

CWS/13/15

الأصل: بالإنكليزية

التاريخ: 8 سبتمبر، 2025

## اللجنة المعنية بمعايير الويبو

الدورة الثالثة عشرة

جنيف، من 10 إلى 14 نوفمبر 2025

اقترح بشأن معيار جديد للويبو يدعم تنقية بيانات الأسماء

وثيقة من إعداد المشرفين المشاركين على فرقة العمل المعنية بتوحيد الأسماء

### ملخص

1. تقدّم فرقة العمل المعنية بتوحيد الأسماء بموجب هذه الوثيقة مسودة نهائية لمعيار جديد للويبو يُعنى بتنقية بيانات الأسماء، لكي تنظر فيها اللجنة المعنية بمعايير الويبو (لجنة المعايير) في دورتها الثالثة عشرة وتعتمدها.

### معلومات أساسية

2. وافقت لجنة المعايير، في دورتها الحادية عشرة المعقودة في عام 2023، على الوصف المنقح للمهمة رقم 55، الذي أصبح نصه الآن كما يلي:

"إعداد مقترح الإجراءات المستقبلية الرامية إلى تحقيق توحيد الأسماء في وثائق الملكية الفكرية، بنية وضع معيار للويبو يساعد مكاتب الملكية الفكرية على تحسين "الجودة في المنبع" فيما يتعلق بالأسماء."

(انظر (ي) الفقرات 75 إلى 78 من الوثيقة CWS/11/28).

3. وناقشت لجنة المعايير في الدورة نفسها المعقودة في عام 2023 مجموعة جديدة من المبادئ التوجيهية التي اقترحتها فرقة العمل المعنية بتوحيد الأسماء لدعم تنقية أسماء المودعين. ووافقت لجنة المعايير على استخدام مصطلح "التوصيات" عوضاً عن "المبادئ التوجيهية" في اسم معيار الويبو الجديد المقترح، لأنه أكثر وضوحاً. وأحاطت لجنة المعايير علماً كذلك بالاسم المقترح من الأمانة، "معيار الويبو ST.93" للتوصيات بشأن تنقية بيانات الأسماء (انظر (ي) الفقرة 135 من الوثيقة CWS/11/28).

4. ولكن لجنة المعايير لم تعتمد المعيار المقترح وأحالته إلى فرقة العمل من أجل مواصلة المناقشة والتحسين. وأشارت لجنة المعايير أيضاً إلى أن الأمانة سوف تنظر في إمكانية نشر مجموعة من جداول النقل الحرفي على موقع الويبو الإلكتروني (انظر (ي) الفقرتين 136 و 137 من الوثيقة CWS/11/28).

5. واستعرضت لجنة المعايير في دورتها الثانية عشرة مسودة محسنة بشأن معيار الويبو المقترح بشأن تنقية بيانات الأسماء، بالصيغة المقدمة من فرقة العمل المعنية بتوحيد الأسماء. وأعربت عدة وفود عن تأييدها المعيار المقترح. ولكن لجنة المعايير لم تعتمد المعيار المقترح لأن أحد الوفود التمس المزيد من الوقت لإجراء تقييم معمق يتناول الأثر المحتمل الذي يمكن أن يحدثه المعيار المقترح تنفيذه ولعقد مشاورات داخلياً ومع عملائه. وطلبت اللجنة من فرقة العمل المعنية بتوحيد الأسماء إعادة النظر في مسودة المعيار ومواصلة تحسينه حسب الحاجة. (انظر (ي) الفقرات من 85 إلى 91 من الوثيقة (CWS/12/29)).

6. وفي الدورة نفسها المعقودة في عام 2024، التمس لجنة المعايير أيضاً من المكتب الدولي تنظيم حلقة عمل بشأن تنقية بيانات الأسماء يمكن لجميع الأطراف المهتمة حضورها. وشجعت لجنة المعايير الدول الأعضاء والمراقبين أيضاً على دعم المكتب الدولي من خلال المشاركة بحلقة العمل والترويج لها بنشاط. (انظر (ي) الفقرة 92 من الوثيقة (CWS/12/29)).

7. وأشارت لجنة المعايير في دورتها الثانية عشرة إلى كون معايير الويبو توصيات قائمة على أفضل الممارسات. فيجوز لمكاتب الملكية الفكرية والأطراف الفاعلة في مجال الملكية الفكرية تنفيذ معايير الويبو بالوتيرة والطرق التي تناسبها وبما يتماشى مع احتياجاتها بصورة عامة. ويجوز في حالات استثنائية أن توافق مكاتب الملكية الفكرية على تنفيذ أحد المعايير كما هو وفي آن واحد، كما هو الحال بالنسبة لمعيار الويبو ST.26، أو أن تنفذه بطريقة منسقة كما هو الحال بالنسبة لمعيار الويبو ST.92. وتكون كل معايير الويبو قابلة للتحسين في وقت لاحق مع مراعاة الخبرات والطلبات الجديدة الواردة من مكاتب الملكية الفكرية فيما يتعلق بالتنفيذ، وحرصاً على بقاء المعايير عملية وفعالة واستمرارها في الاستجابة للاحتياجات المتغيرة. ونوهت لجنة المعايير أيضاً إلى كون معيار الويبو ST.93 يكتسي طابعاً عاماً بالنسبة لسائر معايير الويبو، وأنه يجوز لمكاتب الملكية الفكرية أن تنفذه على الفور، أو تدريجياً، أو ألا تنفذه على الإطلاق إذا اعتُبرت النظم القائمة في وقت اعتماد المعيار كافية. ويكون المعيار قابلاً للتحسين في وقت لاحق مع مراعاة التعليقات الواردة من مكاتب الملكية الفكرية فيما يتعلق بتجارب التنفيذ كما هو معتاد.

8. وعملاً بالقرار الذي اتخذته لجنة المعايير في دورتها الثانية عشرة، نظم المكتب الدولي حلقة عمل بشأن توحيد الأسماء، وعُقدت في 12 مايو/أيار 2025. وقامت فرقة العمل المعنية بتوحيد الأسماء في اجتماعها اللاحق، الذي عُقد في 13 مايو 2025، بتحليل نتائج حلقة العمل وأجرت مناقشة نهائية بغية إعداد المسودة النهائية التي تتناول "التوصيات بشأن تنقية بيانات الأسماء". ويمكن الاطلاع على المزيد من التفاصيل حول سيرة فرقة العمل والتقدم الذي أحرزته منذ الدورة الأخيرة للجنة المعايير في الوثيقة CWS/13/7.

9. وفيما يتعلق بنظام النقل الحر في الذي تستخدمه مكاتب الملكية الفكرية، أبلغت فرقة العمل لجنة المعايير في دورتها الثانية عشرة أن المكاتب الأعضاء في فرقة العمل مدعوة إلى تزويد الأمانة بنظم النقل الحر في التي تعمل بها، إذا كانت متاحة. وهذا يسمح للعلاء ومكاتب الملكية الفكرية الأخرى بالاطلاع على النظم التي تستخدمها المكاتب العاملة بلغات وحروف مختلفة. ومن شأن هذا أيضاً أن يسهل التواصل بصورة منسقة وفعالة بين مكاتب الملكية الفكرية وعملائها بدلاً من تعديل قواعد البيانات القائمة. وبغية التشجيع على بلوغ ذلك الهدف، تُحثّ مكاتب الملكية الفكرية على مشاركة أنظمتها، إذا كانت متاحة.

اقترح لمعيار جديد

#### الأهداف

10. تواجه مكاتب الملكية الفكرية صعوبات في تحديد أفراد الأسر داخل أسر البراءات إذ قد تُستخدم أسماء مختلفة للمودعين داخل أسرة البراءات نفسها. وإضافة إلى ذلك، قد تتضمن أسماء المودعين أخطاء إملائية أو مطبعية. وعلاوة على ذلك، من المعترف به على نطاق واسع أن الحصول على بيانات واضحة لأسماء المودعين أمر مستحسن للأغراض الإحصائية.

#### المنفعة

11. بإنشاء معيار لتنقية بيانات أسماء المودعين في سياق الملكية الفكرية يمكن تحقيق فوائد جمة من شأنها أن تعزز الكفاءة التشغيلية وسلامة البيانات. وهذا يتيح تتبع أصول الملكية الفكرية وإدارتها بشكل فعال، حتى في مختلف الولايات القضائية ومع مرور الزمن. وبفضل بيانات الأسماء النظيفة يصبح من الأسهل إقامة روابط موثوق بها بين مجموعات البيانات المتباينة، ودعم الأنشطة على غرار إدارة المحفوظات، والامتثال القانوني، وتتبع الملكية، والعناية الواجبة. وتساعد البيانات النظيفة والموحدة على تحسين قابلية للبحث، وتقليل التكرار، وتسمح بإجراء تحليلات أكثر دقة وتعزز إمكانية جمع المعلومات المتعلقة بالمنافسين. بالإضافة إلى ذلك، فهي تتيح الأتمتة وتدعم تطوير نماذج الذكاء الاصطناعي والتعلم الآلي الرامية إلى تحليل الملكية الفكرية. والأهم من ذلك أن هذا المعيار يساهم في جهود الموازنة العالمية من خلال حل القضايا المتعلقة باختلاف صيغ الأسماء والاختصارات والاختلافات بين اللغات المتعددة. وفي نهاية المطاف، فإن المعيار الرامي إلى تنقية بيانات أسماء المودعين يشكل خطوة أساسية في المساعي إلى تحسين عمليات اتخاذ القرار، والتخفيف من وطأة المخاطر، وإدارة الملكية الفكرية بشكل استراتيجي.

## النطاق

12. يقدم المعيار المقترح توصيات عامة بشأن تلقي بيانات الأسماء النظيفة ومعالجتها وتنقيتها ونشرها. ولا يقدم هذا المعيار توصيات بشأن نُهج محددة تُتبع في تنقية البيانات أو تحديد الأسماء أو تحويلها مثل النقل الحرفي أو النسخ أو الترجمة. وبالإضافة إلى ذلك، لا يتيح المعيار أي توجيهات بشأن النُهج المتبعة في توحيد الأسماء مثل اختيار الخوارزميات، أو المكان والزمان لتطبيق عمليات التحويل، أو الوثيرة، أو استراتيجيات الدمج.

### التغييرات التي أُجريت على المسودة الأخيرة

13. على ضوء المناقشات بشأن اقتراح تنقية بيانات الأسماء وبناءً على نتائج حلقة العمل المعنية بتوحيد الأسماء، اضطلعت فرقة العمل بمراجعة المسودة الأصلية التي تتناول المبادئ التوجيهية المقترحة (انظر ملحق للوثيقة CWS/12/16 Rev). وقد أُجريت التغييرات التالية:

- حذف المرفق: حُذف مرفق المعيار المقترح مراعاةً للشواغل التي عبّر عنها خلال حلقة العمل المعنية بتوحيد الأسماء بشأن عدم اكتماله واحتمالية التحيز غير المقصود.
- تغييرات تحريرية في عدد من الفقرات لزيادة الوضوح، واستجابةً للردود والتحليلات الواردة في التعليقات خلال ورشة العمل، بما في ذلك:

- في الفقرة 11: أوصى المشاركون في الورشة بتعزيز التوجيهات بشأن إدراج أسماء المودعين بحروفها الأصلية، لا سيما حين يتعلق الأمر بالنقل الحرفي. وجرت الإشارة إلى أن حذف الحروف الأصلية قد يؤدي إلى تناقضات كبيرة بين أعضاء أسرة البراءات بسبب اختلاف نظم النقل الحرفي. فعلى سبيل المثال، يمكن نقل الاسم "Чугаев" حرفياً إلى الأحرف اللاتينية بالطرق التالية: "Tschugaeff" أو "Tchugaev" أو "Tchougavev" أو "Cugaev" أو "Chugaev"، وذلك يصعب عمليات الربط الدقيق للسجلات ذات الصلة ببعضها البعض أو عمليات تحديد المودع الصحيح. وقد خضعت الفقرة 11 للمراجعة بناءً على ذلك.

- في الفقرة 22: تجنباً لاحتمال حدوث أي تأويل خاطئ، حُذفت الجملة الثانية. وقد أُشير إلى أن استبدال رموز التعريف الحالية للمودعين بأرقام فريدة من إنشاء النظام في مرحلة النشر قد يتسبب بالتباس لدى المستخدمين ويقوّض اتساق عملية تتبع المودعين. وقد أُدخلت تحديثات على الفقرة 22 تماشياً مع هذه المشاغل.

14. ويرد المعيار المقترح في مرفق هذه الوثيقة، وترد جميع التغييرات المدخلة منذ المسودة الأخيرة. ويشير النص المحذوف إلى الأجزاء التي أُزيلت، والنص المسطر إلى الأجزاء التي أُضيفت.

15. واقتُرِح الاسم التالي لمعيار الويبو الجديد:

"معيار الويبو ST.93 - التوصيات بشأن تنقية بيانات الأسماء"

16. وإذا اعتمدت لجنة المعايير المعيار الجديد في دورتها الحالية، فمن المقترح أن تطلب من الأمانة نشر هذه التوصيات في [الجزء 3 من دليل الويبو](#).

مقترح لنشر نظم النقل الحرفي التي تستخدمها مكاتب الملكية الفكرية في الجزء 7 من دليل الويبو

17. بعد إجراء المناقشات حول استخدام نظم النسخ الحرفي، اقترح جمع النظم التي تستخدمها مكاتب الملكية الفكرية لنسخ أسماء العملاء نسخاً حرفياً ونشرها على موقع الويبو، في الجزء 7 من دليل الويبو بشأن المعلومات والوثائق المتعلقة بالملكية الفكرية. ومن المتوقع أن يدعم هذا المورد المركزي دقة عمليات التحقق والتأكد من صحة نسخ أسماء العملاء في مختلف الولايات القضائية.

18. ومن المقترح تشجيع مكاتب الملكية الفكرية على تقديم العناوين الشبكية لنظم النسخ الحرفي التي تتبعها إلى الأمانة ليتم نشرها في الجزء 7 من دليل الويبو.

19. واقترحت الأمانة إدراج قسم فرعي جديد في الجزء 7 من دليل الويبو، لئُنشر فيه نظم النسخ الحرفي التي تستخدمها مكاتب الملكية الفكرية تحت عنوان: "نظم النقل الحرفي التي تستخدمها المكاتب".

20. إن لجنة المعايير مدعوة إلى الاضطلاع بما يلي:

(أ) الإحاطة علماً بمضمون هذه الوثيقة ومرفقها؛

(ب) والنظر في اسم معيار الويبو الجديد، على النحو المذكور في الفقرة 15 أعلاه، والموافقة عليه؛

(ج) والنظر في معيار الويبو الجديد ST.93 واعتماده، على النحو المشار إليه في الفقرات من 10 إلى 15 أعلاه، وعلى النحو الوارد في مرفق هذه الوثيقة؛

(د) والطلب من الأمانة نشر معيار الويبو الجديد ST.93 في الجزء 3 من دليل الويبو، على النحو المشار إليه في الفقرة 16 أعلاه؛

(هـ) والطلب من الأمانة إصدار تعميم يدعو المكاتب إلى تقديم نظم النقل الحرفي الخاصة بها، ونشر هذه النظم في الجزء 7 من دليل الويبو، على النحو المذكور في الفقرات من 17 إلى 19 أعلاه.

[يلي ذلك المرفق]

**WIPO STANDARD ST.93**

RECOMMENDATIONS ON THE DATA CLEANING OF NAMES

*Proposal presented for approval by the Committee on WIPO Standards (CWS)  
at its thirteenth session*

TABLE OF CONTENTS

Introduction .....	2
DEFINITIONS .....	2
INTAKE .....	2
TRANSFORMATION OF NAMES .....	3
VALIDATION AND DISAMBIGUATION.....	3
MAINTENANCE .....	3
PUBLICATION AND DATA EXCHANGE .....	4
STATISTICAL PURPOSES .....	4
References .....	4
<u>ANNEX .....</u>	<u>1</u>
<u>Transliteration examples:.....</u>	<u>1</u>
<u>Transcription examples:.....</u>	<u>2</u>

## WIPO STANDARD ST.93

### RECOMMENDATIONS ON NAME DATA CLEANING

*Proposal presented for adoption by the Committee on WIPO Standards (CWS)  
at its thirteenth session*

#### INTRODUCTION

1. This Standard provides general recommendations on the intake, processing, cleaning, and publication of clean name data. This Standard does not provide recommendations on details in relation to approaches to data cleaning, name localization or transformation, such as transliteration, transcription or translation, or approaches to name standardization, such as selection of algorithms, where and when transformations are applied, frequency, or merging strategies. Decisions on ~~these~~ such details will vary greatly depending on the party applying the ~~m~~ approach, the purpose of transformations, and the quickly evolving nature of matching algorithms.

2. WIPO Standard ST.20 should be referred to for recommendations to produce indexes to patent documents giving names of applicants and other customers, and to promote a uniform presentation of names occurring in name indexes as well as a uniform method of ordering the names in the index itself.

#### DEFINITIONS

3. In the context of this document:

- (a) "IPO" refers to an Intellectual Property Office, which manage the application and registration process for intellectual property rights.
- (b) "Customer data" means data on applicants, registrants, owners, legal representatives, or other parties held by an IPO in connection with an IP right, application, registration, or other instrument. This standard is primarily concerned with customer name data: personal names, business names, and related information such as city, address, or email that can be used to disambiguate potential name matches.
- (c) "Clean data" means data that is accurate, consistent and reliable. As the degree of cleanness in a large complex data set is difficult to measure, various metrics may be used as proxies for cleanness or related properties, such as fitness for purpose.
- (d) "Transliteration" means the mapping of a source language character(s) to a target language (phonetic) character(s).
- (e) "Transcription" means the mapping of a source language character/logogram/syllable/phoneme to something that corresponds to the sound in the respective system of the target language.
- (f) "Translation" means representing the meaning of a word or concept in the source language with something that corresponds to ~~that~~ meaning in the target language.

#### INTAKE

4. IPOs may provide the ability for customers to create and manage electronic customer records containing published name information: personal names, business names, names of legal representatives, and related information such as city, address, or email.

5. IPOs should allow a customer record to be associated with multiple applications or registrations for IP rights, so that customers may reuse the same name information for multiple applications or registrations and update their name information in one place.

6. IPOs may provide ~~a~~ form(s) which for customers to use to request the IPOs to create or change their name or related information. IPOs may ~~also~~ allow customers to enter and update their name or related information themselves, or may require a designated party, such as employees, contractors, or an external service to enter and update customer records at the customer's request.

7. Multiple records for one customer may be created and managed by different entities, such as different legal representatives. IPOs should consider this when designing their customer record systems, as multiple records for a single customer may contain slight variations ~~of~~ on the same data or be updated at different times by different representatives.

8. IPOs may support entry of the customer's name in native characters of the customer's language, in addition to the customer's name in the language(s) of operation for an IPO, which should be stored using UTF-8<sup>1</sup> encoding. For instance, an IPO that works in English could allow separate fields for an applicant name in English and the original applicant name in Korean.

9. IPOs may optionally use identification ~~codes numbers~~ to identify customers. Identification ~~numbers codes~~ may be created by the IPO or ~~used taken~~ from an external source, such as a registered business number or passport number. Identification ~~numbers codes~~ alone do not resolve issues with clean customer data, such as duplicate entries, name changes, and outdated or incorrect information. IPOs using identification ~~numbers codes~~ should continue to pay attention to and address the considerations in other parts of this Standard.

#### TRANSFORMATION OF NAMES

10. For data exchange and processing, including the receipt of international applications or registrations, IPOs may consider the name transformation (~~see the Annex to this document~~). It is recommended that IPOs should send and receive name data using UTF-8 encoding.

11. It should be noted that the localization or conversion of customer names is extremely ~~prone to error prone~~ as there are no generally accepted or uniformed standards. For localization or transformation of names, there are three ways referred to in this Standard: transliteration, transcription and translation. If IPOs transliterate, transcribe or translate ~~characters names~~ from one language ~~or character set~~ (such as ~~Greek Korean or Latin~~) to another (such as English ~~or Cyrillic~~), they should publish their scheme of transliteration, transcription or translation. If IPOs transform a customer's name, it is recommended to retain the applicant's name in the native characters or language(s), in accordance with paragraph 8 of this Standard. ~~The~~ Transliterated, transcribed or translated documents, or parts of ~~the~~ documents, should be made available to the customer for review, and customers should have ~~the opportunity a way~~ to submit corrections if the transliteration, transcription or translation is flawed.

12. Reverse transliteration should be avoided, if possible; instead, it is recommended ~~to use that~~ the original name ~~should be used~~ instead. For instance, an application filed by "Phony Corp" ~~in Latin characters~~ might be transliterated to Greek characters as "Φονι Κορπ" in an IPO system, and on publication might be reverse transliterated from Greek back to Latin characters as "Foni Corp", leading to mismatches. ~~Examples of common issues arising from reverse, or re-transliteration, re-transcription or re-translation are available in the Annex to this Standard.~~

#### VALIDATION AND DISAMBIGUATION

13. Validation and disambiguation approaches should be designed to meet specific objectives, either administrative or statistical, and appropriate methods applied given the objectives. Approaches to name matching and disambiguation should be appropriately scoped and risk assessed ~~given in the light of~~ their design objective to ensure appropriate levels of disambiguation are achieved for the use case.

14. IPOs may choose to perform validation of submitted customer information, including automated checks. The validation of such results should be made available communicated to the customer. ~~Where necessary, any corrections must be approved by the customer prior to implementation in the system and corrections accepted by the customer if needed, including ways to bypass an automated validation mechanism, in case it provides incorrect or incomplete results. Furthermore, provisions should be made to allow bypassing the automated validation mechanism in instances where it produces inaccurate or incomplete results.~~

15. IPOs attempting to disambiguate name records (i.e., find duplicate entries) may wish to consider more than just ~~the~~ customer names. Names are not inherently unique. For example, there may be multiple individuals named "John Smith" or multiple companies named "Data Corp". Comparing related data points such as city, post code, birthdate, or other information, where available, can increase the likelihood of successful matches.

16. Any validation or disambiguation process initiated by ~~the an~~ IPO that ~~could~~ potentially ~~could~~ have legal effects, such as correcting or standardizing the name of the registered owner of an IP right, should be confirmed by the customer before the change is made in the IPO's system.

#### MAINTENANCE

17. IPOs should develop a strategy to periodically clean data in customer name databases, including searching for and attempting to resolve duplicate records, i.e., multiple records for the same ~~entity customer~~. In some instances, ~~the~~ duplicates may be merged or combined, for instance, records with slight unintentional differences in spelling such as "ABC Corp" and "ABC Corp." ~~could be consolidated~~. In other instances, maintaining separate records might be preferable. Each IPO should decide what approach ~~best~~ fits ~~best for their its~~ own name record management system. The strategy may include ~~the~~

---

<sup>1</sup> UTF-8 is an encoding system for Unicode.

involvement of the ~~concerned~~ customers concerned of with the records in the data cleaning process and the responsibility of the cleaned data.

18. IPOs should provide a mechanism for customers to update their name information on multiple applications or IP rights by entering the information once. For instance, this could be achieved by associating each application or IP right with a single customer record containing name information, or by allowing customers to select multiple applications or IP rights and submit one instance of updated name information to be applied to all of them.

19. IPOs may designate someone to be responsible for data cleaning ~~data~~ issues, including the development of metrics for measuring ~~clean~~ data cleanness, regular monitoring and reporting of those metrics, and taking action to improve customer data when needed.

#### PUBLICATION AND DATA EXCHANGE

20. IPOs should make available updates to name information that are ~~made submitted~~ after an IP right document has been published. For instance, if "ABC Corp" changes ~~their~~ its name to "XYZ Corp" in ~~their~~ its customer record, then the name "XYZ Corp" should be associated with the IP right in online publications. The original name may also appear ~~on the~~ in published IP ~~right documents~~, ~~according depending to the~~ legal requirements of the IPO.

21. If an IPO ~~has holds~~ other forms of a customer's name, such as an original name expressed using native characters, these should be included in published ~~data-IP documents~~ and ~~the data documents~~ exchanged with other IPOs.

22. If an IPO uses identification ~~numbers codes~~ to identify entities, the ~~numbers codes~~ should be included in published data and data exchanged with other IPOs, ~~unless~~. If the identification ~~numbers codes~~ are sensitive and cannot be shared, ~~then the IPO should indicate which customer data uses these identification numbers, such as by replacing the sensitive numbers with generated unique numbers for publication.~~

#### STATISTICAL PURPOSES

23. For statistical purposes, IPOs may attempt to match customer data with variations in customer names, or other fields, to achieve counts that are more accurate. In such cases, IPOs should publish their matching strategy or algorithm along with the statistical results so that others can understand the methodology used.

#### REFERENCES

24. References to the following Standard are of relevance to this Standard:

WIPO Standard [ST.20](#) Recommendation for the pPreparation of name indexes to patent documents

ANNEX

DIFFERENT MEANS OF NAME TRANSFORMATION

Although transliteration and transcription are different concepts from a linguistic perspective, the result is usually very similar for character-based writing systems. However, transcription provides a more practical result, because only standard characters from the target language are required for the conversion.

As English is a language that is adopted as a common language between speakers whose native languages are different, it is generally overlooked that transcription is rarely standardized between any pair of languages. In the best case there are official definitions for [xx] → [en] leading to the assumption that [xx] → [en] → [yy] is equal to [xx] → [yy], which is usually not correct.

TRANSLITERATION EXAMPLES<sup>2</sup>:

Figure 1 shows below an example of letter correspondence and remarks regarding this transliteration.

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

Figure 1: Transliteration example

<sup>2</sup>Machine Transliteration Survey

<https://www.researchgate.net/figure/Transliteration-examples-in-four-language-pairs-Letter-correspondence-shows-how-the-fig1-220566444>

**TRANSCRIPTION EXAMPLES:**

Shown below are examples where transcription can lead to inaccuracies:

[ru]: Ш → [de]: sch<sup>3</sup>

[ru]: Ш → [en]: sh

[ko]: ㅟ → [de]: ja<sup>4</sup>

[ko]: ㅟ → [en]: ya

[gr]: Ω → latin: O<sup>5</sup>

[da]: /E → [de]: Ä or AE, [en]: AE<sup>6</sup>

**TRANSLATION EXAMPLES:**

In the first example, it is clear that the direct translation can lead to issues:

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

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

A second example below, which demonstrates typical borderline cases of the Romanization of a Chinese company name shown in Figure 2 are:

— [zh]: 北京东土科技股份有限公司 → [en] transliterated (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] transcribed (pinyin): beijing dongtu keji gufen youxian gongsi

— [zh]: 北京东土科技股份有限公司 → [en] translated (English): Beijing, China Science and Technology Joint stock Limited Company

— [zh]: 北京东土科技股份有限公司 → in reality : 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: Romanization of Chinese company name**

[End of Annex to the proposed Standard and of  
Standard]

[End of the proposed Standard]

[End of Annex and the document]

<sup>3</sup> [https://de.wikipedia.org/wiki/Kyrillisches\\_Alphabet#Russisch](https://de.wikipedia.org/wiki/Kyrillisches_Alphabet#Russisch)

<sup>4</sup> [https://de.wikipedia.org/wiki/Koreanisches\\_Alphabet](https://de.wikipedia.org/wiki/Koreanisches_Alphabet)

<sup>5</sup> [https://en.wikipedia.org/wiki/Romanization\\_of\\_Greek](https://en.wikipedia.org/wiki/Romanization_of_Greek)

<sup>6</sup> [https://en.wikipedia.org/wiki/Dania\\_transcription](https://en.wikipedia.org/wiki/Dania_transcription)