F

OMPI
ORGANISATION MONDIALE
DE LA **PROPRIÉTÉ**
INTELLECTUELLE

CWS/1/5
ORIGINAL : ANGLAIS
DATE : 13 OCTOBRE 2010

Comité des normes de l'OMPI (CWS)

Première session
Genève, 25 – 29 octobre 2010

PROPOSITION CONCERNANT L'ÉLABORATION D'UNE NOUVELLE NORME RELATIVE À LA PRÉSENTATION DES LISTAGES DES SÉQUENCES DE NUCLÉOTIDES ET D'ACIDES AMINÉS EN LANGAGE XML (EXTENSIBLE MARKUP LANGUAGE)

Document établi par le Secrétariat

1. Le 17 septembre 2010, l'Office européen des brevets (OEB) a envoyé au Secrétariat document invitant le Comité des normes de l'OMPI (CWS), à sa première session prévue en octobre 2010, à examiner une proposition relative à une nouvelle norme concernant la présentation des listages des séquences de nucléotides et d'acides aminés en langage XML (eXtensible Markup Language). La demande de l'OEB est reproduite à l'annexe du présent document.

2. Conformément à la demande de l'OEB, le Secrétariat propose les points suivants pour examen et approbation par le CWS :

   a) la création d'une nouvelle tâche libellée comme suit :

      "Établir une recommandation concernant la représentation des listages des séquences de nucléotides et d'acides aminés en langage XML (eXtensible Markup Language) pour adoption en tant que norme de l'OMPI. La proposition relative à l'établissement de cette nouvelle norme de l'OMPI devrait être assortie d'une étude de l'incidence de ladite norme sur la norme ST.25 actuelle de l'OMPI, indiquant notamment les modifications à apporter à la norme ST.25";

   b) l'établissement d'une nouvelle équipe d'experts chargée de mener à bien cette nouvelle tâche et la désignation de son responsable; et

c) une demande invitant l’équipe d’experts à soumettre au CWS pour examen et approbation à sa deuxième session prévue en 2011 la proposition relative à la nouvelle norme de l’OMPI et aux modifications à apporter à la norme ST.25.

3. Le CWS est invité

a) à prendre note de la demande de l’Office européen des brevets reproduite à l’annexe du présent document, concernant l’établissement d’une nouvelle norme de l’OMPI;

b) à examiner et approuver la proposition concernant la création de la tâche et le calendrier correspondant indiqués au paragraphe 2.a) et c); et

c) à examiner et approuver la création de la nouvelle équipe d’experts, et la désignation de son responsable, ainsi qu’il est indiqué au paragraphe 2.b).

[L’annexe suit]

DEMANDE DE L'OEB EN FAVEUR DE L'ÉTABLISSEMENT D'UNE ÉQUIPE D'EXPERTS DE L'OMPI

Proposition de l'OEB relative à une nouvelle norme concernant la présentation des listages des séquences au format XML

GÉNÉRALITÉS

1. Les séquences biologiques divulguées dans les demandes de brevet sont présentées sous forme de "listages des séquences". Ces listages sont actuellement présentés conformément à la norme ST.25 de l'OMPI que ce soit dans le cadre du PCT (annexe C des instructions administratives) ou de la plupart des procédures nationales ou régionales en matière de brevets.

2. L'OEB considère que, pour différentes raisons techniques et pratiques, la norme ST.25 de l'OMPI devrait être remplacée, ou du moins complétée, par une nouvelle norme en XML. Cette nouvelle norme devrait pallier les inconvénients de la norme ST.25 de l'OMPI et offrir des avantages supplémentaires aux déposants et aux offices, dans la mesure où l'établissement et la présentation de listages des séquences de qualité est de nature à renforcer l'efficacité des procédures en aval.

3. L'OEB entend soumettre sa proposition relative à l'établissement d'une nouvelle norme de l'OMPI d'ici à la fin de 2010 pour adoption en 2011. L'OEB demande donc par la présente qu'une équipe d'experte spécifique soit créée par le Comité des normes de l'OMPI (CWS) en vue d'élaborer la nouvelle norme sur la base de la proposition de l'OEB. À l'appui de la présente demande, l'OEB souhaite fournir les informations préliminaires ci-après.

PROBLÈMES RENCONTRÉS AVEC LA NORME ST.25 ACTUELLE DE L'OMPI

4. La norme ST.25 de l'OMPI a été adoptée il y a des années, alors que les sciences de la vie sont un domaine où la technique évolue très rapidement. La norme ST.25 de l'OMPI définit un format propre au monde des brevets. Elle n'est pas aisément utilisable par les inventeurs travaillant directement sur des séquences biologiques, qui sont généralement plus familiers de la présentation des séquences selon les formats utilisés dans les répertoires publics. En particulier, la norme ST.25 de l'OMPI ne permet pas de tirer parti des logiciels libres.

5. La norme ST.25 ne répond pas aux derniers critères scientifiques, notamment :

   - Les fournisseurs de bases de données publiques (Laboratoire européen de biologie moléculaire (EMBL), DNA Databank of Japan (DDBJ), National Center for Biotechnology Information (NCBI)) prennent en charge davantage de types d'emplacements que la norme ST.25 de l'OMPI. Les déposants doivent trouver des solutions non conventionnelles et souvent imprécises pour annoter les séquences.

   - Les clés et qualificatifs de caractérisation (vocabulaires contrôlés utilisés pour décrire les caractéristiques des séquences) utilisés dans la norme ST.25 de l'OMPI ne sont pas utilisés dans les répertoires mondiaux de séquences.

   - Certaines des abréviations standard de l'Union internationale de chimie pure et appliquée (UICPA) ne sont pas non plus prises en charge.

6. La norme ST.25 de l'OMPI est propice aux erreurs. Dans la mesure où cette norme est censée être déchiffrable à la fois par l'être humain et par ordinateur, il est facile de faire des erreurs qui sont difficiles à détecter. La facilité relative avec laquelle il est possible de créer un listage des séquences au moyen d'un logiciel non spécialisé, tel qu'un traitement de texte, est à l'origine de nombreuses erreurs, notamment :

- séquences invalides;
- numérotation erronée des séquences;
- noms d'organismes invalides;
- caractères non autorisé;
- caractéristiques erronées; et
- syntaxe et informations scientifiques invalides en raison de la complexité de la présentation dite en mode mixte des séquences biologiques.

7. Il s'ensuit que les fournisseurs de bases de données publiques laissent de côté les informations fournies par les déposants et les reconstituent intégralement. Les offices qui communiquent des listages des séquences aux répertoires publics consacrent par conséquent beaucoup de temps et d'efforts à la correction des données.

NOUVELLE NORME PROPOSÉE

8. La nouvelle norme proposée par l'OEB remédierait aux inconvénients de la norme ST.25 de l'OMPI et contribuerait à favoriser la qualité des séquences figurant dans les bases de données de brevets et dans les bases de données de séquences publiques. Il convient également de noter que la norme ST.36 de l'OMPI recommande l'utilisation des normes en vigueur dans l'industrie[1].

9. La nouvelle norme aura donc les caractéristiques suivantes :

- universalité : un format unique pour les spécialistes des brevets et les autres utilisateurs; et
- fiabilité et souplesse : un format XML basé sur une DTD convenue.

Universalité

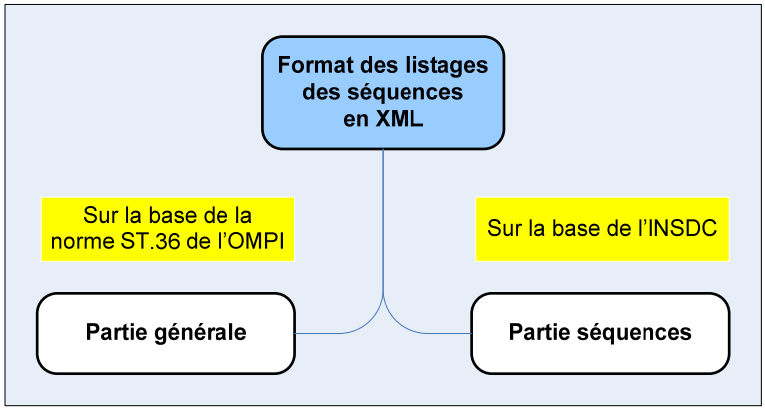
10. Un listage des séquences est constitué d'une "partie information générale" décrivant l'information relative à la demande de brevet et d'une "partie séquences" composée d'un certain nombre de séquences biologiques.

11. Les fournisseurs de données publiques faisant partie de l'International Nucleotide Sequence Database Collaboration (INSDC) (c'est-à-dire, European Bioinformatics Institute (EBI), DDBJ et NCBI) ont mis en point un format XML appelé INSDSeq pour l'échange de séquences et d'informations connexes. Ce format est censé faciliter l'adjonction de l'information de brevet pertinente.

---

[1] Voir le paragraphe 93 de la norme ST.36 de l'OMPI : Lorsque cela convient au contenu du document, c'est-à-dire lorsque le contenu n'est pas propre au domaine de la propriété industrielle, utiliser des DTD courantes dans l'industrie.

12. La nouvelle norme relative à la présentation des listages des séquences devrait par conséquent utiliser le format INSDSeq pour décrire l'information relative aux séquences et la modifier en fonction de l'information connexe en matière de brevets (information générale des listages des séquences) conformément à la norme ST.36 de l'OMPI. L'objectif est d'assurer la synchronisation de la partie séquences du format des listages des séquences de l'OMPI avec le format INSDSeq.



13. Concrètement, la nouvelle norme en XML devrait être plus facile d'utilisation, dans la mesure où les scientifiques utiliseraient pratiquement le même format à la fois pour les demandes de brevet et pour la communication aux bases de données publiques, sans qu'il soit nécessaire de procéder à des conversions. La simple familiarité avec la syntaxe de la norme pourrait en elle-même réduire le nombre d'erreurs.

Fiabilité et souplesse

14. La syntaxe de la nouvelle norme assurée par la DTD sera à la fois plus précise et plus facile à vérifier par des moyens automatisés. La vérification pourra s'effectuer grâce à un large éventail de logiciels gratuits existants.

15. Le contenu d'un fichier XML, bien que plus complexe à déchiffrer par l'homme sans feuilles de style, est en fait plus facile à manipuler par ordinateur, et il existe un grand nombre de bibliothèques à cet effet.

ÉTAT D'AVANCEMENT DES TRAVAUX

16. Un projet de DTD est à l'examen entre les offices de la coopération trilatérale. La DTD retenue constituera la base de la proposition.

17. L'OEB a mis au point, en coopération avec l'EBI, un logiciel client pour la communication des séquences biologiques (Biological Sequence Submission Application for Patents (BiSSAP)), qui prévoit un module de vérification et une option pour la création de fichiers XML. Ce logiciel est donc compatible à la fois avec les normes ST.25 et INSDSeq/ST.36. Il a été testé avec des résultats positifs par des utilisateurs européens au moins d'août. Le calendrier pour la mise en œuvre et la diffusion au public de ce nouvel outil est en cours d'examen à l'OEB.

[Fin de l'annexe et du document]