

## *Check Yo'self Before You Wreck Yo'self (or Your Sequence Listing)*

Author: [Harbor Consulting IP Services, Inc.](#) Team Member

The USPTO provides two programs that are necessary for Sequence Listing preparation: [PatentIn](#) and [Checker](#). For those unfamiliar with these tools, the PatentIn software generates a sequence listing which is subsequently run through the Checker software to search for errors. To ensure that a sequence listing is fully compliant under 37 C.F.R. §§ [1.821-1.825/WIPO ST.25](#) sequence listing rules, it is important to understand exactly what the Checker software is “checking,” as well as the limitations of this program.

When verifying a sequence listing, what does the Checker software search for, and what are considered errors? There are numerous and very specific errors the software finds, and they primarily fall within the following three categories:

1. Sequence numbering errors. Examples are as follows:
  - displaying an incorrect total number of sequences when compared to the actual number of sequences disclosed in the sequence listing;
  - incorrect nucleotide/residue numbering within a sequence; and
  - displaying an incorrect total number of nucleotides/residues in a sequence header when compared to the total number literally disclosed in a single sequence.
2. Formatting errors. Examples are as follows:
  - omission of a required sequence header such as the <210> sequence number filed;
  - inclusion of non-ASCII characters; and
  - incorrect grouping of nucleotides in a DNA sequence.
3. Sequence data errors. Examples are as follows:
  - omission of a feature key for an undefined nucleotide/residue (this check is fairly limited as it does not confirm the correct number or position of feature keys); and
  - inclusion of non-standard nomenclature for amino acids or nucleotides within a sequence.

The Checker software is an extremely useful tool for sequence listing preparers; however, it is important to be aware that the software does have limitations. In a perfect world, the Checker software would catch every single error that could result in a rejection. Unfortunately, problems that are more substantive in nature will not be considered errors by the Checker software. These types of issues include, but are not limited to, the inclusion of incorrect sequences or organism designations, the misspelling of organism names, incorrect modification positions, nucleotide sequences shown coded through the stop codons, and non-compliant modification definitions.

We often see error reports for sequence listings that successfully ran through the Checker software but still end up with a rejection later on when validated by the USPTO as a result of the limitations in the Checker software. The validation software utilized by the USPTO is essentially a "supped up" version of the Checker software and will frequently detect issues that the Checker software does not indicate as errors. Additionally, further substantive review of the application during prosecution may result in additional sequence listing rejections, usually the result of discrepancies between the sequence listing and sequences disclosed in the application as filed.

At some point in the near future, WIPO ST.25 will be replaced with the new standard, [WIPO ST.26](#). A major change from ST.25 to ST.26 is the format of sequence listings from an ASCII text-based file to an extensible mark-up, language-based ("XML") file. XML is a computer code language and is considered more "computer friendly" in terms of database upload and analysis.

The manner in which this transition from ASCII to XML will be dealt with by the requisite patent offices has not been finalized and as the current Checker software evaluates sequence listings prepared under 37 C.F.R. §§ 1.821-.825/WIPO ST.25, there are some questions to consider, including both: (1) will the software equivalent to the PatentIn and Checker software be created to support this format change? and (2) will there be a software to convert ASCII format to XML format? While the answers to these questions are unknown, XML format will certainly complicate the preparation of sequence listings and increase the potential for introducing errors, which will likely result in additional time and cost incurred by applicants.

Absent a new version of the Checker software, XML format will make spotting errors prior to submission difficult for preparers. Alternatively, because XML format is more computer-friendly, this change may enable software developers to create a new, improved version to mitigate or possibly eliminate sequence listing errors altogether.

In summary, as there are limitations to the scope of errors the current Checker software is able to detect, it is best practice to avoid editing the sequence listing after preparation in the PatentIn software, as well as thoroughly reviewing the sequence listing prior to submission. Filing an erroneous sequence listing can result in rejections, higher fees, and prosecution delays. Further, if the erroneous sequence listing is not detected by the USPTO, the error may compromise the integrity of prior art searches.