E

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

PCT/WG/5/14 ADD.
ORIGINAL: ENGLISH
DATE: MAY 14, 2012

Patent Cooperation Treaty (PCT)
Working Group

Fifth Session
Geneva, May 29 to June 1, 2012

TASK FORCE ON SEQUENCE LISTINGS STANDARD

Status report by the European Patent Office

BACKGROUND

1. The Task Force on Sequence Listings was created by the Committee on WIPO Standards (CWS), at its first session (October 25-29, 2010), to deal with Task No. 44:

> "Prepare a recommendation on the presentation of nucleotide and amino acid sequence listings based on eXtensible Markup Language (XML) for adoption as a WIPO standard. The proposal of the new WIPO standard should be presented along with a report on the impact of the said standard on the current WIPO Standard ST.25, including the proposed necessary changes to Standard ST.25."

2. The Task Force was also requested:

> "To liaise with the appropriate PCT body with regard to the possible impact of such standard on Annex C to the Administrative Instructions under the PCT."

3. The EPO was assigned the role of Task Force Leader and has since then held five rounds of discussions on WIPO's wiki. The principle of differentiating the technical aspects of ST.25 from Annex C (PCT Administrative Instructions) was agreed upon at last PCT MIA 2011 and at the PCT Working Group 2011.

PROGRESS REPORT

4. The Task Force started operating in February 2011 on the basis of drafts prepared by the EPO. Many Offices participated in the process and posted useful comments on WIPO's related wiki. After more than one year of intensive discussions, the Task Force achieved substantial progress and the final draft of the main body and its various annexes have been posted on WIPO's wiki ready for use by Offices requiring public hearing.

5. To summarize, the differences with current ST.25 are the following:

- All (PCT) procedural issues are transferred to the PCT Administrative Instructions: the new standard would concentrate on technical aspects only to enable an optimal presentation of the sequence listings (the biotech-related part) and the appropriate format of the submission (namely XML).

- The biotech-related part has been considerably improved to reflect modern industry standards, for example:

    - inclusion of modified nucleic acids and amino acids not previously provided for (e.g. D-amino acids, PNA, morpholinos etc.) which have gained importance in industry and need to be electronically searchable
    - clear instructions for gapped sequences and sequence variants
    - clarification with regard to features and annotations
    - consistency with latest public biological sequence repositories consortia requirements (INSDC and UniProt)

- The XML definition will be self-contained and ST.26 would be dependent on neither ST.36 nor on ST.96 (XML4IP).

TRANSITION FROM ST.25 TO ST.26

6. At its 19th session, the Meeting of International Authorities (MIA) (see report, in document PCT/MIA/19/14, paragraph 87) agreed that:

(a) it would be preferable if the CWS Task Force, before concluding its work on the development of the new XML standard, would also look into the issue of whether it will be possible for any tool to be developed which would allow for the easy and complete conversion of sequence listings filed in one format (ST.25 Text or ST.26 XML) into the other;

(b) based on the conclusions reached by the Task Force on the issue of the feasibility of developing a conversion tool, the appropriate PCT bodies should commence a discussion on the most appropriate mechanism for transition from ST.25 to the new XML standard.

7. The EPO took over the task of assessing whether the conversion tool of BiSSAP from ST.25 standard to XML format, and vice versa, was reliable for the purpose of re-utilization of the converted sequence listing. As the draft ST.26 was only frozen late March, the work could unfortunately not start on time for a proper evaluation of the conversion tool by end May. At this stage only a very preliminary report could be produced (enclosed in Annex). Further work is ongoing with the aim of establishing a full report this Summer prior to the last consultation round on WIPO's wiki to initiate discussions with Task Force members.

8. In view of the above, the EPO suggests to stick to the position adopted at last MIA, that is the CWS Task Force will assess, on the basis of the report by the EPO, whether the conversion tool is reliable enough, and report back to the next PCT body to commence the discussion on the best transition mechanism.

9. Should the conversion tool prove to be reliable, the EPO would favor a sequential implementation in order for the Office and its applicants to benefit as soon as possible from the substantial improvements contained in the proposed new standard (e.g. inclusion of modified sequences, removal of ambiguities, more and better structured information, electronic file end-to-end, compatibility with other XML standards, flexibility). Also, delaying implementation by 3-5 years from now bears the risk of the standard being outdated.

NEXT STEPS BEFORE ADOPTION

10. A status report was presented at the 2nd session of the Committee on WIPO Standards (CWS/2). The Task Force work plan is now defined as follows:

- March-July 2012: Comments from users on the draft standard (public hearings)
- August: Consolidation of the draft standard by the EPO
- September - Nov.: Last round of consultations on WIPO's wiki
- Early 2013 Adoption of the draft standard at 3rd session of the Committee on WIPO Standards (CWS/3)

11. The Working Group is invited to take note of the status report set out in this document.

[Annex follows]

PRELIMINARY REPORT ON CONVERSION TOOL BETWEEN ST.25 AND ST.26

1. The new WIPO Standard ST.26 is far from limited to only defining a way of representing in XML the information which is currently provided in ST.25 format, but rather has taken the opportunity to modify and improve the Standard. The intended benefit of the proposed new ST.26 XML standard is its close alignment with worldwide standards for representing biological sequence information listings to ensure better interoperability.

2. Unsurprisingly, under those circumstances, a seaming less conversion between ST.25 and ST.26 can only occur with information provided by the user with a dedicated software during the conversion process.

3. In a nutshell, the structures of ST.25 and ST.26 are interchangeable. The biological content between both standards will need input from the user to move the data from one standard to the other in a limited number of situations. This assistance of the end user will be especially needed for annotations: the amount of keys and qualifiers within ST.26 reflects the type of possibilities available for the annotation of sequences within the scientific community and takes latest types of information needed into account.

Differences of elements of information between both standards:

General Information part	ST25	ST26	Convertibility
Application File Reference			
Application Number (Office Code + Number)			X
Application Filing Date			
Priority Application Number (Filing Office Code + Number)			X
Earliest priority date			
Applicant name			
Applicant Name Characters			
Inventor Name			
Inventor Name Characters			
Invention Title			
Invention Title Characters			
Number of sequences			

Sequence(s) part			
Sequence ID			
Sequence Length			
Sequence Type /molecule Type			
Sequence Division		C	
Publication information			X
Organism /sequence source			
Feature Key			
Feature Location			
Feature Qualifier			
Sequence			

Presence of elements

Mandatory		C in the box, indicates a constant value
Optional		
Absent		

Convertibity

Automatically		X indicates, a different structure between standards
With Manual Intervention		
Loss of information possible upon conversion		

[End of Annex and of document]