

Committee on WIPO Standards (CWS)

Fourteenth Session
Geneva, November 23 to 27, 2026

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 presents the final draft of a new WIPO standard on the data cleaning of names for consideration and adoption at the fourteenth 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 a view to developing a WIPO standard to assist IP offices in ensuring 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 the cleaning of applicant names. The CWS agreed to use the term “recommendations” instead of “guidelines” in the title of the proposed new WIPO standard. The CWS also noted the title 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 Name Standardization 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 on 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 assess thoroughly the potential impact of implementing the proposed standard and to conduct consultations both internally and with its customers. The CWS requested that the Task Force revisit the draft standard and continue to improve it, 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, open to all 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 based on best practices. In general, IP offices and the IP industry implement WIPO Standards at their own pace and in a manner suited to their needs. In exceptional cases, IP offices agree to implement a particular Standard as adopted and within the same timeframe, as in the case of WIPO Standard ST.26, or in a harmonized manner, as in the case of WIPO Standard ST.92. All WIPO Standards may subsequently be improved in light of implementation experience gained by IP offices or emerging needs, to ensure that they remain practical, effective and responsive to evolving requirements. The CWS also noted that the proposed WIPO Standard ST.93 shares the general nature of WIPO Standards and that, once adopted, IP offices may implement it immediately or gradually, or may decide not to implement it where existing systems are sufficient. As with other WIPO Standards, it may subsequently be improved in light of feedback from implementation experience.
8. Following the decision taken by the CWS at its twelfth session, the International Bureau organized a 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". Further details on the history of the Task Force and the progress made since the previous session of the CWS are provided in document CWS/13/7.
9. With regard to the transliteration schemes used by IP offices, the Task Force informed the twelfth session of the CWS that Task Force members had been invited to provide their transliteration schemes to the Secretariat, where available. This would enable customers and other IP offices to consult the schemes used by Offices operating in different languages and scripts, thereby facilitating consistent and effective communication between IP offices and their customers without requiring modifications to existing databases. To support this objective, IP offices are encouraged to share their schemes, where available.
10. At its thirteenth session, the CWS reviewed an improved draft of the proposed WIPO Standard on the data cleaning of names but did not adopt the proposed WIPO Standard ST.93, as consensus was not reached. The CWS requested that the Name Standardization Task Force revisit and improve the draft WIPO Standard ST.93 and present a final proposal for consideration at the next session of the Committee (see paragraphs 105 and 106 of document CWS/13/33).

PROPOSAL FOR A NEW STANDARD

Objectives

11. The objective of the proposed Standard is to establish general recommendations for the cleaning of applicant name data by IP offices, with a view to improving the consistency, accuracy and reliability of such data. The proposed standard is intended to support the identification of patent-family members, statistical analysis and other uses of IP information.

Benefit

12. The proposed standard would provide IP offices with a common framework for improving the quality, consistency and usability of applicant name data. Its implementation would facilitate the identification and linking of records across datasets, jurisdictions and time periods, thereby supporting more efficient searching, portfolio management, statistical analysis and exchange of IP information. It would also help to reduce duplication and inconsistencies, support the automation of data-processing activities and the development of data-driven tools, and promote greater harmonization in addressing name variations, abbreviations and differences arising from the use of multiple languages and scripts.

Scope

13. The proposed standard sets out general recommendations for the intake, processing, cleaning and publication of applicant name data by IP offices. Its scope does not extend to specific methods or technical solutions for data cleaning, name localization or name transformation, including transliteration, transcription and translation. Nor does it extend to particular approaches to name standardization, including the selection of algorithms, the stage at which transformations are applied, the frequency of their application or strategies for merging records.

Changes made to the last draft

14. Following the decision taken by the CWS at its thirteenth session, the members of the Name Standardization Task Force were invited to review and comment on the draft WIPO standard ST.93 presented at that session. On the basis of the comments received, the following changes were made to the draft:

- Adding definitions of “Validation” and “Disambiguation” to paragraph 3; and
- Revising paragraph 11 to clarify the distinction between languages and character sets.

15. The Task Force also considered a proposal to include a statement in the Standard clarifying its voluntary nature exceptionally. Considering the Task Force’s discussion, and ensuring consistency with other WIPO Standards, the Secretariat proposes the following statement to be exceptionally included in the editorial note of the International Bureau:

“Implementation of this Standard is voluntary and non-binding. Each Office may determine, according to its circumstances, whether and when to implement it, either fully or partially, or not at all.”

16. The proposed Standard is reproduced in the Annex to the present document. All changes made to the draft presented at the thirteenth session of the CWS are shown as tracked changes, with deletions indicated by strikethrough and additions by underlining.

17. The following title is proposed for the new WIPO Standard:

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

18. Subject to the adoption of the new Standard by the CWS at its current session, it is proposed that the CWS request the Secretariat publish the Standard in [Part 3](#) of the WIPO Handbook for Intellectual Property Information and Documentation.

PROPOSAL FOR REVISION OF TASK NO. 55 DESCRIPTION

19. If the proposed new WIPO Standard ST.93 is adopted by the CWS, the work under the current description of Task No. 55 would be considered completed. It is proposed that the Task Force continue its work with a view to supporting the implementation of the Standard by IP offices, and maintaining and further improving the Standard in the light of experience gained through its implementation. The continued collection and exchange of information on practices relating to the cleaning of customer name data, would contribute to that work. Accordingly, the Task Force Co-leaders propose that the description of Task No. 55 be revised as follows:

“Share practices for cleaning customer name data, including the algorithms used and the purposes for which cleaned or standardized data are applied; develop practical guidance for IP offices; and maintain and update WIPO Standard ST.93, as necessary.”

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

20. Following discussions on the use of transliteration schemes, it is proposed that the transliteration schemes used by IP offices for customer names be collected and published on the WIPO website under Part 7 of the WIPO Handbook on Intellectual Property Information and Documentation. This centralized resource is expected to support accurate cross-referencing and verification of customer names across jurisdictions.

21. It is proposed that IP offices be encouraged to provide the Secretariat with the URLs of their transliteration schemes for publication in Part 7 of the WIPO Handbook.

22. The Secretariat proposes to create a new subsection in Part 7 of the WIPO Handbook, entitled *“Transliteration Schemes Used by Offices”*, in which the transliteration schemes used by IP offices would be published.

23. *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 adding of the editorial note to the proposed new WIPO standard as indicated in paragraph 15 above;*

(c) *consider and approve the name of the new WIPO standard as indicated in paragraph 17 above;*

(d) *consider and adopt the new WIPO Standard ST.93 as referred to in paragraphs 11 to 16 above and as reproduced in the Annex to the present document;*

(e) *request the Secretariat to publish the new WIPO Standard ST.93 in Part 3 of the WIPO Handbook, as referred to in paragraph 18 above;*

(f) *consider and approve the proposed new description of Task No. 55, as indicated in paragraph 19 above; and*

(g) *request the Secretariat to issue a circular inviting Offices to provide their transliteration schemes and to publish the schemes received in Part 7 of the WIPO Handbook, as referred to in paragraphs 20 to 22 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 fourteenth session*

TABLE OF CONTENTS

INTRODUCTION	2
DEFINITIONS	2
INTAKE	2
TRANSFORMATION OF NAMES	3
VALIDATION AND DISAMBIGUATION	3
MAINTENANCE	4
PUBLICATION AND DATA EXCHANGE	4
STATISTICAL PURPOSES	4
REFERENCES	4

WIPO STANDARD ST.93

RECOMMENDATIONS ON NAME DATA CLEANING

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

Editorial Note by the International Bureau:

Implementation of this Standard is voluntary and non-binding. Each Office may determine, according to its circumstances, whether and when to implement it, either fully or partially, or not at all.

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 such details will vary greatly depending on the party applying the 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.
 - (g) "Validation" means the process of conducting automated or manual checks on customer name data to confirm its accuracy, completeness, and compliance.
 - (h) "Disambiguation" means the processes of comparing names to identify and distinguish identical or similar names that may refer to different entities, so as to eliminate ambiguity in name matching.

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 forms for customers to use to request the IPO to create or change their name or related information. IPOs may 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 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-8¹ 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 to identify customers. Identification codes may be created by the IPO or taken from an external source, such as a registered business number or passport number. Identification codes alone do not resolve issues with clean customer data, such as duplicate entries, name changes, and outdated or incorrect information. IPOs using identification 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. 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 as there are no generally accepted or uniform 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 names from one language or character set (such as Korean or Latin) to another (such as English or Cyrillic) respectively, 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. Transliterated, transcribed or translated documents, or parts of documents, should be made available to the customer for review, and customers should have the opportunity to submit corrections if the transliteration, transcription or translation is flawed.

12. Reverse transliteration should be avoided, if possible; instead, it is recommended 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.

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 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 communicated to the customer. Where necessary, any corrections must be approved by the customer prior to implementation in the system. 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 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 an IPO that could potentially 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.

¹ UTF-8 is an encoding system for Unicode.

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 customer. In some instances, 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 its own name record management system. The strategy may include involving the customers concerned 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 issues, including the development of metrics for measuring 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 submitted after an IP document has been published. For instance, if “ABC Corp” changes its name to “XYZ Corp” in its customer record, then the name “XYZ Corp” should be associated with the IP right in online publications. The original name may also appear in published IP documents, depending to the legal requirements of the IPO.

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

22. If an IPO uses identification codes to identify entities, the codes should be included in published data and data exchanged with other IPOs, unless the identification codes are sensitive and cannot be shared.

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 preparation of name indexes to patent documents

[End of the proposed Standard]

[End of Annex and document]