

Comité des normes de l'OMPI (CWS)

Onzième session
Genève, 4 – 8 décembre 2023

PROPOSITION RELATIVE AUX PRINCIPES DIRECTEURS CONCERNANT LE NETTOYAGE DES DONNEES DES NOMS

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

RAPPEL

1. À sa cinquième session tenue en 2017, le Comité des normes de l'OMPI (CWS) a ajouté une nouvelle tâche au programme de travail du CWS : la tâche n° 55. L'Équipe d'experts chargée de la normalisation des noms a été créée lors de la même session et s'est vu confier cette tâche n° 55, décrite comme suit :

“Envisager l'élaboration d'une norme de l'OMPI visant à aider les offices de propriété industrielle à améliorer la 'qualité à la source' des noms des déposants, établir une proposition concernant la poursuite des travaux relatifs à la normalisation des noms des déposants dans les documents de propriété industrielle, et la présenter au CWS pour examen”.

(Voir les paragraphes 85 à 88 du document CWS/5/22.)

2. À sa septième session, tenue en 2019, le CWS a noté que le Secrétariat avait diffusé en novembre 2018 une enquête sur l'utilisation des identifiants des déposants par les offices de propriété industrielle, comme convenu à la sixième session du CWS. L'enquête concernant l'utilisation d'identifiants étant terminée, l'équipe d'experts a proposé de réviser la description de la tâche n° 55 en tenant compte de cette enquête. Le CWS a approuvé la description révisée de la tâche n° 55, désormais libellée comme suit :

“Envisager l'élaboration d'une norme de l'OMPI visant à aider les offices de propriété industrielle à améliorer la 'qualité à la source' des noms des déposants, établir une proposition visant la poursuite des travaux relatifs à la normalisation des noms des déposants dans les documents de propriété industrielle, et la présenter pour examen par le CWS”.

(Voir les paragraphes 72 à 79 du document CWS/7/29.)

3. Un atelier a été organisé par le Bureau international en 2019 pour examiner les questions relatives à la normalisation des noms. Les conférenciers ont été invités par les offices et les acteurs du secteur à mettre en commun des pratiques et à proposer d'éventuelles solutions. Lors de cet atelier, le Bureau international a noté que les offices de propriété industrielle avaient de nombreuses approches différentes en matière de nettoyage des données à des fins de normalisation des noms.

4. Des informations détaillées concernant les progrès réalisés par l'équipe d'experts depuis la dernière session du CWS sur leur programme de travail sont présentées dans le document CWS/11/22.

PROPOSITION DE PRINCIPES DIRECTEURS

5. Dans le cadre de la tâche n° 55, l'Équipe d'experts chargée de la normalisation des noms a élaboré une proposition finale relative à un ensemble de principes directeurs visant à aider les offices de propriété industrielle qui entreprennent un nettoyage des données des noms. Ces principes directeurs figurent en annexe du présent document.

6. Ces principes directeurs ont été établis afin de fournir des recommandations générales et de haut niveau. Les variations observées dans les facteurs tels que les exigences juridiques, les pratiques en matière de données, les objectifs du nettoyage, l'utilisation prévue des données, les exigences en matière de ressources et les considérations techniques signifient qu'il n'y a pas d'approche unique à privilégier pour tous les offices de propriété industrielle. Ces recommandations visent à rendre compte des pratiques générales applicables à tout office de propriété industrielle pour faciliter le nettoyage des données relatives au nom des clients, qui améliore, quant à lui, la normalisation des noms et les techniques d'appariement par les utilisateurs en aval.

7. L'Équipe d'experts note que la norme ST.20 de l'OMPI fournit des recommandations pour établir des index de noms propres des déposants et autres clients figurant sur les documents de brevet, et pour unifier la présentation et la méthode de classement des noms dans ces index. Sous leur forme définitive, les principes directeurs comprendront par conséquent une référence à cette norme, dont ils encouragent l'utilisation.

8. Si ces principes directeurs sont approuvés par le CWS durant la session en cours, il est proposé que le CWS prie le Secrétariat de les publier dans la partie 3 du Manuel de l'OMPI, à l'adresse suivante : https://www.wipo.int/standards/fr/part_03_standards.html.

9. *Le CWS est invité*

a) à prendre note du contenu du présent document et de son annexe,

b) à examiner et à approuver les principes directeurs tels que présentés aux paragraphes 5 à 7 ci-dessus et qui figurent à l'annexe du présent document et

c) à prier le Secrétariat de publier les principes directeurs dans la partie 3 du Manuel de l'OMPI comme indiqué au paragraphe 8 ci-dessus.

[L'annexe suit]

PRINCIPES DIRECTEURS CONCERNANT LE NETTOYAGE DES
DONNÉES DES NOMS

*Proposition présentée au Comité des normes de l'OMPI (CWS) pour approbation
à sa onzième session*

Introduction.....	2
Définitions.....	2
Collecte	2
Transformation des noms	3
Validation et levée des ambiguïtés	3
Maintenance	4
Publication et échange de données.....	4
Utilisation à des fins statistiques.....	4
Références	4
Annexe	5
Exemples de translittération	5
Exemples de transcription	6
Exemples de traduction	6

PRINCIPES DIRECTEURS CONCERNANT LE NETTOYAGE DES DONNÉES DES NOMS

*Proposition présentée au Comité des normes de l'OMPI (CWS) pour approbation
à sa onzième session*

INTRODUCTION

1. La présente série de principes directeurs 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 concernant le nettoyage des données, la localisation ou la transformation des noms, telle que la translittération, la transcription ou la traduction, ou des approches concernant la normalisation des noms, telles que la sélection 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.
2. Il convient de noter que la norme ST.20 de l'OMPI fournit des recommandations concernant l'établissement d'index des documents de brevet donnant le nom des déposants et d'autres clients, et vise à unifier la présentation et la méthode de classement des noms dans ces index.

DEFINITIONS

3. Dans le présent document, on entend par :
 - a) "office de propriété intellectuelle", l'office de propriété intellectuelle chargé de la gestion des procédures de dépôt et d'enregistrement relatives aux droits de propriété intellectuelle;
 - b) "données clients", les 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;
 - c) "données propres", des données précises, cohérentes et fiables, exemptes d'erreurs et de doublons. Comme il est difficile de mesurer le degré de propreté d'un vaste ensemble de données complexes, diverses mesures peuvent être utilisées comme substituts de la propreté ou de propriétés connexes, telles que leur utilité;
 - d) "translittération", la mise en correspondance d'un ou plusieurs caractères de la langue source avec un ou plusieurs caractères (phonétiques) de la langue cible;
 - e) "transcription", la mise en correspondance d'un caractère, d'un logogramme, d'une syllabe ou d'un phonème de la langue source avec quelque chose qui correspond au son dans le système correspondant de la langue cible;
 - f) "traduction", la représentation du sens d'un mot ou d'une notion dans la langue source par quelque chose qui correspond au sens dans la langue cible.

COLLECTE

4. Les offices de propriété intellectuelle peuvent 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 proposer un ou plusieurs formulaires que les clients peuvent utiliser pour demander aux offices de propriété intellectuelle de créer ou de modifier leur nom ou des informations connexes. Les offices de propriété intellectuelle peuvent également permettre aux clients de saisir et de mettre à jour eux-mêmes les informations nominatives ou connexes, 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 peuvent 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 travail de l'office de propriété intellectuelle, qui doit être stocké en utilisant le codage UTF-8¹. 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 du déposant en coréen.

9. Les offices de propriété intellectuelle peuvent en outre utiliser des numéros d'identification pour identifier les clients. Les numéros d'identification 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 des présents principes directeurs.

TRANSFORMATION DES NOMS

10. Pour l'échange et le traitement des données, y compris la réception de demandes internationales ou d'enregistrements internationaux, les offices de propriété intellectuelle peuvent envisager la possibilité de transformer les noms (voir l'annexe du présent document). Il est recommandé que les offices de propriété intellectuelle envoient et reçoivent les données relatives aux noms en utilisant le jeu de caractères UTF-8.

11. Il convient de noter que la localisation ou la conversion des noms de clients est fortement sujette à erreur, car il n'existe pas de normes généralement acceptées ou uniformes. Pour la localisation ou la conversion des noms, il existe trois méthodes mentionnées dans les présents principes directeurs : la translittération, la transcription et la traduction. Si les offices de propriété intellectuelle procèdent à la translittération, à la transcription ou à la traduction de caractères d'une langue (comme le grec) à une autre (comme l'anglais), ils devraient publier leur schéma de translittération, de transcription ou de traduction. Le document produit par la translittération, la transcription ou la traduction, ou des parties de ce document, doivent être mis à la disposition du client pour examen et ce dernier doit avoir la possibilité de soumettre des corrections si la translittération, la transcription ou la traduction est entachée d'erreurs.

12. La translittération inversée doit être évitée si possible; il est recommandé d'utiliser plutôt le nom original. Par exemple, pour une demande déposée par "Phony Corp", une translittération en caractères grecs donnerait "Φονι 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. Des exemples de problèmes courants liés à la translittération, à la transcription ou à la traduction inverse, ou à la retranslittération, à la retranscription ou à la retraduction, figurent dans l'annexe de ces principes directeurs.

VALIDATION ET LEVEE DES AMBIGUÏTES

13. Les méthodes de validation et de levée des ambiguïtés doivent être conçues pour répondre à des objectifs spécifiques, administratifs ou statistiques, et les méthodes appropriées doivent être appliquées en fonction de ces objectifs. Les méthodes de mise en correspondance et de levée des ambiguïtés concernant les des noms doivent être correctement définies et les risques doivent être évalués en fonction de leur objectif de conception, afin de garantir des niveaux appropriés de levée des ambiguïtés pour le cas d'utilisation considéré.

14. 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.

15. 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.

16. 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.

¹ UTF-8 est un système de codage pour Unicode.

MAINTENANCE

17. Les offices de propriété intellectuelle devraient élaborer une stratégie de nettoyage périodique des données dans les bases de données des noms des clients, 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. La stratégie peut faire intervenir les clients concernés par les dossiers dans le processus de nettoyage des données et la responsabilité des données nettoyées.

18. 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.

19. Les offices de propriété intellectuelle peuvent 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 ECHANGE DE DONNEES

20. 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.

21. Si un office de propriété intellectuelle a d'autres formes de noms de clients, comme le nom original exprimé à l'aide de 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.

22. 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 ces numéros d'identification, par exemple en remplaçant les numéros confidentiels par des numéros uniques générés pour la publication.

UTILISATION A DES FINS STATISTIQUES

23. À des fins statistiques, les offices de propriété intellectuelle peuvent tenter de rapprocher les données clients et des variations des noms des clients, 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.

REFERENCES

24. Aux fins de la présente série de principes directeurs, il est utile de se reporter aux normes suivantes :

Norme ST.20 de l'OMPI Établissement d'index des noms propres apparaissant sur les documents de brevet

[L'annexe suit]

ANNEXE

DIFFERENTS MOYENS DE TRANSFORMATION DES NOMS

*Proposition présentée au Comité des normes de l'OMPI (CWS) pour approbation
à sa onzième session*

Bien que la translittération et la transcription soient des notions distinctes d'un point de vue linguistique, le résultat est généralement très similaire pour les systèmes d'écriture basés sur des caractères. Cependant, la transcription offre un résultat plus pratique, car seuls les caractères standard de la langue cible sont nécessaires pour la conversion.

L'anglais étant la lingua franca de l'économie mondialisée, on oublie généralement que la transcription est rarement normalisée entre deux langues. Dans le meilleur des cas on trouve des définitions officielles pour [xx] -> [en], ce qui permet de supposer que [xx] -> [en] -> [yy] est égal à [xx] -> [yy], ce qui n'est généralement pas correct.

EXEMPLES DE TRANSLITTÉRATION²

La Figure 1 présente ci-dessous un exemple de courrier et des remarques concernant cette translittération.

Source and Target words	Letter Correspondence	Description
English to Persian		
John /dʒɒn/	J o h n	<i>h</i> is a silent letter (no sound is associated to the letter) and is not transliterated
جان /dʒɒn/	ج ا ن	
Arabic to English		
نجيب /nædʒiːb/	ن ج ي ب	short vowel /æ/ on N is normally not written in Arabic script
Najib /nædʒiːb/	Na j i b	
English to Japanese		
Bill /bi:l/	B i l l	each syllable in Japanese is a consonant-vowel sequence
ビル [bi-ru]	ビ ル	
English to Hindi		
Adam /ædəm/	A d a m	the second "a" is not transliterated in Hindi
अदम /ædəm/	अ द म	

Figure 1 : Exemple de translittération

² Enquête sur la translittération automatique, disponible à l'adresse suivante :

https://www.researchgate.net/figure/Transliteration-examples-in-four-language-pairs-Letter-correspondence-shows-how-the_fig1_220566444.

EXEMPLES DE TRANSCRIPTION

Voici des exemples où la transcription peut entraîner des inexactitudes :

[ru] : Ш → [de] : sch³

[ru] : Ш → [en] : sh

[ko] : ㅸ → [de] : ja⁴

[ko] : ㅸ → [en] : ya

[gr] : Ω → latin : O⁵

[da] : Æ → [de] : Ä ou AE, [en] : AE⁶

EXEMPLES DE TRADUCTION

Dans le premier exemple, il est clair que la traduction directe peut poser des problèmes :

[de] : Aktiengesellschaft → [en] : corporation, stock co, ...

[ru] : ОАО Силовые машины → [en] : OJSC "Power Machines" – OR – [en] : Open Joint-stock Company "Power Machines"

Voici un deuxième exemple, qui montre un cas limite typique de la romanisation d'un nom d'entreprise chinois figurant dans la Figure 2 :

- [zh] : 北京东土科技股份有限公司 → [en] translittération (pinyin) : běi jīng dōng tǔ kē jì gǔ fèn yǒu xiàn gōng sī;
- [zh] : 北京东土科技股份有限公司 → [en] transcription (pinyin) : beijing dongtu keji gufen youxian gongsi
- [zh] : 北京东土科技股份有限公司 → [en] traduction (anglais) : Beijing, China Science and Technology Joint-stock Limited Company
- [zh] : 北京东土科技股份有限公司 → en réalité : Kyland Technology Co., Ltd.

(71) 申请人: 北京东土科技股份有限公司 (KYLAND TECHNOLOGY CO., LTD) [CN/CN]; 中国北京市石景山区实兴大街30号院2号楼8层901, Beijing 100041 (CN)。

Figure 2 : Romanisation d'un nom d'entreprise chinois

[Fin de l'annexe et des principes directeurs]

³ https://de.wikipedia.org/wiki/Kyrillisches_Alphabet#Russisch

⁴ https://de.wikipedia.org/wiki/Koreanisches_Alphabet

⁵ https://en.wikipedia.org/wiki/Romanization_of_Greek

⁶ https://en.wikipedia.org/wiki/Dania_transcription