

Comité des normes de l'OMPI (CWS)

Treizième session
Genève, 10 – 14 novembre 2025

NUMERISATION DE LA DOCUMENTATION EN MATIERE DE PROPRIETE INTELLECTUELLE

Document établi par le Bureau international

RESUME

1. Les membres du Comité des normes de l'OMPI (CWS) ont indiqué que l'un de leurs principaux défis était la numérisation de leur documentation en matière de propriété intellectuelle. Parallèlement, les administrations des brevets qui souhaitent inclure leurs documents de brevets dans la documentation minimale du Traité de coopération en matière de brevets (PCT) s'efforcent actuellement de satisfaire aux exigences énoncées dans les instructions administratives du PCT. Le présent document expose de façon détaillée certains des défis auxquels sont confrontés les offices et souligne la nécessité d'élaborer un ensemble de lignes directrices en matière de numérisation, en se concentrant dans un premier temps sur les collections de brevets.

RAPPEL

2. En 2024, le Secrétariat a mené une enquête auprès des membres du CWS afin de comprendre les problèmes qu'ils peuvent rencontrer lors de l'échange de données sur la propriété intellectuelle et les solutions potentielles à ces problèmes. Les résultats de l'enquête ont révélé que les plus gros problèmes rencontrés par les offices, quelle que soit leur taille, sont l'absence de données des bulletins de propriété intellectuelle dans un format lisible par machine et l'insuffisance des ressources, tant en termes de compétences du personnel qu'en termes de ressources informatiques, pour mener à bien ces activités (voir le paragraphe 8 du document CWS/12/23).

3. L'Équipe d'experts chargée de la documentation minimale du PCT (ci-après dénommée "équipe d'experts"), dirigée par l'Office européen des brevets (OEB) et l'Office des brevets et des marques des États-Unis d'Amérique (USPTO), a été créée en 2005 par la Réunion des

administrations internationales instituées en vertu du PCT afin de procéder à un examen complet de la documentation minimale du PCT. Toutefois, les travaux n'ont véritablement commencé qu'au début de l'année 2017, date à laquelle un programme de travail a été défini en vue d'atteindre les quatre objectifs ci-après (voir l'annexe du document PCT/MIA/24/4) :

- a) Objectif A : créer un inventaire actualisé des éléments de la littérature brevet et non-brevet qui font partie de la documentation minimale du PCT actuelle.
- b) Objectif B : recommander les conditions et les normes requises pour l'inclusion d'une collection de documents de brevet dans la documentation minimale du PCT.
- c) Objectif C : proposer des éléments bibliographiques et textuels clairement définis de données de brevet qui devraient figurer dans toutes les collections de brevets faisant partie de la documentation minimale du PCT.
- d) Objectif D : recommander les conditions et les normes requises pour la révision, l'ajout et la tenue à jour de la littérature non-brevet et de l'état de la technique découlant de savoirs traditionnels et évaluer ensuite, sur la base des critères qui auront été établis, la proposition révisée des autorités indiennes au sujet de la bibliothèque numérique des savoirs traditionnels.

4. À la suite des discussions menées au sein de l'équipe d'experts et du Groupe de travail du PCT, l'Assemblée de l'Union du PCT a adopté en juillet 2023 des modifications apportées aux règles 34, 36 et 63 du règlement d'exécution du PCT, qui entreront en vigueur le 1^{er} janvier 2026 (document PCT/A/55/2). Les modifications apportées aux instructions administratives du PCT ont été promulguées en juin 2024 (circulaire C. PCT 1672), y compris l'adjonction d'une nouvelle annexe H. L'annexe H expose en détail les exigences techniques tendant à ce que toutes les collections de brevets et de modèles d'utilité appartenant à la documentation minimale du PCT soient mises gratuitement à la disposition de chaque administration chargée de la recherche internationale dans un format lisible par machine, au moins en ce qui concerne les publications à compter du 1^{er} janvier 2026, les publications remontant au moins à 1991 devant être mises à disposition en format texte intégral au cours des 10 années suivantes.

5. L'annexe H précise qu'il existe trois formats autorisés pour les données consultables par recherche textuelle : les instances XML conformes à la norme ST.36 ou à la norme ST.96 de l'OMPI, ou le format texte brut. Les administrations peuvent choisir le format qu'elles jugent le plus approprié, à condition qu'il permette une recherche de haute qualité dans leurs collections de brevets. Le Bureau international recommande l'utilisation du format XML, au moins pour la fourniture de données bibliographiques. Les offices qui choisissent le format texte brut comme format de préférence sont invités à envisager de migrer vers l'une des normes de l'OMPI afin d'améliorer la qualité de leurs publications de brevets et leurs possibilités de recherche.

PREPARATIFS

6. Afin d'aider les administrations chargées de la recherche internationale à satisfaire aux exigences techniques avant la date limite du 1^{er} janvier 2026, le Bureau international a organisé en 2025 une série de séminaires bilatéraux sur la documentation minimale du PCT. Les offices ont préparé à l'avance les questions qu'ils souhaitaient poser au Bureau international, et ce dernier s'est assuré que l'expert compétent était présent au séminaire pour y répondre. Au cours de ces séances, il a été noté que les instructions administratives du PCT ne fournissaient aucune indication sur la qualité des documents en texte intégral. À la demande de l'Équipe d'experts du CWS chargée du fichier d'autorité, le Bureau international a fait part de son idée d'établir un ensemble d'exigences minimales pour ce qu'il considère comme un document déchiffrable par machine "acceptable".

7. En ce qui concerne les offices qui numérisent leurs documents de brevets, les indications ci-après donnent des orientations permettant de se conformer aux exigences de l'annexe H :

- Les instances XML conformes à la norme ST.36 de l'OMPI doivent être bien formées, avoir un élément racine cohérent et présenter un minimum d'erreurs liées à la reconnaissance optique des caractères. Les données bibliographiques doivent inclure au moins la référence de la demande, la référence de la publication et le titre de l'invention.
- Les instances XML conformes à la norme ST.96 de l'OMPI doivent être conformes à la composante de document `PatentPublication` de cette norme, mais il n'est pas nécessaire d'appliquer `MathML` ou `OASISTable`.
- Les fichiers texte brut doivent présenter un minimum d'erreurs liées à la reconnaissance optique de caractères, le texte doit être présenté sous forme de paragraphes corrects, sans numéros de ligne ou de page, et permettre une lecture dans l'ordre approprié. Ils doivent de préférence être accompagnés de données bibliographiques au format XML.

8. L'Équipe d'experts chargée de la documentation minimale du PCT est convenue d'une procédure de certification en trois phases pour les administrations chargées de la recherche internationale ou les autres offices de brevets souhaitant que leur collection fasse partie de la documentation minimale du PCT. La deuxième phase, ou "phase de vérification", a permis de déterminer si un office était en mesure de fournir un référentiel de données contenant des documents conformes à l'un des trois formats indiqués à l'annexe H pouvant être téléchargés et de produire un fichier d'autorité conforme à la norme ST.37 de l'OMPI, y compris les indicateurs de possibilité d'effectuer une recherche textuelle.

DIFFICULTES RENCONTREES PAR LES OFFICES

9. Au cours des ateliers sur la documentation minimale du PCT et de la phase de vérification, le Bureau international a formulé des commentaires afin d'aider à l'élaboration de publications de brevets déchiffrables par ordinateur et à la production de fichiers d'autorité conformes à la norme ST.37 de l'OMPI. À l'issue de ce processus d'examen, les difficultés communément rencontrées par les offices lors de la production de documents déchiffrables par ordinateur étaient les suivants :

- a) pour tous les formats, une mauvaise qualité de reconnaissance optique de caractères entraînant de nombreuses erreurs et, par conséquent, polluant les index de recherche, en particulier pour les documents de brevets contenant des formules chimiques;
- b) pour tous les formats, une mauvaise détection ou gestion de la mise en page, entraînant l'inclusion d'informations d'en-tête ou de pied de page dans le texte, ou l'apparition de numéros de ligne dans le résultat;
- c) pour tous les formats, une mauvaise segmentation des sections relatives à la description, les revendications et les dessins, entraînant par exemple l'inclusion du texte de la description dans le texte intégral des revendications;
- d) pour tous les formats, un mauvais balisage linguistique des sections de texte intégral, entraînant des problèmes d'indexation critiques dans les systèmes de recherche;
- e) pour les instances conformes à la norme ST.96 de l'OMPI, les instances n'ayant pas été validées par rapport à la composante de document `PatentPublication` et la confusion autour de l'utilisation des codes de langue dans la hiérarchie;

f) pour les instances conformes à la norme ST.36 de l'OMPI, une mauvaise compréhension de ce qui est considéré comme une instance "valide" et de la norme minimale à respecter pour déterminer si le document XML est "acceptable". par exemple, si le DTD type de la norme ST.36 doit être utilisé; et

g) pour le format texte brut, le texte n'est pas présenté sous forme de paragraphes appropriés, mais sous forme de lignes, ce qui rend le contenu difficile à rechercher, et les données bibliographiques sont fournies dans des formats non conformes à la norme de l'OMPI.

ASSISTANCE TECHNIQUE

10. Afin de résoudre certaines des difficultés mentionnées ci-dessus, le Bureau international fournit une assistance technique aux offices qui en font la demande. La *solution ROC de l'OMPI* permet de produire des instances XML conformes à la norme ST.36 de l'OMPI à partir de fichiers PDF contenant des signets, consultables par image et par texte. Toutefois, le CWS doit noter que cet outil ne prend pas en charge la détection et le balisage des numéros de revendication et traite les tableaux et les formules chimiques et mathématiques complexes comme des dessins intégrés.

11. *WIPO Publish* permet aux offices d'extraire et de contrôler la qualité de leurs données au format ST.36 de l'OMPI et, prochainement, les résultats seront migrés pour être conformes à la norme ST.96 de l'OMPI. Cela facilite l'échange d'informations issues des bulletins officiels pour de nombreux offices de propriété intellectuelle. En 2025, le Bureau international a ajouté des fonctionnalités permettant d'extraire des documents de brevet pour les offices qui appliquent les exigences en matière de documentation minimale du PCT. Un projet pilote a été mené à bien avec l'Office de la propriété intellectuelle des Philippines. La même approche pourrait être utilisée pour d'autres offices de brevets dont la collection fait partie de la documentation minimale du PCT; tout office intéressé peut contacter le Bureau international.

12. Le Bureau international fournit également une assistance directe à une trentaine d'offices afin de les aider à numériser leur documentation, y compris par reconnaissance optique de caractères en texte intégral pour certains offices. Ce processus comprend la numérisation des documents, leur conversion numérique, puis un contrôle qualité des données bibliographiques. Dans tous les cas, ce travail est sous-traité à des prestataires tiers.

PROPOSITIONS RELATIVES AUX PROCHAINES ETAPES

13. Compte tenu des difficultés susmentionnées, le Bureau international souhaite proposer l'élaboration d'un ensemble de lignes directrices relatives à la numérisation des documents de propriété intellectuelle, afin de contribuer à combler le déficit de compréhension évoqué plus haut. Afin d'examiner plus en détail les difficultés rencontrées par les offices qui procèdent actuellement à la numérisation de leurs documents de propriété intellectuelle, le Bureau international propose de mener une enquête au moyen d'un questionnaire publié dans une circulaire conjointe du CWS et du PCT. Le Bureau international a l'intention d'établir ce questionnaire et d'inviter les membres des deux comités à répondre à l'enquête au cours du premier trimestre 2026.

14. Bien que les discussions à ce stade se concentrent sur la numérisation des collections de brevets, il est probable que ces lignes directrices soient élargies afin d'inclure des orientations supplémentaires sur la numérisation des documents relatifs aux marques et aux dessins et modèles industriels.

15. À l'issue de l'enquête, le Bureau international a l'intention d'organiser une réunion afin de discuter de l'élaboration des lignes directrices en matière de numérisation au cours du premier semestre 2026. Tous les membres du CWS et du PCT seront invités, mais les informations relatives à la réunion, notamment son format et sa ou ses dates, seront communiquées en

temps utile. Lors de cette réunion, le Bureau international a l'intention de présenter les résultats de l'enquête ainsi qu'un calendrier prévisionnel pour l'élaboration des lignes directrices relatives à la numérisation des documents de propriété intellectuelle. Les membres du CWS sont encouragés à participer à l'enquête et à cette réunion, car leurs commentaires sont nécessaires à l'élaboration des lignes directrices relatives à la numérisation. Le Bureau international a également l'intention de rendre compte des résultats de cette réunion à la quatorzième session du CWS.

16. En outre, le Bureau international prévoit d'organiser un atelier sur les normes ST.36 et ST.96 de l'OMPI au cours du deuxième trimestre 2026. Les invitations à cet atelier seront envoyées aux membres et observateurs du CWS et l'événement sera annoncé sur le site Web de l'OMPI. Le Bureau international invite également les offices qui mettent en œuvre les normes de l'OMPI à solliciter une assistance technique en cas de besoin.

17. *Le CWS est invité*

a) à prendre note du contenu du présent document, en particulier des difficultés rencontrées par les offices, comme indiqué au paragraphe 9,

b) à approuver la proposition de mener l'enquête par l'intermédiaire de la circulaire conjointe qui serait diffusée auprès des membres du CWS et du PCT, comme indiqué au paragraphe 13 ci-dessus,

c) à encourager ses membres à participer à l'enquête et à la réunion, et à noter que les résultats de la réunion seront présentés à la quatorzième session du CWS, comme indiqué au paragraphe 15 ci-dessus, et

d) à encourager ses membres et observateurs à participer à l'atelier sur les normes OMPI ST.36 et ST.96 qui se tiendra au deuxième trimestre 2026, comme indiqué au paragraphe 16 ci-dessus.

[Fin du document]