

## **Comité de Normas Técnicas de la OMPI (CWS)**

### **Primera sesión**

**Ginebra, 25 a 29 de octubre de 2010**

### **PROPUESTA DE PREPARACIÓN DE UNA NUEVA NORMA TÉCNICA DE LA OMPI RELATIVA A LA PRESENTACIÓN DE LISTAS DE SECUENCIAS DE NUCLEÓTIDOS Y AMINOÁCIDOS MEDIANTE EL LENGUAJE EXTENSIBLE DE MARCADO (XML)**

*Documento preparada por la Secretaría*

1. El 17 de septiembre de 2010, la Oficina Europea de Patentes (OEP) envió un documento a la Secretaría solicitando que el Comité de Normas Técnicas de la OMPI (CWS) emprenda, en su primera sesión de octubre de 2010, los debates sobre la propuesta de preparación de una nueva norma técnica relativa a la presentación de listas de secuencias de nucleótidos y aminoácidos mediante el lenguaje extensible de marcado (XML). La petición de la OEP figura en el Anexo del presente documento.
2. Con arreglo a la petición de la OEP, la Secretaría somete a examen y aprobación del CWS lo siguiente:
  - a) establecer la siguiente nueva tarea:

“Preparar una recomendación sobre la presentación de listas de secuencias de nucleótidos y aminoácidos en lenguaje extensible de marcado (XML) para su adopción como norma técnica de la OMPI. Además de la propuesta de una nueva norma técnica de la OMPI, debe presentarse un informe sobre el efecto de dicha norma en la actual Norma ST.25 de la OMPI, además de los cambios necesarios propuestos de dicha Norma ST.25”;
  - b) crear un nuevo equipo técnico, con su responsable correspondiente, para que se ocupe de la nueva tarea; y

- b) que el equipo técnico someta la propuesta de preparación de una nueva norma técnica de la OMPI y los cambios necesarios de la Norma ST.25 a examen y aprobación del CWS en su segunda sesión, que se celebrará en 2011.

3. *Se invita al CWS a:*

- a) *tomar nota de la petición de la Oficina Europea de Patentes, relativa a la preparación de una nueva norma técnica de la OMPI, que se recoge en el Anexo del presente documento;*
- b) *examinar y aprobar la propuesta relativa a la creación de la tarea y el calendario correspondiente referidos en los apartados a) y c) del párrafo 2; y*
- c) *examinar y aprobar la creación del nuevo equipo técnico, con su responsable correspondiente, referido en el apartado b) del párrafo 2.*

[Sigue el Anexo]

## PETICIÓN DE LA OEP DE CREACIÓN DE UN EQUIPO TÉCNICO DE LA OMPI

*Propuesta de la OEP sobre una nueva norma técnica relativa a la presentación de listas de secuencias en formato XML*

### ANTECEDENTES Y PETICIÓN

1. Las secuencias biológicas que se presentan en las solicitudes de patente figuran como "listas de secuencias". Actualmente, las listas de secuencias se presentan con arreglo a la Norma ST.25 de la OMPI, tanto en el marco del PCT (Anexo C de las Instrucciones Administrativas) como en el de la mayoría de los procedimientos nacionales y regionales de obtención de patentes.
2. La OEP opina que, por diversas razones técnicas y prácticas, la Norma ST.25 de la OMPI debería ser sustituida, o al menos complementada, por una nueva norma basada en el formato XML. Esa nueva norma remediaría las deficiencias de la Norma ST.25 de la OMPI y aportaría ventajas adicionales a los solicitantes y las oficinas, ya que la redacción y presentación de listas de secuencias de alta calidad favorecería la eficiencia de los procesos ulteriores.
3. La OEP tiene previsto presentar su propuesta sobre una nueva norma técnica de la OMPI a finales de 2010 para que sea aprobada en 2011. Así, pide por este medio pide que el Comité de Normas Técnicas de la OMPI (CWS) cree un equipo técnico específico que se encargue de elaborar la nueva norma basándose en la propuesta de la OEP. Para fundamentar su petición, facilita la información preliminar que se expone a continuación.

### PROBLEMAS QUE PLANTEA LA ACTUAL NORMA ST.25 DE LA OMPI

4. La Norma ST.25 de la OMPI se aprobó hace ya muchos años, si bien las biociencias son un ámbito técnico de rápida evolución. La Norma ST.25 de la OMPI constituye un formato específico del mundo de las patentes. Se trata de un formato escasamente conocido por los inventores que trabajan directamente con secuencias biológicas, quienes están generalmente más familiarizados con la presentación de secuencias en los formatos de los bancos públicos. Así, la Norma ST.25 de la OMPI no permite sacar partido de los programas de código abierto.
5. La Norma ST.25 de la OMPI no satisface los últimos requisitos científicos; por ejemplo:
  - Los proveedores de bases de datos públicas (Laboratorio Europeo de Biología Molecular (EMBL), DNA Databank of Japan (DDBJ), National Center for Biotechnology Information (NCBI)) tienen en cuenta más tipos de localización que la Norma ST.25 de la OMPI. Los solicitantes se ven obligados a buscar formas poco convencionales y muchas veces imprecisas para anotar secuencias.
  - Las claves y calificadores de caracterización (vocabulario controlado que se utiliza para representar las características de las secuencias) que se utilizan en la Norma ST.25 de la OMPI no son los que se utilizan en los bancos mundiales de secuencias.
  - Tampoco figuran algunas de las abreviaturas tipificadas por la Unión Internacional de Química Pura y Aplicada (IUPAC).

6. La Norma ST.25 de la OMPI favorece el error. Al estar diseñada de modo que pueda ser legible a la vez por personas y por computadoras, es fácil cometer errores, que además son difíciles de detectar. La relativa facilidad con que se puede elaborar una lista de secuencias mediante programas informáticos no especializados, como los procesadores de textos, es causa de muchos errores; por ejemplo:
- secuencias no válidas;
  - numeración errónea de las secuencias;
  - nombres de organismos no válidos;
  - caracteres no autorizados;
  - características erróneas;
  - sintaxis e información científica incorrectas debido a la complejidad de la llamada presentación mixta de secuencias biológicas.
7. En consecuencia, los proveedores de bases de datos públicas omiten los datos facilitados por los solicitantes y los vuelven a generar. Y las oficinas que proporcionan listas de secuencias a los bancos públicos tienen que dedicar mucho tiempo y esfuerzo a corregir los datos.

#### **LA NUEVA NORMA PROPUESTA**

8. La nueva norma que propone la OEP remediaría los inconvenientes que plantea la Norma ST.25 de la OMPI y contribuiría a mejorar la calidad de las secuencias de las bases de datos de patentes y las bases de datos públicas de secuencias. Cabe también señalar que la Norma ST.36 de la OMPI recomienda la utilización de las normas vigentes en la industria.<sup>1</sup>

La nueva norma tendrá las siguientes características:

- universalidad: un formato único para los especialistas en el ámbito de las patentes y para los demás usuarios;
- precisión y flexibilidad: un formato XML basado en un modelo de DTD convenido.

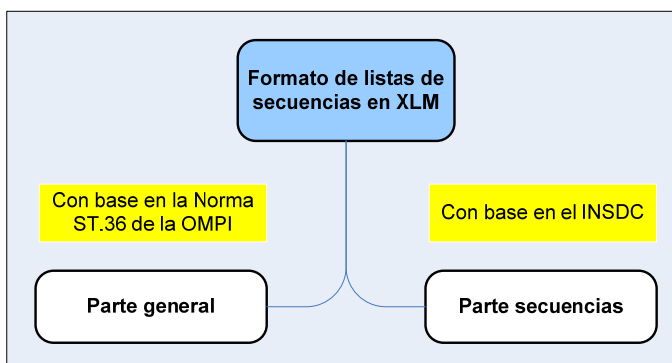
#### *Universalidad*

9. Una lista de secuencias se compone de dos partes: "parte de información general", esto es, información relacionada con la solicitud de patente, y "parte de secuencias", esto es, un número variable de secuencias biológicas.
10. Los proveedores de datos públicos, que integran la International Nucleotide Sequence Database Collaboration, INSDC (proyecto común entre el Instituto Europeo de Bioinformática (EBI), la DDBJ y el NCBI), han diseñado un formato XML para el intercambio de secuencias y de información conexa llamado INSDSeq. Dicho formato se elaboró para facilitar la adición de información pertinente relativa a patentes.

---

<sup>1</sup> Véase el párrafo 93 de la Norma ST.36 de la OMPI: Cuando sea adecuado al contenido de un documento, es decir, cuando el contenido no sea propio del dominio de la propiedad industrial, se utilizarán DTD de uso corriente en la industria.

11. Por esa razón, la nueva norma sobre presentación de listas de secuencias debería utilizar el formato INSDSeq para presentar información sobre secuencias y para modificarla con arreglo a la información relativa a patentes (información general de listas de secuencias) en virtud de la Norma ST.36 de la OMPI. Y ello a fin de asegurar la armonización de la parte de secuencias del formato de la lista de secuencias de la OMPI con el formato INSDSeq.



12. La nueva norma, basada en el formato XML, sería más fácil de aplicar, ya que los científicos utilizarían prácticamente un mismo formato en las solicitudes de patente y en la aportación de datos a las bases de datos públicas, sin necesidad de efectuar conversiones. La simple familiaridad con la sintaxis de la norma reduciría la posibilidad de error.

#### *Precisión y flexibilidad*

13. La sintaxis de la nueva norma que proporciona la DTD será a la vez más precisa y más fácil de comprobar mediante herramientas automáticas. Podría comprobarse mediante el amplio repertorio de programas gratuitos disponibles.
14. El contenido de un archivo XML, aunque de lectura más complicada para una persona sin hojas de estilo, es más accesible mediante computadora, y existe un gran número de bibliotecas al respecto.

#### **ESTADO DE LA CUESTIÓN**

15. Las Oficinas de la Cooperación Trilateral están discutiendo actualmente un borrador de DTD. El modelo de DTD que se acuerde será la base de la propuesta.
16. La OEP ha diseñado, en cooperación con el EBI, un programa cliente para la comunicación de secuencias biológicas (Biological Sequence Submission Application for Patents (BiSSAP)) que dispone de un módulo de comprobación y de una opción para crear ficheros XML. El programa es, así, compatible con la Norma ST.25 y los formatos INSDSeq / ST.36 XML. Además, fue probado en agosto por usuarios europeos con buenos resultados. Actualmente la OEP estudia la fecha de entrada en funcionamiento y puesta a disposición del público de esta nueva herramienta.

[Fin del Anexo y del documento]