

## **Comité des normes de l'OMPI (CWS)**

**Dixième session**  
**Genève, 21 – 25 novembre 2022**

### **RAPPORT DE L'ÉQUIPE D'EXPERTS CHARGÉE DE LA NORMALISATION DES NOMS (TACHE N° 55)**

*Document établi par les responsables de l'Équipe d'experts chargée de la normalisation des noms*

#### **RAPPEL**

1. À sa neuvième session, tenue en 2021, le Comité des normes de l'OMPI (CWS) a pris note des progrès accomplis par l'Équipe d'experts chargée de la normalisation des noms. Plus précisément, l'équipe d'experts a fait état de son travail de collecte d'informations sur les activités de nettoyage des données à l'appui de la normalisation des noms. L'équipe d'experts a fait part de son intention de présenter des recommandations à la dixième session du CWS. (Voir les paragraphes 117 à 118 du document CWS/9/25.)

#### **RAPPORT SUR LES ACTIVITES**

2. L'équipe d'experts a continué à recueillir des informations auprès de ses membres sur leur expérience en matière de nettoyage des données à des fins de normalisation des noms. Des questions plus détaillées ont été soulevées au regard des données précédemment collectées afin d'obtenir des informations supplémentaires utiles à l'équipe d'experts. Six membres de l'équipe d'experts ont soumis des propositions au cours du premier trimestre de 2022.

3. Sur la base des informations collectées, l'équipe d'experts a commencé à travailler sur des projets de recommandations relatives aux meilleures pratiques. Les recommandations portent sur des considérations générales pour la collecte, le traitement, le nettoyage et la publication des données sur les noms nettoyés. Elles n'abordent pas les nombreuses questions complexes liées à des approches particulières du nettoyage des données, de la translittération ou de la normalisation des noms, telles que le choix des algorithmes, le lieu et le

moment où les transformations sont appliquées, la fréquence ou les stratégies de fusion. Ces types de décisions varieront considérablement en fonction de la partie qui les applique, de l'objectif des transformations et de la nature en constante évolution des algorithmes de correspondance.

4. Une version préliminaire des recommandations figure dans l'annexe du présent document. Le projet de recommandations est à un stade très précoce et ne reflète pas encore l'accord ou le consensus de l'équipe d'experts. Il est présenté au CWS à des fins d'information et de commentaires. Les recommandations finales pourraient changer considérablement.

5. L'équipe d'experts prévoit de poursuivre les travaux sur le projet de recommandations en 2023, avec plusieurs séries de discussions. L'équipe d'experts entend présenter une proposition finale de recommandations à la prochaine session du CWS.

6. *Le CWS est invité*

a) *à prendre note du contenu du présent document;*

b) *à prendre note de l'état d'avancement du projet de recommandations relatives aux données propres à l'appui de la normalisation des noms figurant dans l'annexe du présent document; et*

c) *à formuler des observations sur le projet de recommandations.*

[L'annexe suit]

## RECOMMANDATIONS CONCERNANT LE NETTOYAGE DES DONNÉES POUR LA NORMALISATION DE NOMS

### PROJET DE DOCUMENT DE TRAVAIL

*Note établie par le Bureau international*

*Le présent projet de document de travail est établi par l'Équipe d'experts 3D et présenté à la huitième session du CWS pour information uniquement en anglais. Ce projet sera actualisé par l'équipe d'experts et une version finale sera soumise au CWS pour examen à sa prochaine session.*

#### PORTÉE

1. La présente norme couvre les considérations d'ordre général pour la collecte, le traitement, le nettoyage et la publication des données propres relatives aux noms. Elle ne traite pas des nombreuses questions complexes liées à des approches particulières concernant le nettoyage des données, la translittération ou la normalisation des noms, telles que le choix des algorithmes, le lieu et le moment où les transformations sont appliquées, la fréquence ou les stratégies de fusion. Ces décisions varieront considérablement en fonction de la partie qui les applique, de l'objectif des transformations et de la rapidité avec laquelle les algorithmes de correspondance évoluent.

#### DÉFINITIONS

Dans le présent document :

2. Les données clients sont des données sur les déposants, les titulaires d'enregistrements, les propriétaires, les représentants légaux ou d'autres parties, détenues par un office de propriété intellectuelle en rapport avec un droit de propriété intellectuelle, une demande, un enregistrement ou tout autre instrument. Cette norme concerne principalement les données relatives aux noms des clients : les noms de personnes, les noms d'entreprises et les informations connexes telles que la ville, l'adresse ou l'adresse électronique qui peuvent être utilisées pour lever toute ambiguïté sur d'éventuelles correspondances de noms.

3. Les données propres sont des données précises et fiables, exemptes d'erreurs et de doublons. Étant donné qu'aucun ensemble de données complexes ne sera jamais parfaitement propre, la propreté doit être considérée en termes relatifs, avec des mesures telles que le pourcentage de données contenant des erreurs assorties d'un facteur d'incertitude.

#### COLLECTE

4. Les offices de propriété intellectuelle doivent donner aux clients la possibilité de créer et de gérer des dossiers clients électroniques contenant des informations nominatives publiées : noms de personnes, noms d'entreprises, noms de représentants légaux, et informations connexes telles que la ville, l'adresse ou l'adresse électronique.

5. Les offices de propriété intellectuelle devraient permettre d'associer le dossier d'un client à plusieurs demandes ou enregistrements de droits de propriété intellectuelle, afin que les clients puissent réutiliser les mêmes informations nominatives pour plusieurs demandes ou enregistrements et mettre à jour leurs informations nominatives à un seul et même endroit.

6. Les offices de propriété intellectuelle peuvent permettre aux clients de saisir et de mettre à jour eux-mêmes les informations nominatives, ou peuvent exiger qu'une partie désignée, telle que des employés, des sous-traitants ou un service externe, saisisse et mette à jour les dossiers des clients à la demande de ces derniers.

7. Plusieurs dossiers pour un même client peuvent être créés et gérés par différentes entités, comme des représentants légaux différents. Les offices de propriété intellectuelle devraient en tenir compte lors de la conception de leurs systèmes de dossiers clients, car il se pourrait que plusieurs dossiers concernant un même client contiennent des données légèrement différentes ou soient mis à jour à des moments différents par des représentants différents.

8. Les offices de propriété intellectuelle devraient permettre la saisie du nom du client dans les caractères propres à la langue du client, en plus du nom du client dans la ou les langues de l'office de propriété intellectuelle. Par exemple, un office de propriété intellectuelle qui travaille en anglais pourrait prévoir des champs distincts pour le nom du déposant en anglais et le nom original dans d'autres caractères, selon que de besoin.

9. Les offices de propriété intellectuelle peuvent utiliser des numéros d'identification pour identifier les clients s'ils le souhaitent. Les numéros peuvent être créés par l'office de propriété intellectuelle ou tirés d'une source externe, comme un numéro d'entreprise enregistré ou un numéro de passeport. Les numéros d'identification ne permettent pas à eux seuls de résoudre de nombreux problèmes liés à la propreté des données clients, tels que les doublons, les changements de nom et les informations obsolètes ou incorrectes. Les offices de propriété intellectuelle qui utilisent des numéros d'identification devraient continuer à prêter attention et à tenir compte des considérations figurant dans les autres parties de la présente norme.

#### TRANSLITTÉRATION

10. Pour l'échange de données, y compris la réception de demandes internationales ou d'enregistrements internationaux, les offices de propriété intellectuelle devraient envoyer et recevoir les données en utilisant le jeu de caractères UTF-8.

11. Si un office de propriété intellectuelle procède à la translittération de caractères d'une langue (comme le grec) à une autre (comme l'anglais), il devrait publier son schéma de translittération. Le document produit par la translittération doit être mis à la disposition du client pour examen et ce dernier doit avoir la possibilité de soumettre des corrections si la translittération est entachée d'erreurs.

12. La translittération inversée doit être évitée si possible, et l'on préférera utiliser le nom original. Par exemple, pour une demande déposée par "Phony Corp", une translittération en caractères grecs donnerait "Φωνι Κορπ" dans le système d'un office de propriété intellectuelle et, lors de la publication et d'une translittération inverse en caractères latins, donnerait "Foni Corp", ce qui entraînerait des incohérences.

#### TRANSCRIPTION

*L'équipe d'experts doit examiner les recommandations...*

#### TRADUCTION

*L'équipe d'experts doit examiner les recommandations...*

#### VALIDATION ET LEVÉE DES AMBIGUÏTÉS

13. Les offices de propriété intellectuelle peuvent choisir de procéder à la validation des informations soumises concernant les clients, y compris à des contrôles automatisés. Les résultats de validation devraient être mis à la disposition du client, et les corrections acceptées par le client selon que de besoin, de même qu'il faudrait disposer de moyens permettant de contourner les mécanismes de validation automatisée si celui-ci fournit des résultats incorrects ou incomplets.

14. Les offices de propriété intellectuelle qui tentent de lever des ambiguïtés dans les dossiers nominatifs (c'est-à-dire de trouver des doublons) pourraient vouloir tenir compte d'autres éléments que les seuls noms des clients. Les noms ne sont pas uniques par définition. Par exemple, il existe plusieurs personnes nommées "John Smith" ou plusieurs entreprises nommées "Data Corp". Si l'on compare des éléments de données connexes, tels que la ville, le code postal, la date de naissance ou d'autres informations, lorsqu'elles sont disponibles, on peut augmenter la probabilité d'obtenir des correspondances.

15. Tout processus de validation ou de levée des ambiguïtés lancé par l'office de propriété intellectuelle et susceptible d'avoir des conséquences juridiques, comme la correction ou la normalisation du nom du propriétaire enregistré d'un droit de propriété intellectuelle, doit être validé par le client avant que le changement ne soit effectué dans le système de l'office de propriété intellectuelle.

#### MAINTENANCE

16. Les offices de propriété intellectuelle devraient élaborer une stratégie de nettoyage périodique des données, comprenant la recherche et la résolution des doublons, c'est-à-dire les enregistrements multiples pour la même entité. Dans certains cas, les doublons pourront être fusionnés ou combinés, comme dans le cas d'enregistrements présentant de légères différences d'orthographe involontaires, tels que "ABC Corp" et "ABC Corp.". Dans d'autres cas, il peut être préférable de tenir des dossiers séparés. Chaque office de propriété intellectuelle devrait décider de l'approche qui convient le mieux à son propre système de gestion des dossiers clients.

17. Les offices de propriété intellectuelle devraient proposer aux clients un mécanisme qui leur permette de mettre à jour les informations nominatives sur plusieurs demandes ou droits de propriété intellectuelle en saisissant ces informations une seule fois. Pour ce faire, on pourrait, par exemple, associer chaque demande ou droit de propriété intellectuelle à un seul dossier client contenant des informations sur le nom, ou permettre aux clients de sélectionner plusieurs demandes ou droits de propriété intellectuelle et de soumettre une seule fois des informations actualisées sur le nom à utiliser pour toutes ces demandes ou droits.

18. Les offices de propriété intellectuelle devraient désigner une personne chargée de traiter les questions relatives à la propreté des données, notamment d'élaborer des critères pour déterminer la propreté des données, d'assurer un suivi et d'établir des rapports réguliers sur ces critères, et de prendre des mesures pour améliorer la qualité des données clients selon que de besoin.

#### PUBLICATION ET ÉCHANGE DE DONNÉES

19. Les offices de propriété intellectuelle devraient mettre à disposition les mises à jour des informations nominatives qui sont effectuées après la publication d'un droit de propriété intellectuelle. Par exemple, si "ABC Corp" remplace son nom par "XYZ Corp" dans le dossier client, le nom "XYZ Corp" devrait être associé au droit de propriété intellectuelle dans les publications en ligne. Le nom original peut également apparaître sur le droit de propriété intellectuelle publié, selon les exigences légales de l'office de propriété intellectuelle.

20. Si un office de propriété intellectuelle a d'autres formes de noms de clients, comme le nom original en caractères propres à la langue du client, celles-ci doivent être incluses dans les données publiées et les données échangées avec les autres offices de propriété intellectuelle.

21. Si un office de propriété intellectuelle utilise des numéros d'identification pour identifier les entités, ces numéros doivent figurer dans les données publiées et les données échangées avec les autres offices de propriété intellectuelle. Si les numéros d'identification sont confidentiels et ne peuvent être partagés, l'office de propriété intellectuelle doit indiquer quelles données clients utilisent les mêmes numéros d'identification, par exemple en remplaçant les numéros confidentiels par des numéros uniques générés pour la publication.

#### UTILISATION À DES FINS STATISTIQUES

22. À des fins statistiques, les offices de propriété intellectuelle peuvent tenter de rapprocher les données clients et des variations d'un nom ou d'autres champs afin d'obtenir des chiffres plus précis. Dans ce cas, les offices de propriété intellectuelle devraient publier leur stratégie ou leur algorithme de correspondance ainsi que leurs résultats statistiques afin que les autres puissent comprendre la méthodologie employée.

[Fin de l'annexe et du document]