

Committee on WIPO Standards (CWS)

Twelfth Session Geneva, September 16 to 19, 2024

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 a final draft for a new WIPO standard on supporting the data cleaning of names, for consideration and adoption at the twelfth 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. Further details regarding the history of the Task Force and the progress made since the last session of the CWS can be found in document CWS/12/8.

4. At its eleventh session in 2023, the CWS considered a new set of guidelines to support the cleaning of applicant names presented by the Name Standardization Task Force. The CWS agreed to using the term ‘recommendations’ instead of ‘guidelines’ in the name of the proposed new WIPO standard, as it considered it clearer in scope. The CWS also noted the Secretariat’s proposal of the name: “WIPO Standard ST.93” (see paragraph 135 of document CWS/11/28).

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

PROPOSAL FOR A NEW STANDARD

6. Intellectual Property Offices (IPOs) experience issues with being able to identify family members within a patent family, as different applicant names may be used within the same patent family. Additionally spelling or typographical errors may be made when entering the applicant names. The desire to have clean applicant name data for statistical purposes is well accepted.

7. The Name Standardization Task Force, under the framework of Task No. 55, has prepared a final proposal for a new WIPO standard to support the cleaning of names, to achieve clean applicant data. This proposal is provided as the Annex to the present document.

Objectives

8. These recommendations aim to provide general and high-level guidance. Variations in factors such as legal requirements, data practices, the purpose of the clean-up, the intended use of the data, resource requirements, and technical considerations mean that there is no single approach that works best across all IPOs. These recommendations reflect general practices that can be applied at any IPO to support cleaning customer name data, thereby enhancing name standardization and name matching techniques for downstream users.

Scope

9. The proposed 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.

10. The proposed Standard is structured as follows:

- Main body: defines the general recommendations for the processing of applicant names in order to achieve clean data; and
- Annex: provides examples on transliteration, transcription and translation to support the recommendations provided in the Main Body.

11. The following name for the new WIPO Standard is proposed:

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

Changes made since last draft

12. In the light of the discussions on the proposal for name data cleaning and outstanding interventions from several delegations at the eleventh session of the CWS, the Task Force revised original draft of the proposed guidelines (see Annex to CWS/11/23). The following changes were made:

- The Task Force notes that the previous definition of “clean data” as “free from errors and duplication” was problematic because it is unrealistic to guarantee that data is 100 per cent free from errors and duplications. Consequently, the Task Force agreed to modify the definition of “clean data” to: “*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.”

- In the ‘Transformation of names’ section, the Task Force agreed to change term “conversion” to “transformation” to better align with the section title and allow for a more flexible interpretation.
- In the ‘References’ section, the Task Force discussed referencing ISO standards for the romanization of various languages, as suggested by the International Bureau. The Task Force concluded that the proposed Standard should include only relevant WIPO Standards as a general approach, rather than incorporating relevant ISO standards, because IPOs may not consistently follow ISO standards and may change their practices over time.

13. Regarding the transliteration tables used by IPOs, the Task Force notes that the main objective is to provide a reference for reasonable discussions with applicants, not to change the whole database according to the transliteration tables. The Task Force Offices were asked to provide their transliteration tables, if available, so that applicants, representatives, or IPOs can refer to the tables from other IPOs using different languages when submitting names or cleaning name data. It is proposed that the CWS request its Members to provide their transliteration table(s). It is also proposed to publish transliteration tables provided by IPOs in Part 7 of the WIPO Handbook.

14. If the new Standard is adopted by the CWS at the current session, it is proposed that the CWS requests the Secretariat to publish these recommendations in [Part 3 of the WIPO Handbook](#) .

15. *The CWS is invited to:*

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

(b) request the Name Standardization Task Force revisit and improve the draft WIPO Standard ST.93;

(c) encourage Offices and IP Industry to nominate experts to join the Name Standardization Task Force;

(d) request the International Bureau organize a workshop regarding the data cleaning of names where all interested parties may attend; and

(e) request its Members and Observers to support the International Bureau by promoting this workshop.

[Annex follows]