (INSDC) requirements.

4. During the ninth session, concerns were expressed by delegations regarding the potential impacts of revisions to WIPO ST.26 after all Member States had implemented WIPO ST.26, in case their IT systems required retooling for instance (see paragraph 62 of document CWS/9/25). The International Bureau indicated that they would provide guidance on this in the tenth session.

5. Several editorial changes were made to Annex VI and the Appendix to Annex VII after it was noted that there were several referencing errors for the SEQ IDs and incorrect feature locations provided after the revision made at the ninth session. These amendments were made and published in May 2022.

PROPOSED REVISION OF WIPO STANDARD ST.26

6. The SEQL Task Force has prepared a proposal for revision to WIPO ST.26 for consideration and where appropriate approval by the CWS. Where indicated strikethrough text indicates a change and bold text indicates an addition.

7. The proposed changes to the Main Body of WIPO Standard ST.26 can be summarized as follows:

(a) Replace the first sentence of the editorial note to the following regarding the entry into force for the new version of the Standard, version 1.6:

The Tenth Session of the Committee on WIPO Standards decided that version 1.6 of WIPO Standard ST.26 will enter into force on July 1, 2023. Meanwhile, version 1.5 of Standard ST.26 should continue to be used.

(a)(b) Update paragraph 44 to improve the filename value provided such that it includes '-' in place of '_': *filename="US11-405455-SEQL.xml"* considering that some IPOs do not allow the underscore '_' within the filename of submitted electronic files;

(b)(c) Update paragraph 45 to improve the definition of ApplicationNumberText to read: *'The application number as provided by the office of filing (e.g., PCT/IB2013/099999).'*;

(d) Update paragraph 73 to ensure consistency with the description of Section 6 by including the following phrase in the second half of the first sentence: *'Section 8 provides the exclusive listing of qualifiers, **and their specified value formats, if any**, for each amino acid sequence feature key.'*;

(e)(e) Update paragraph 76, to indicate that there is a definition provided for 'and one attribute':

(d)(f) Update paragraph 76 to provide a new definition of the 'id' attribute that is provided for language dependent free text qualifiers as part of new fourth row of the included table. The description reads as follows: *A qualifier with language-dependent free text value may be uniquely identified by using the optional XML attribute 'id' in the element INSDQualifier (see paragraph 87(d)). The value of the 'id' attribute must start with the letter 'q' and continue with any positive integer. The value of an 'id' attribute must be unique to one INSDQualifier element, i.e. the attribute value must only be used once in a sequence listing file.*~~*A qualifier with language-dependent free text may be uniquely identified by using the optional XML attribute 'id' in the element INSDQualifier. The value of the 'id' attribute must start with the letter 'q' and continue with any positive integer. 'id' values do not need to be numbered consecutively but an 'id' attribute must be unique to*~~

~~one INSDQualifier element, i.e., the attribute value must only be used once in a sequence listing file.~~;

- ~~(e)~~(g) Update paragraph 77 to include a space and a " between 'organism' and 'for' in the first line: '*...and "organism" for amino acid sequences...*';
- (h) Update paragraph 79, example 2: remove the space before 'mol_type';
- ~~(f)~~(i) Update paragraph 83, in the example, to remove the space before 'mol_type';
- ~~(e)~~(i) Update paragraph 86 to read: '*The use of free text must be limited to a few short terms indispensable for the understanding of a characteristic of the sequence. For each qualifier **other than the "translation" qualifier**, the free text must not exceed 1000 characters.*';
- ~~(h)~~(k) Update the start of paragraph 87(d) to '*For qualifiers with a language-dependent free text value, ...*'; ~~and~~
- (l) Update all instances of '*industrial property*' to '*intellectual property*', after the update at the ninth session of the CWS to increase the scope of the WIPO Handbook was approved; and
- ~~(f)~~(m) Update all references to the 'international, national or regional procedures' such that it appears in this particular order throughout the Main Body.

8. The proposed changes to the Annexes of WIPO Standard ST.26 can be summarized as follows:

- (a) Annex I, Section 6: update Table 5 to include frequency as a qualifier with language dependent free text value;
- (b) Annex I, Section 6.21: update the mandatory value format to indicate that the frequency qualifier is language dependent;
- (c) Annex I, Section 6: update the title of Table 5 to '*List of qualifiers ~~values~~ for nucleotide sequences with language dependent free text values*';
- (d) Annex I, Sections 6 and 8: update the note to include '*international*', which currently only refers to '*national or regional procedures*';
- (e) Annex I, Section 6.77: strip the spaces in the example after the anticodon value but before the less-than symbol;
- (f) Annex I, Section 8: update the note to be in line with the note provided in Section 6. The updated note will commence with: '*Any qualifier value provided for a qualifier with a language dependent "free text" value format may require translation...*';
- (g) Annex I, Section 6 and 8: all qualifiers with a mandatory value format which indicates they are language dependent now indicates that translation should be for: '*International/National/Regional procedures*' by adding '*International*';
- (h) Annex I, Section 6 and 8: note at the top of each section edited so that '*language dependent "free text"*' is corrected to '*language-dependent free text*' for consistency with the rest of the Standard;
- (f)(i) Annex I, Section 6 and 8: correct the title of the second column replacing 'value' with 'qualifer' to read '*Language-Dependent Free Text Qualifier*'
- (g)(i) Annex II: include the following comment within the DTD to provide clarification on the value which should be provided within the `fileName` attribute: '*By default this will be set to the value provided for the project name in WIPO Sequence. If the value is identical to the actual ST.26 XML filename, it should be noted that Offices may enforce their requirements for the filename used which may restrict which characters are allowable for submitted electronic files. It is also acceptable for the value of the filename attribute and the actual file name to be different. Please refer to the WIPO Sequence and ST.26 Knowledge Base for further details on Offices' naming conventions for electronic files*';
- (h)(k) Annex VI, Example 3(c)-1, Question 3: correct the term 'acylation' to be 'acetylation';
- (l) Annex VI, Example 3(k)-2, Question 3: correct the reference to SEQ ID NO. 12 in the third paragraph to: '*the feature key "misc_difference" with feature location "10" should be used together with two "replace" qualifiers where the value for one would be "gk" and the second would be "c"*';
- (m) Annex VI, Example 7(b)-2: remove the first of the two figures;
- (n) A-;
- (i)(o) Annex VI, Example 28: correct SEQ ID 46 to correspond to the residues provided in the corresponding sequence in the Appendix to Annex VI;
- (j)(p) Annex VI, Example 7(b)-4 and 7(b)-5: correct SEQ ID NO. 89, SEQ ID NO.90 and NO. 91 to SEQ ID NO. 90, SEQ ID NO. 91 and SEQ ID NO. 92;

- (k)(g) Annex VI and Appendix to Annex VI: include a series of three new examples, Example 94-2, Example 30-2 and Example 12-1. The new Examples are provided in full as Annex I to the present document;
- (h)(r) Annex VII, Scenario 4: replace the first sentence with the following text, '*For both nucleotide sequences and amino acid sequences, ST.26 has the mandatory feature key "source" with two mandatory qualifiers*:';
- (m)(s) Annex VII, Scenario 4: replace the table caption for the Nucleotide sequences table with: '*ST.26 - feature key 5.37 source; mandatory qualifier 6.39 mol_type (see ST.26 paragraph 75)*';
- (n)(t) Annex VII, Scenario 4: replace the table caption for the Amino acid sequences table with '*ST.26 - feature key 7.30 source; mandatory qualifier 8.1 mol_type (see ST.26 paragraph 75)*';
- (u) Annex VII, Scenario 8: update row 18 of the included table to add in the fourth column the following text: '**TATA signal and (if <223> present)**:', and also remove from the note the following text: '(if <223> present)';
- (v) Annex VII, Scenario 8: replace the second sentence of the footnote to read: *In such a case, in order to avoid addition of subject-matter that may lead to partial loss of priority, it is recommended to include the more limited term "TATA signal" in a "note" qualifier as shown in the above table (item N° 18). If in rare cases the Applicant considers that the use of the "TATA box" value for the "regulatory class" qualifier is not appropriate, the value:"other" may be used instead of "TATA box". In this case, the term "TATA signal" must be included in a "note" qualifier associated to the "regulatory" feature key;*
- and
- (p)(w) Appendix to Annex VI: any corresponding changes required due to updates to Annex VI. The Appendix to Annex VI is provided as Annex II to the present document.

GUIDELINES ON UPDATING ST.26 VERSION

9. As indicated above, the International Bureau noted, at the ninth session of the CWS, the concerns regarding regular updates to WIPO ST.26 and the impact that this may cause (see paragraph 4 of the present document).
10. It should be noted that WIPO ST.26 will only be updated at most once a year, as required and as approved by the CWS. In the proposed revision presented at CWS/10, there are both editorial and substantive amendments. Editorial changes made in the Main Body should have no substantive impact on users including IPOs. Substantive changes to Annexes VI and VII should also have no impact on IPOs and users.
11. The International Bureau will, in future versions of the WIPO ST.26, provide an appropriate version number, X.Y, which will indicate to users of the Standard whether this revision is considered to be backwards compatible with the previous version. Minor updates to the WIPO ST.26 are indicated by an increment to the second digit only ('Y') and these are considered backwards compatible. Major updates, which are indicated by an increment to the first digit ('X') will likely have an impact on your IT systems and may require re-tooling. The version number of the DTD may be different to that of the Main Body and other Annexes.
12. As there are no changes proposed to the WIPO ST.26 DTD except adding a comment in this new version and the changes proposed to the Main Body and Annexes are mostly editorial in nature or for the purposes of clarification, the version number for the DTD in this instance will remain as version 1.3 and the version number of the Standard will be updated to version 1.6.

13. If the proposed revision is approved by the CWS, in collaboration with the SEQL Task Force, the International Bureau will incorporate the approved contents into WIPO Standard ST.26 and publish it in 2022.

14. *The CWS is invited to:*

(a) note the content of this document and Annexes to this document; and

(b) consider and approve the proposed revisions to WIPO Standard ST. 26 as referred to in paragraphs 7 and 8 above and in the Annexes I and II to this document.

[Annex I (the proposed new Examples for Annex VI) follows]

PROPOSED NEW EXAMPLES FOR ANNEX VI OF STANDARD ST.26

Paragraph 94 – Variant sequence disclosed as a single sequence with enumerated alternative residues

Example 94-2 – Representation of single sequence with enumerated alternative amino acids that may be modified amino acids

A patent application describes the following polypeptide:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Xaa-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

where Xaa can be Ile, Ala, Phe, Tyr, alle, Melle, or Nle.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide provides 16 specifically defined amino acids. Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph (7)(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

The most restrictive ambiguity symbol that can encompass “Ile, Ala, Phe, Tyr, alle, Melle, or Nle” is “X”. Therefore, the sequence must be included in a sequence listing as:

LEYCLKRWXETISHCAW (SEQ ID NO: xx)

ST.26 paragraph 30 requires that “[a] modified amino acid must be further described in the feature table”. However, paragraph 30 does not require any specific feature key be used to describe modified amino acids. While paragraph 30 describes the use of feature keys “CARBOHYD”, “LIPID”, “MOD_RES”, and “SITE”, these feature keys are more appropriate for scenarios where the modified amino acid is not within a list of alternatives for a specific location. In this example, the feature key “VARIANT” satisfies the requirement of paragraph 30 since it allows for the inclusion of all of the alternatives for the variant site. So, the feature key “VARIANT” with the qualifier “note” ~~and a qualifier value~~ “Ile, Ala, Phe, Tyr, alle, Melle, or Nle” as a qualifier value should be used to describe the variant site at position 9. The use of a second feature key such as “SITE” with a qualifier “note” may be used to further identify the modified amino acids found at position 9.

Relevant ST.26 paragraph(s): 3(a), 7(b), 27, 30, **94**, 96, and Annex I, Section 4, Table 4

Paragraph 30 – Annotation of a modified amino acid

Example 30-2 – Post-translationally modified amino acids

A patent application describes the following polypeptide:

Leu-Glu-Tyr-Cys-Leu-Lys-**Arg**-Trp-Glu-Thr-Ile-Ser-His

wherein the Arg at position 7 may be post-translationally deiminated to citrulline.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide provides 13 specifically defined amino acids. Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph (7)(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

According to ST.26 paragraph 29, a modified amino acid should be represented in the sequence as the corresponding unmodified amino acid whenever possible.

Therefore, the sequence should be included in a sequence listing as:

LEYCLKRWETISH (SEQ ID NO: xx)

where the symbol “R” is used to represent the arginine at position 7.

A further description indicating that the arginine at position 7 may be modified to citrulline is required. The modification of arginine to citrulline is a post-translational modification. Therefore, the feature key “MOD_RES” should be used together with the mandatory qualifier “note” to indicate that the arginine may be deiminated to form citrulline. The location descriptor in the feature location element is the residue position number of the modified arginine.

Relevant ST.26 paragraph(s): 3(a), 7(b), **30**, and Annex I, Section 7, Feature Key 7.18

Paragraph 12 – Circular nucleotide sequence

Example 12-1: Circular nucleotide sequence

A patent application contains the following figure, disclosing the DNA sequence of plasmid pCIRC1:



Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated nucleotide sequence has more than 10 specifically defined nucleotides. Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph (7)(a).

Question 3: How should the sequence(s) be represented in the sequence listing?

According to ST.26 paragraph 12, when nucleotide sequences are circular in configuration, the applicant must choose the nucleotide in residue position number 1. For the purposes of this example, the “a” residue identified by the arrow in the figure will be used as position 1. However, any residue may be chosen as position 1. With the residue indicated by the arrow as position 1, the sequence should be included in a sequence listing as:

atggataatgaagaagttaacgaagaatgtatgagattat
aaacatctaacatcaagggtgacatgcatgaaatgaaatgcatatatcacgttcagatgcttctg
gatggaatacatcgcaaactactaggttctcgaagagctactttgaaacaagaaaatgtaccac
caatg (SEQ ID NO: xx)

The sequence should be further described using feature key “misc_feature” with a location of “212^1”, which indicates that the last residue in the sequence, position 212, is linked to residue 1. A “note” qualifier must be included with a value indicating that the molecule is circular.

Relevant ST.26 paragraphs: 7(a), 12, and Annex I, Section 5, Feature Key 5.15

[Annex II (Appendix to Annex VI) follows]