E

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

CWS/13/15
ORIGINAL: ENGLISH
DATE: SEPTEMBER 8, 2025

Committee on WIPO Standards (CWS)

Thirteenth Session
Geneva, November 10 to 14, 2025

PROPOSAL FOR A NEW WIPO STANDARD SUPPORTING THE DATA CLEANING OF NAMES

Document prepared by the Name Standardization Task Force Co-Leaders

SUMMARY

1. The Name Standardization Task Force hereby presents a final draft of a new WIPO standard on data cleaning of names for consideration and adoption at the thirteenth session of the Committee on WIPO Standards (CWS).

BACKGROUND

2. At its eleventh session in 2023, the CWS approved the revised description of Task No. 55 which reads:

"Prepare a proposal for future actions aimed at achieving the standardization of names in Intellectual Property (IP) documents, with the view to developing a WIPO standard to assist IP offices in providing a better "quality at source" in relation to names"

(see paragraphs 75 to 78 of document CWS/11/28).

3. At the same session in 2023, the CWS considered a new set of guidelines proposed by the Name Standardization Task Force to support cleaning applicant names. The CWS agreed to use the term "recommendations" instead of "guidelines" in the name of the proposed new WIPO standard because it is clearer. The CWS also noted the name proposed by the Secretariat, "WIPO Standard ST.93" for recommendations on the data cleaning of names (see paragraph 135 of document CWS/11/28).

4. However, the CWS did not adopt the proposed standard and referred it back to the Task Force for further discussion and improvement. The CWS also noted that the Secretariat would explore the possibility of publishing a collection of transliteration tables on the WIPO website. (See paragraphs 136 and 137 of document CWS/11/28.)

5. At its twelfth session in 2024, the CWS reviewed an improved draft of the proposed WIPO Standard for the data cleaning of names, as presented by the Name Standardization Task Force. Several delegations expressed their support for the proposed standard. However, the CWS did not adopt it, because one delegation requested additional time to thoroughly assess the potential impact of the proposed standard to implement and to conduct consultations both internally and with its customers. The CWS requested that the Task Force revisit and continue to improve the draft standard as necessary. (See paragraphs 85 to 91 of document CWS/12/29.)

6. At the same session in 2024, the CWS requested that the International Bureau organize a workshop on the data cleaning of names in 2025, which would be open to any interested parties. The CWS also encouraged its members and observers to support the International Bureau by actively promoting and participating in the workshop. (See paragraph 92 of document CWS/12/29.)

7. At its twelfth session, the CWS noted that WIPO Standards are recommendations on the basis of best practices. IP offices and IP industry implement WIPO Standards with their own pace and their own way upon their needs in general. Exceptionally IP offices agree on the implementation of a certain Standards as-is at the same time, for examples, WIPO Standard ST.26 or in a harmonized way such as WIPO Standard ST.92. All WIPO Standards can be improved later considering implementation experiences of IP offices or new demands to ensure that they remain practical, effective and responsive to evolving needs. The CWS also noted that the proposed WIPO Standard ST.93 has a general nature of WIPO Standards and IP offices may implement it immediately, gradually, or not at all if existing systems suffice once it is adopted. The Standard can be also improved later considering feedback from implementation experiences of IP offices as usual.

8. Following the decision of the CWS made at its twelfth session, the International Bureau organized the workshop on name standardization, which was held on May 12, 2025. At its subsequent meeting held on May 13, 2025, the Name Standardization Task Force analyzed the outcomes of the workshop and held a final discussion to prepare the final draft of the “Recommendations on Name Data Cleaning”. Details of the history of the Task Force and the progress made since the last session of the CWS can be found in document CWS/13/7.

9. With regard to the transliteration scheme used by IP offices, the Task Force informed the twelfth session of the CWS that the Task Force Offices were invited to provide their transliteration schemes to the Secretariat, if available. This would allow customers and other IP offices to consult the schemes used by offices operating in different languages and scripts. This would facilitate consistent and effective communication between IP offices and their customers rather than modifying existing databases. To support this goal, IP offices are encouraged to share their schemes, if available.

PROPOSAL FOR A NEW STANDARD

Objectives

10. IP offices experience difficulties identifying family members within a patent family because different applicant names may be used within the same patent family. Additionally, the names of applicants may include spelling or typographical errors. Furthermore, the desirability of clean applicant name data for statistical purposes is widely recognized.

Benefit

11. Establishing a standard for cleaning applicant name data in the context of IP provides significant benefits that enhance both operational efficiency and data integrity. This enables effective tracking and management of IP assets, even across jurisdictions and over time. Cleaned name data facilitates reliable linkage between disparate datasets, supporting activities such as portfolio management, legal compliance, ownership tracking and due diligence. Clean, standardized data improves searchability, reduces redundancy and allows for more accurate analytics and competitive intelligence. Additionally, it enables automation and supports the development of artificial intelligence and machine learning models for IP analysis. Importantly, such a standard contributes to global harmonization efforts by resolving issues related to name variations, abbreviations and multilingual discrepancies. Ultimately, a data cleaning standard for applicant names is a foundational step towards improving decision-making, mitigating risk, and managing IP strategically.

Scope

12. The proposed Standard provides general recommendations on the intake, processing, cleaning and publication of clean name data. This standard does not provide recommendations on specific approaches to data cleaning, name localization or transformation, such as transliteration, transcription and translation. Additionally, it does not offer guidance on approaches to name standardization, such as the selection of algorithms, the location and timing of the application of transformations, their frequency, or merging strategies.

Changes made to the last draft

13. In light of discussions regarding the proposal for name data cleaning and the outcome of the Name Standardization Workshop, the Task Force revised the original draft of the proposed guidelines (see Annex to document CWS/12/16 Rev.). The following changes were made:

- Removal of the Annex: The Annex to the proposed standard was removed based on concerns about its incompleteness and the potential for unintended bias expressed at the Name Standardization Workshop.

- Editorial changes to a number of paragraphs to improve clarity, reflecting feedback and analysis of the comments received during the workshop, including:

    - In paragraph 11: Participants at the workshop recommended strengthening the guidance on including applicant names in their native characters, especially when transliteration is involved. It was noted that omitting native characters could result in significant inconsistencies among patent family members due to various transliteration systems. For instance, the name “Чугаев” could be transliterated in Latin script as “Tschugaeff,” “Tchugaev,” “Tchougaev,” “Cugaev,” or “Chugaev,” complicating efforts to accurately link related records or identify the correct applicant. Paragraph 11 has been revised accordingly.

    - In paragraph 22: To avoid potential misinterpretation, the second sentence was removed. It was suggested that replacing an established applicant identification code with a system-generated unique number at the publication stage could cause confusion for users and undermine the consistency of applicant tracking. Paragraph 22 has been updated to reflect this concern.

14. The proposed Standard is included as Annex to this document and all changes made since the last draft are tracked. Strike-through text indicates deletion and underlined text indicates addition.

15. The following name is proposed for the new WIPO standard:

"WIPO Standard ST.93 – Recommendations on the data cleaning of names"

16. If the new standard is adopted by the CWS at the current session, it is proposed that the CWS request the Secretariat to publish these recommendations in Part 3 of the WIPO Handbook.

PROPOSAL FOR PUBLICATION OF TRANSLITERATION SCHEMES USED BY IP OFFICES IN THE WIPO HANDBOOK PART 7

17. Following discussions on the use of transliteration schemes, it is proposed to collect and publish the transliteration schemes used by IP offices for customers names on WIPO website, under Part 7 of the WIPO Handbook on Intellectual Property Information and Documentation. It is expected that this centralized resource will support accurate cross-referencing and verification of customer names across jurisdictions.

18. It is proposed that IP offices are encouraged to provide the URL of their transliteration schemes to the Secretariat for the publication on the WIPO Handbook Part 7.

19. The Secretariat proposes to create a new subsection under Part 7 of the WIPO Handbook, where the transliteration schemes used by IP offices will be published, entitled: *"Transliteration* scheme*s used by Offices"*

20. The CWS is invited to:

(a) take note of the content of the present document and the Annex to the present document;

(b) consider and approve the name of the new WIPO Standard as indicated in paragraph 15 above;

(c) consider and adopt the new WIPO Standard ST.93 as referred to in paragraphs 10 to 14 above and as reproduced in the Annex to the present document;

(d) request the Secretariat to publish the new WIPO Standard ST.93 in Part 3 of the WIPO Handbook as referred to paragraph 16 above; and

(e) request the Secretariat to issue a circular inviting Offices to provide their transliteration schemes and publish the provided transliteration schemes in Part 7 of the WIPO Handbook as referred to in paragraphs 17 to 19 above.

[Annex follows]

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

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 ~~those~~ 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 represent~~s~~ing~~s~~ the meaning of a word or concept in the source language with something that corresponds to th~~e~~at~~e~~ 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 IPO~~s~~ 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 [1]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 uniform~~ed~~ 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 t~~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 ~~V~~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~~

[1] UTF-8 is an encoding system for Unicode.

involving~~ement 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 p~~P~~reparation 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[2]:

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

Source and Target words	Letter Correspondence				Description
English to Persian					
John /dʒɒn/	J	o	h	n	*h* is a silent letter (no sound
	|	|		|	is associated to the letter)
جان /dʒɒn/	ج	ا		ن	and is not transliterated
Arabic to English					
نجيب /nædʒi:b/	ن	ج	ی	ب	short vowel /æ/ on N is
	|	|	|	|	normally not written in
Najib /nædʒi:b/	Na	j	i	b	Arabic script
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 /ˈædəm/	A	d	a	m	the second "a" is not
	|	|		|	transliterated in Hindi
अदम /ˈædəm/	अ	द		म	

Figure 1: Transliteration example

[2] 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[3]~~

~~[ru]: Ш → [en]: sh~~

~~[ko]: ㅑ → [de]: ja[4]~~

~~[ko]: ㅑ → [en]: ya~~

~~[gr] : Ω → latin: O[5]~~

~~[da]: Æ → [de]: Ä or AE, [en]: AE[6]~~

~~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 Силовы́е маши́ны → [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]

[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~~