

# **Brief on the draft recommendations on data cleaning of names**

**Name Standardization Workshop  
Geneva, Switzerland  
May 12, 2025**

Erjola MURATAJ  
IP Information Officer, Standards Section

# Importance of a Standard on Data Cleaning of Names

Having a recognized standard for applicant names provides multiple key benefits:

**Interoperability:** Allows data sharing and integration across IP offices, legal databases, and research tools worldwide.

**Consistency:** Ensures that different organizations and systems refer to harmonized names.

**Efficiency:** Reduces time and manual effort in cleaning or reconciling data.

# Committee on WIPO Standard (CWS)

## Task No. 55

- Task Force: Name Standardization
- Co-leaders: Korean Intellectual Property Office (KIPO) and the International Bureau

### **Description of the Task No.55**

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

# Proposal Overview

- Intellectual Property Offices (IPOs) experience challenges identifying members within a patent family due to:
  - Inconsistent or varied applicant names.
  - Spelling or typographical errors.
- Accurate and clean applicant name data is crucial for reliable statistics and efficient processing.

# Proposal on Recommendations on Name Data Cleaning for WIPO ST.93

- Name Standardization Task Force developed and presented at the 12th session of the CWS a set of recommendation [CWS/12/16 REV.](#)
- The proposal aims to provide general recommendations on collection, cleaning, and publication of customer name data in IP systems.
- CWS ask Task Force to revisit and improve the proposal and International Bureau to organize a workshop on data cleaning of names

# Scope of the Proposed Standard

- Intake and management of personal and business names
- Data cleaning and duplication handling
- Name transformations: transliteration, transcription, translation
- Publication and data exchange

# Not Covered by the Scope of the Proposed Standard

- Details on data cleaning approaches
- Details on name localization or transformation methods
- Specific algorithms, merging strategies, or matching logic
- Frequency or automation details

These choices are left to each IPO or authority, based on internal systems and processes, purpose of the transformation (legal, administrative, statistical) and the fast-changing nature of data matching algorithms

# Intake of Customer Names

- Allow customers to create and manage electronic records containing published name information
- Allow customers to be associated with multiple applications or registrations for IP rights, so that they may reuse the same name information
- Support customer name input in both native characters and IPO language
- Use UTF-8 encoding to handle multilingual data
- Allow centralized updates to name records used across applications



# Transformation of Customer Names

- Transliteration – spelling-based conversion
- Transcription – sound-based conversion
- Translation – meaning-based conversion
- Scheme of transliteration, transcription or translation should be published
- Draft Standard recommends to avoid reverse transformations — they often introduce data integrity issues.

# Validation and Disambiguation

- Design of validation approaches
- Perform validation of submitted customer information, including automated checks
- Use automated checks and allow manual overrides
- Confirm name changes with the customer if they affect legal records
- Names are not unique — use contextual data (e.g., address, birthdate where available )
- Enable detection of duplicates like “ABC Corp” vs. “ABC Corp.”

# Maintenance

- Implement periodic cleaning routines, including searching for and attempt to resolve duplicate records
- Appoint data quality stewards to be responsible for cleaning data issues, including development of metrics for measuring clean data, regular monitoring and reporting of those metrics, and taking action to improve customer data when needed.
- Provide a mechanism for customers to update their name information on multiple applications or IP rights by entering the information once

# Publication, Exchange & Statistics

- All changes of names on published record should be reflected
- IPO has other forms of a customer name, such as original name expressed using native characters, these should be included in published data and data exchanged with other IPOs
- When matching methodology uses for statistical purpose to improve data accuracy, matching strategy or algorithm should be published along with the statistical results so others can understand the methodology used.

# Next steps

- The Secretariat will gather input and suggestions from this Workshop and present them to the Name Standardization Task Force.
- The Task Force will present a revised proposal at the 13th Session of the Committee on WIPO Standards (CWS).
- Send your suggestions for improving the [draft standard](#) to: [cws.mail@wipo.int](mailto:cws.mail@wipo.int)

Thank you!  
Any questions or comments?

Contact us at [cws.mail@wipo.int](mailto:cws.mail@wipo.int)