

Patent Cooperation Treaty (PCT) Working Group

Eighteenth Session
Geneva, February 18 to 20, 2025

PROCESSING SEQUENCE LISTINGS

Document prepared by the International Bureau

SUMMARY

1. WIPO Standard ST.26 is the required format for nucleotide and amino acid sequence listings in international applications filed on or after July 1, 2022. Following the successful rollout and transition, the system operates largely efficiently and as expected, and work continues on the further development and enhancement of tools and procedures for processing of sequence listings in this format.
2. The International Bureau nevertheless wishes to simplify processing and ensure that the results are high quality, support the requirements of patent applicants and are suitable for use by International Authorities, designated Offices and database providers. For instance, it may be possible to reduce the number of document types. Analysis continues regarding options for improved validation arrangements and to identify whether language requirements are met effectively. The Sequence Listing Task Force of the Committee on WIPO Standards (CWS) is considering substantive changes to the Standard itself to expand disclosure requirements and include short sequences.

BACKGROUND

3. WIPO Standard ST.26 is the Standard which recommends an XML format for the presentation of nucleotide and amino acid sequences in patent applications, and it entered into force on July 1, 2022. Amendments to the PCT Regulations and modifications to the Administrative Instructions, including relevant forms, also entered into force at the same time, requiring sequence disclosures to be presented in compliance with WIPO Standard ST.26 in any international application filed on or after that date. This document provides updates on the

recent developments and presents certain challenges and possible ways to improve the current practices for processing of sequence listings.

WIPO SEQUENCE SUITE

4. The International Bureau has developed a software suite known as WIPO Sequence Suite which is comprised of the following two components:

(a) WIPO Sequence: a standalone desktop application, available for Windows, Linux and MacOS, to assist applicants to author and validate sequence listings in compliance with WIPO Standard ST.26; and

(b) WIPO Sequence Validator: a web service that runs internally at patent Office environments to check filed sequence listings for compliance with WIPO Standard ST.26.

5. Development and testing of the WIPO Sequence Suite is prioritized and undertaken in coordination with the Sequence Listings Task Force.

6. The latest version 3.0.0 of WIPO Sequence Validator was released in October 2024, offering significantly improved performance over previous versions and removing outdated dependencies. The operations manual was also made available in HTML format in 10 languages, offering a more accessible and streamlined resource for Offices, supporting the downloading of code segments to help get started more quickly and reliably. Since this release, Offices have reported some bugs which are currently being resolved. A patch version, 3.0.1, is scheduled to be released in February 2025.

7. WIPO Sequence version 3.0.0 is still being tested. Similarly, deprecated components have been upgraded and the WIPO Sequence Validator has now been incorporated ensuring 100% consistency in validation for applicants and their representatives.

DEVELOPMENT OF WIPO STANDARD ST.26

8. The latest version of WIPO Standard ST.26, version 1.7, came into force on July 1, 2024 and should be used for all international, national and regional applications filed on or after that date. The changes compared to version 1.6 are essentially clarifications and the addition of several new examples in the guidance document. The Sequence Listings Task Force is currently considering two substantive revisions for consideration by the CWS at its thirteenth session, to be held in November 2025. The first, proposed by the European Patent Office, would be to lift the minimum length requirement which does not allow short sequences to be included within a sequence listing. The other, proposed by the United States Patent and Trademark Office, would require certain peptide and nucleotide analogs to be included in a sequence listing. If these substantive changes are approved, the new version of the WIPO Standard ST.26 will be version 2.0. The first of these changes will also require significant updates to the WIPO Sequence Suite before the entry into force date of the new version of the Standard.

DATA PACKAGE FORMAT FOR THE ELECTRONIC EXCHANGE OF PRIORITY DOCUMENTS

9. At its twelfth session, held from September 16 to 19, 2024, the CWS adopted new WIPO Standard ST.92 on the data package format for the electronic exchange of priority documents. This Standard aims to improve the processing of priority documents by allowing the electronic exchange of structured text formats and especially sequence listings in the XML format of WIPO Standard ST.26, instead of only PDF files. It was tentatively agreed to aim for implementation of the new Standard by July 1, 2027. This will involve IT development work by the International

Bureau for PCT systems and the WIPO Digital Access Service for Priority Documents (DAS), as well as by national Offices.

PROCESSING OF SEQUENCE LISTINGS UNDER THE PCT

10. Since July 2022, the International Bureau has received approximately 32,000 international applications with a sequence listing. For most, the processing was simple and in line with expectations, but a significant number included issues that would need further input from staff.

Sequence Listing Validation

11. The WIPO Sequence Validator provides two types of validation: “formality” and “full”. The formality validation checks that the file format is in XML format and that it is compliant with the WIPO Standard ST.26 Document Type Definition (DTD). The full validation commences with the formality validation check and then checks compliance with a series of verification rules derived from the body of WIPO Standard ST.26. Even with the improved performance of the new version, the WIPO Sequence Validator operations manual recommends using full validation only as part of a batch process since it takes much longer than the formality validation.

12. Sequence listings prepared using WIPO Sequence should not contain errors identifiable using WIPO Sequence Validator (though there may be warnings) since the same validation rules are used when the sequence listing is generated by the applicant. However, a significant number of sequence listings have been received where it is apparent that, although they are indicated as having been created by WIPO Sequence, some manual post-processing using a text editor has been performed without the result being validated by WIPO Sequence, to ensure it still complies with WIPO Standard ST.26.

13. At present, the International Bureau’s systems do not use the full validation check conducted by the WIPO Sequence Validator either when sequence listings are uploaded into ePCT or when they are received from receiving Offices and International Authorities through PCT-EDI or other forms of transmission; these need to be performed by national Offices where they need such detail. Furthermore, the current PCT Receiving Office Guidelines and Administrative Instructions (Annex C) do not require either the receiving Office or the International Bureau to validate the sequence listings. However, the International Bureau is evaluating version 3.0.0 of WIPO Sequence Validator to determine whether the full validation can be made available through ePCT and PATENTSCOPE, either as the result of a routine action performed on receipt of a sequence listing, or else on demand.

Sequence Listings under Rule 13ter

14. The use of Rule 13ter to request sequence listings for the purpose of international search has decreased following the introduction of WIPO Standard ST.26, but not as dramatically as had been expected. It had been hoped that international applications disclosing sequences requiring a sequence listing would all contain a standard-compliant sequence listing, validated prior to submission and considered suitable for international search. However, around 14 per cent of international applications with a sequence listing include a submission under Rule 13ter. This includes:

- (a) cases where the sequence listing included in the international application as filed was presumably found to be defective, requiring a standard-compliant listing for search; and
- (b) cases where no listing had originally been submitted, but the international application disclosed sequences that should have been included in a listing.

15. Table 1 below presents a snapshot taken in late December 2024 of the number of international applications filed on or after July 1, 2022, for which the International Bureau had received a sequence listing of some type, broken down by receiving Office. The second to fourth columns show:

- (a) the number of international applications with any type of sequence listing on file;
- (b) the number of international applications including a Rule 13*ter* sequence listing in addition to a listing that had been included in the international application as filed; and
- (c) the number of international applications including a Rule 13*ter* sequence listing where the international application as filed contained no sequence listing.

Table 1: Number of International Applications with Sequence Listings by Receiving Office

| Receiving Office | Total with listings | Rule 13<i>ter</i> in addition to listing as filed | Rule 13<i>ter</i> submitted later only |
|-------------------------|----------------------------|--|---|
| AU | 289 | | 13 |
| BR | 46 | | 1 |
| CA | 281 | 5 | 33 |
| CL | 6 | | |
| CN | 6,795 | 7 | 83 |
| CU | 12 | 9 | |
| CZ | 28 | | 4 |
| DE | 10 | | |
| DK | 7 | | 1 |
| EP | 4,722 | 94 | 289 |
| ES | 60 | | 3 |
| FI | 30 | 1 | |
| FR | 46 | 2 | 1 |
| GB | 558 | 18 | 73 |
| GR | 4 | 1 | |
| HU | 13 | | 2 |
| IB | 1,235 | 66 | 129 |
| ID | 1 | 1 | |
| IL | 313 | 39 | 3 |
| IN | 89 | 4 | 10 |
| IT | 18 | 1 | 3 |
| JP | 1,882 | 1 | |
| KR | 2,237 | 1 | |
| MX | 3 | 1 | |
| MY | 5 | | |
| NL | 89 | | 46 |
| NO | 3 | | 1 |
| NZ | 23 | | 5 |
| PE | 1 | | |
| PH | 2 | | |
| PL | 43 | 6 | 6 |
| PT | 5 | 1 | |

| Receiving Office | Total with listings | Rule 13ter in addition to listing as filed | Rule 13ter submitted later only |
|------------------|---------------------|--|---------------------------------|
| QA | 1 | | |
| RU | 44 | 3 | 1 |
| SE | 26 | 1 | 1 |
| SG | 190 | 1 | 7 |
| SK | 2 | | 1 |
| TH | 11 | 2 | 4 |
| TR | 43 | 4 | 10 |
| US | 12,179 | 991 | 2,336 |
| ZA | 1 | | |

16. The reasons for Rule 13ter submissions require further investigation, but the variations are a factor of the International Searching Authority as well as the origin of the listing. It is more common in some States than others for applicants to prepare international applications containing sequences that are required by Rule 5.2(a) to be included in a sequence listing without furnishing such a listing as part of the international application. Some International Authorities check the sequence listings filed more thoroughly than others and are more likely to require a standard-compliant listing under Rule 13ter for the purpose of international search if a defect is found in a listing or if the application body contains sequences that are not included in a listing.

Manual Checks Conducted During Processing of Sequence Listings

17. The European Patent Office and United States Patent and Trademark Office have supplied details to the Sequence Listing Task Force of checks that they carry out on listings beyond the checks performed automatically by WIPO Sequence and WIPO Sequence Validator, which trigger many of the requests for Rule 13ter listings. Information on these checks may be valuable to other national Offices to improve their processing of listing files received at their Office.

18. Further analysis is required of the common types of defects and their consequences for international search to determine what action, if any, should be considered in relation to:

- (a) education of applicants about the requirement to include a sequence listing as part of the description and the requirements of WIPO Standard ST.26;
- (b) improvements needed for the WIPO Sequence Suite;
- (c) the validation to be undertaken by receiving Offices, International Authorities and the International Bureau;
- (d) guidance on when it would be useful to request a listing for the purpose of international search (noting that the invitation under Rule 13ter is at the discretion of the Authority and a Rule 13ter listing does not become part of the international application and consequently the formal defect may still need to be corrected in Chapter II or the national phase).

Document types related to Sequence Listings

19. Table 2 below shows different types of document codes related to sequence listings currently available and the corresponding number of usages of each document type for applications filed since July 2022. This list may not be complete for the PCT in general since the International Bureau does not receive all the relevant documents (particularly under

Rule 13ter), but it should be a good indication. It is not easy for applicants or formalities examiners to understand and select the correct document types for the different special cases. Human error may occur due to confusion between multiple similar document types or due to mistakes such as accidentally clicking on the incorrect document type on the dropdown list. Further consideration should be given to whether some of these document types should be eliminated. For example, the document types “Seq List 13ter Corr.,” “Seq List 13ter Rectif.” and “Seq List 13ter rectified (IPEA)” could be considered as unnecessary since it is not part of the international application as such, if a Rule 13ter listing is wrong for any reason, it could simply be replaced with a new one rather than specifying whether the change is due to a correction or a rectification.

Table 2: Use of Sequence Listings Document Types

| Document type | Number |
|---|--------|
| Seq. List. | 28,804 |
| Seq List 13ter | 3,550 |
| Seq List statement 13ter | 3,319 |
| Seq List originally filed-Replaced (R.12, 26, 91) | 253 |
| Seq. List. Info | 213 |
| Seq List Corr. | 128 |
| Seq List (ST.26 non-compliant) | 65 |
| Seq List Rectif. | 63 |
| Seq List 13ter Rectif. | 50 |
| Seq List Trans for pub | 20 |
| Seq. List. Incorpor. By Ref. (Rule 20.6) | 19 |
| Seq List Trans for search | 11 |
| Seq List Validation Report | 8 |
| Sequence Listing - Amendment - Article 34 | 6 |
| IPRP II Seq List | 5 |
| Seq List 13ter Corr. | 3 |
| Seq. list - later submitted (R20.5bis) | 2 |
| Seq. List - later submitted (R20.5) | 2 |
| Certified Copy Seq List Placeholder Page | 1 |
| Seq List rectified (IPEA) | 1 |
| Seq List 13ter rectified (IPEA) | 0 |
| SIS Seq List Trans | 0 |

20. Currently some receiving Offices check the bibliographic data in the sequence listings and invite the applicant to submit corrections under Rule 26. Since July 2022, the International Bureau received 128 sequence listings corrected due to the defects in applicant’s name, priority date, invention title etc. In order to reduce the number of such type of corrections, the bibliographic data in the sequence listings could be either simplified or not necessarily checked by the receiving Offices.

TRANSLATION OF SEQUENCE LISTINGS

21. WIPO Standard ST.26 requires an XML file to be provided in two parts: a general information part and a sequence data part.

Language Elements in General Information Part

22. The general information part of the sequence listing allows information to be provided concerning the application number, filing date, applicant file reference, earliest priority application, applicant name, inventor name, and invention title. The applicant, inventor names and invention title may be supplied in multiple languages or character sets. The information supplied in this information part is intended to help confirm that the sequence listing has been attached to the correct international application. If it differs from the information in the request form or is provided in a different language, the receiving Office may draw this to the applicant's attention, but the Office cannot require the applicant to correct it and the application should proceed on the basis of the information in the request form (paragraph 28 of Annex C of the Administrative Instructions).

Language-Dependent Free Text in Sequence Data Part

23. The sequence data part of the listing is based on an industry standard used by database suppliers and includes elements to contain free text ("qualifier values"). The industry standard expects these values either to be language-neutral, or else to be written in English. While some national Offices indicated that they would be able to accept sequence listings with language-dependent free text in English even though the language of the description may differ, many Offices indicated that it was essential that their applicants be either permitted or required to supply language-dependent free text in the language of the description. Consequently, WIPO Standard ST.26 requires qualifiers expecting language-dependent values to have one or both of the following elements:

- (a) `INSDQualifier_value` – the qualifier value English; and
- (b) `NonEnglishQualifier_value` – the qualifier value in another language, as indicated by an attribute `nonEnglishFreeTextLanguageCode` in the root element of the listing.

24. Under Rule 12.1(d), receiving Offices may specify the languages that they accept for the language-dependent free text qualifier values, which may differ from what they accept for the description in general. Furthermore, they may permit the applicant to file the sequence listing with the language-dependent free text in more than one language, but cannot require them to do so. Technically, the Standard permits only English and one other language within the sequence listing.

25. This flexibility was considered important to ensure that information supplied to databases is in the industry database format (where the industry standard specifically expects English text in the relevant fields) and for the applicant to be able to meet the potentially different language requirements of Offices for the purposes of filing, international search and various national phase entries. Allowing English and a second language simultaneously overcomes the need to supply separate listings for the most common case of different languages for filing and international search.

26. A project created in WIPO Sequence allows multiple language versions to be provided and for a sequence listing to be generated with the desired language on demand. Consequently, it should be possible to prepare a project and easily export appropriate language versions depending on filing requirements. For example, a sequence listing might be generated for the international application with language-dependent free text in Japanese and English; then, on national phase entry, substantively identical listings might be generated for different designated Offices containing the text in English only, Chinese and English, and Korean.

27. Within WIPO Sequence, the different language values can also be exported to an XLIFF file, a standard format used by translators, indicating the source and target languages. This file can be read into standard translation software, the appropriate values filled in and the result imported into WIPO Sequence, rather than manually adding the target language values one-by-one. It is not clear that these features are well understood or used by applicants. It is also noted that sequence listings contain ID attributes for language dependent free-text qualifiers allowing for different language values to be matched when importing XLIFF files. However there have been several problems relating to these attributes that have caused errors. Consideration is being given to whether these ID attributes should be kept or obscured from the user even after generation of the sequence listing.

28. WIPO Sequence does not attempt to validate whether the appropriate language has been specified within a project. The only check performed is that the text in the `INSDQualifier_value` complies with paragraph 40(b) of the Standard and contains only characters from the basic (unaccented) Latin character set.

29. Given that it is possible to submit sequence listings containing free text in two languages (English and one other), and that examiners at International Searching Authorities are mostly proficient in the language of filing, whether or not that is the language of publication, it is perhaps not surprising that translations of sequence listings are rarely submitted during the international phase. Since July 2022, the International Bureau has received just 31 documents indicated as sequence listing translations for publication or for search purpose. However, inspection of samples of listings suggest that the language options are not always being used as intended. In some cases (including documents provided as corrections, as referred to in paragraph 20, above), the translation of the sequence listing was provided simply to give the relevant language in the general information part, which should not be required. Furthermore, it is not clear that the contents of qualifier tags are always in the correct languages and, where it is indicated that both English and a second language is being used, that the qualifier texts are equivalent. The International Bureau hopes to perform a more systematic analysis of the use of language options within sequence listings during the international phase in order to determine whether action should be taken of types equivalent to those referred to in paragraph 18, above. Information would be welcome from designated Offices on the use of translations of sequence listings in WIPO Standard ST.26 format on national phase entry.

PUBLISHING SEQUENCE LISTINGS

30. PATENTSCOPE has for quite some time included a facility to view sequence listings in a human-friendly format similar to the display format of WIPO Standard ST.26. An updated version of PATENTSCOPE was deployed in November 2024 with a new display engine to better facilitate the viewing of very large sequence listings. The transformation is streamed rather than having to load the whole document in order to process it. Whereas before, large listings might take a minute or more to display, or fail entirely, the start of the listing will now appear almost instantaneously and, if not displayed completely, additional parts can be loaded on demand.

31. The Working Group is invited to note the contents of document PCT/WG/18/14 and provide any comments for further review and analysis of processing of sequence listings under the PCT.

[End of document]