From: James Coburn
Sent: Thursday, July 12, 2012 2:57 PM
To: seq_listing_xml
Subject: Comments related to proposed ST. 26

1. Is the main body of the standard comprehensive and clear?
   What's proposed is generally clear, but there are many issues, scenarios, possibilities, etc., that have not been touched upon. To name just a couple: (i) sequences associated with GenBank Accession Numbers/ Gene Name References are often require by various offices even though they are not disclosed as literal sequences; (ii) how to handle the penchant for national and regional offices to impose "internal directives" not accepted by other countries.

2. Does the XML standard include any unnecessary procedural requirements?
   The XML format will cause Sequence Listing files to be on the order of 1-3X the size of current Sequence listings files. Will the EFS portal(s) increase their byte size acceptance limit?

3. Are any feature keys or qualifiers not clear or that are optional but should be mandatory/vice versa?
   There are many feature keys/qualifier requirements that seem unnecessary and/or were not required for inclusion under ST. 25. To name just a couple: (i) why may some modifications be abbreviated, while other modifications must be described with the full modification name? (ii) What is the purpose of making the "Source" feature mandatory?

4. Definition of a Sequence for which a Sequence Listing is Required:
   a. Prohibited sequences:
      i. Does this mean that the **software will not accept** branched sequences and/or sequences fewer than 10 defined nucleotides or 4 defined amino acids, or will a listing be rejected upon filing?
      ii. What about claimed formula sequences wherein the sequence is defined as all "Xaa" residues, but an office mandates inclusion for search purposes?
   b. Modified nucleotides:
      i. It's proposed that sequences with any chemical moiety which provides the same structural function as a phosphate moiety of DNA or RNA be included in a listing. Does this mean that each individual chemical moiety in the sequence must be explicitly defined?
   c. D-amino acids:
      i. Is it a requirement to indicate in the listing which residues are D-amino acids?
      ii. Which are L-amino acids?
      iii. Or is no feature required to describe the configuration?
   d. Variants:
      i. Clarity regarding variant sequences. For example purposes: the sequence "RGSTDM" is has possible mutations of R1E, S3F and M6L
         1. Would the variants be encompassed by a single consensus, e.g., "X1 G X3 T D X6" wherein X1 is R or E, X3 is S or F and X6 is M or L? This disclosure would constitute that the sequence is now less than 4 defined amino acids, but would likely still be required for inclusion.
         2. Should each possible variant sequence be given a unique identifier? This would exponentially increase the amount of data in a given listing. This example contains only 3 possible

variants, but if you assign every possible combination, it would equate to the following 7 sequences.

i. eGSTDM
ii. RGfTDM
iii. RGSTDl
iv. eGfTDM
v. eGfTDl
vi. RGfTDl
vii. eGSTDl

If this scenario contained perhaps 20 variants, the combination of possible sequences would be extensive.

5. Publications (references):
No perceived detriment due to non-inclusion.

6. Transition issues:
   a. A one year grace period, wherein a Sequence Listing may be submitted under ST. 25 or ST. 26 format, would seem appropriate.
   b. National and/or regional compliance according to different standards would only be problematic insomuch as it would add additional time and cost to patent prosecution. Any authoring tools to assist applicants could only be seen as helpful.

General Comments:

- Will people be taught how to produce this new format? Will training sessions be offered?
- There will be many people who do not understand this format. How will the various offices handle the influx of calls/questions/correspondence? Will they hire on more staff? Will the reviewers and Examiners be adequately trained with regards to the new format/regulations?
- The idea of separate authoring tools has come up several times. Right now, the EPO has a tool (BiSSAP) to generate XML listings. The USPTO is also planning on developing a tool. If the point of the standard is to unify sequence listings then what is the point of having all of these separate tools? Won't uniformity decrease?
- The proposed new listing format is difficult to read. Is it necessary to have the information formatted in the exact manner shown or would it be acceptable to have spaces or lines in between to break up the heavy text?
- Due to time/cost, we perform the majority of our post processing outside of Patent-In. Is the patent office planning to make the new listing generation software more efficient to use?