
WIPO/IP/CWS_NS/GE/25/2
ORIGINAL: ENGLISH ONLY
DATE: AUGUST 11, 2025

NAME STANDARDIZATION WORKSHOP

Geneva¹, May 12, 2025

SUMMARY OF DISCUSSIONS

prepared by the International Bureau of WIPO

¹ The headquarters of WIPO, 34, Chemin des Colombettes, Geneva (Room U. Uchtenhagen).

Date: May 12, 2025

Venue: WIPO Headquarters, Geneva and online via Zoom

Organized by: The International Bureau of the World Intellectual Property Organization (WIPO)

Participants

The Workshop was attended by 88 participants (online and in person) representing 35 countries and four regional and international organizations (the African Intellectual Property Organization (OAPI), European Patent Office (EPO), European Union Intellectual Property Office (EUIPO) and World Intellectual Property Organization (WIPO)).

List of speakers

- **Young-Woo Yun**, Head, Standards Section, WIPO
 - **Erjola Murataj**, Intellectual Property Information Officer, Standards Section, WIPO
 - **Mitchell Gray**, Senior Data Scientist, Data and Analytics, IP Australia, Australia
 - **Jumi Lee**, Senior Deputy Director, Korean Intellectual Property Office, Republic of Korea
 - **Alexander Gorbunov**, Head, AI Development Centre, Federal Institute of Industrial Property, Russian Federation
 - **Jouko Berndtson**, Development Specialist, Finnish Patent and Registration Office, Finland
 - **Argentina García González**, Examiner, Business Operations Department, EUIPO
 - **Mosahid Khan**, Head, Statistics and Data Analytics Division, WIPO
 - **Khadar Hosh**, Data Analyst, Canadian Intellectual Property Office, Canada
 - **Yann Cherel**, Consultant, Data Management and Governance Section, WIPO
 - **Nicolas Hauw**, Senior Project Manager, Standards Section, WIPO
 - **Arndt Mecke**, Deputy Chair, IMPACT Working Group, Patent Documentation Group (PDG)
 - **Guido Moradei**, Confederacy of European Patent Information User Groups (CEPIUG)
-

The Workshop program and presentations are available at:
https://www.wipo.int/meetings/en/details.jsp?meeting_id=86494

Introduction and opening remarks

IP data experts and other representatives and stakeholders from intellectual property (IP) offices and the broader IP industry were invited to the WIPO Name Standardization Workshop to address one of the most persistent challenges in the IP system: the consistent identification of applicant names.

Young-Woo Yun, Head of the Standards Section at WIPO, opened the Workshop and noted the complexity and global importance of name standardization. He said that inconsistencies in applicant names caused by typographical errors, differing transliteration systems, legal naming conventions and acronym use had long posed obstacles to achieving clean data, transparency and interoperability in the IP ecosystem.

The Workshop was part of a broader initiative launched through previous events held in 2016 and 2019, paving the way for the proposed WIPO Standard ST.93 (ST.93) on name data cleaning. In addition to sharing best practices, the aim of the Workshop was therefore to gather feedback in order to refine the draft standard before it was considered for adoption at the 13th session of the Committee on WIPO Standards, due to be held in November 2025.

Session 1: Practices and challenges on cleaning and normalizing name data

Brief on the draft recommendations on data cleaning of names

Erjola Murataj, an IP information officer at WIPO, presenting the draft ST.93, said that the mandate of the Committee's Task No. 55 called for a comprehensive framework for managing the quality and consistency of applicant names across IP systems. The aim of ST.93 was to provide practical guidance to IP offices on collecting, cleaning, managing and publishing customer name data. It outlined structured processes for name intake, including recommendations on character encoding (notably UTF-8 for multilingual compatibility), and encouraged the use of standardized methods for name transformations such as through transliteration, transcription and translation. The draft stressed the importance of maintaining data quality through duplicate detection and contextual validation (cross-checking birth dates or addresses), and the need to establish mechanisms allowing applicants to update their name records across multiple applications simultaneously. However, the draft excluded prescriptive algorithms, giving national offices flexibility to implement methods according to local legal and administrative requirements. Specific approaches to automation, matching logic and internal decision-making processes were acknowledged but not mandated, allowing the standard to be adaptable across jurisdictions with diverse technical infrastructures. Feedback gathered at the Workshop would be discussed in Name Standardization Task Force meeting immediately following the event, with the intention of preparing a final version of the draft standard for formal submission in late 2025.

Exchange of current practices, benefits and challenges on cleaning and normalizing name data

The Workshop provided a platform for national IP offices to share a broadly diverse range of practical experiences and shed light on their operational, legal and linguistic challenges in managing applicant name data.

IP Australia

Mitchell Gray, Senior Data Scientist, Data and Analytics at IP Australia, delivered a highly technical presentation on the agency's sophisticated entity resolution strategies. He said that IP Australia operated with large and historically diverse datasets, ranging from manually entered records from the early 20th century to modern, digitally structured applications. In its approach, the agency distinguished between two use cases: legal authentication, which required strict precision, and analytics, which could accept harmonized approximations of identity.

IP Australia leveraged machine learning, network science and community detection algorithms. Its novel "iterative recurrent Leiden" model allowed entities to be clustered at varying levels of granularity, depending on the analytical purpose. For example, the model could differentiate between divisions within a multinational firm (such as Google's artificial intelligence (AI) subsidiary as opposed to its search engine group), while still linking them under a broader corporate umbrella. The agency also maintained two key datasets: a domestic analytics repository ("DAVID") and an international data integration initiative based on PATSTAT (EPO patents statistics database) and DOCDB (EPO worldwide bibliographic data). Such systems, although powerful, were tailored for analytical purposes and were not used to determine legal ownership or rights without proper authentication mechanisms.

Korean Intellectual Property Office

Jumi Lee, Senior Deputy Director, Korean Intellectual Property Office (KIPO), outlined the Office's longstanding use of customer codes, introduced in 1999 to streamline applicant tracking. She said that the codes, linked to national registration numbers, made it possible to uniquely identify applicants, even across multiple filings. The Office's structured procedures ensured that, once a code had been assigned, it must be used consistently. Duplicate codes had been cleaned systematically in 2007 and 2009, and further revisions had been formalized through guidelines introduced in 2011.

KIPO used the Levenshtein Distance Algorithm to detect variant spellings and typographical errors in name entries. One highlight was the successful consolidation of 23 separate entries for Samsung Electronics into a single, authoritative customer code. Those normalized names were now published through the KIPRIS+ system, along with the associated customer identifiers. Like other offices and despite its robust internal systems, KIPO faced challenges in updating address information and managing foreign names. The Office was also exploring algorithmic models tailored to handle non-Latin scripts such as Korean and Chinese.

Russian Federal Institute of Industrial Property

Alexander Gorbunov, Head of the AI Development Centre, Federal Institute of Industrial Property, in an overview of the Russian Federation's dual-track approach to name standardization, said that two national identifiers, the taxpayer ID and the state registration number, formed the basis for linking name data. Many legacy records, however, predated those components, leaving substantial gaps in older datasets.

To address that issue, the Institute had developed a human-machine interface that enabled examiners to manually merge name records and assign standardized forms. In parallel, an automated system pulled data from national registers, enriched existing entries and attempted to match them using fuzzy logic and string comparison techniques. Normalized names were used primarily for analytical and search purposes; legal names, once submitted by the applicant, remained unchanged in official records. An intriguing innovation was the

mapping of corporate hierarchies, which allowed users to view group structures and identify subsidiaries linked to parent entities.

Finnish Patent and Registration Office

Jouko Berndtson, Development Specialist, Finnish Patent and Registration Office (PRH), said that, given the relatively small scale of its operations and the modest volume of national filings, the Finnish Office managed customer data with manual and automated processes. All data was stored in UTF-8 format and integrated with national business registers. Applicant information could be pre-filled from the Finnish YJT business register and filing systems used the same core database to avoid duplicates.

Flagging new entries for review if they did not match existing identifiers or displayed minor inconsistencies was particularly effective. The Office did not engage in name transformation or transliteration, as most of its applicants submitted data in the Latin alphabet. Future developments, such as a global identifier (Global ID or GID) resulting from the ongoing WIPO Global Identifier Project, would be critically important, especially if used in conjunction with the Finnish Business ID.

European Union Intellectual Property Office

Argentina García González, an examiner at the European Union Intellectual Property Office (EUIPO), in her presentation on data validation and name standardization in EUIPO trademark and design systems, said that new representatives were created in a pending status that triggered a back-office task to check their validity and for duplicates.

WIPO Statistics and Data Analytics Division

Mosahid Khan, Head of the WIPO Statistics and Data Analytics Division, said that the main goal of efforts to harmonize applicant names for statistical and analytical purposes was to ensure the accuracy and consistency of IP data. That was crucial for effective policymaking, research and operational planning. Inconsistencies in applicant names, such as typographical errors, formatting differences, acronyms and variations in legal forms, could result in misleading statistics.

Different user groups, including policymakers, researchers and businesses, had diverse IP data needs. The methods used for name harmonization must therefore be tailored to suit the intended purpose of the data. For instance, a statistical analysis might require different handling of name variants compared with legal or administrative uses.

Common problems, such as spelling errors, inconsistent use of punctuation or abbreviations, and non-standard characters, posed a major challenge. Differences in legal forms of company names, such as “Ltd” versus “Limited”, also had to be taken into account. WIPO harmonized data at individual applicant level rather than by aggregating data by corporate group or geographical location. For example, “Philips Netherlands” and “Philips Japan” were treated as separate entities in the statistical dataset.

WIPO had adopted a hybrid approach combining dictionary-based matching with a repository of canonical names and their known variants, and string similarity algorithms such as Levenshtein and Jaro-Winkler to identify close matches. Before applying those techniques, names underwent pre-processing steps, which included converting them to uppercase, removing special characters and standardizing legal form terms.

To ensure high accuracy, especially for top applicants, who accounted for a significant portion of filings, WIPO conducted manual validation, which boosted accuracy levels to

between 97 per cent and 99 per cent. The focus was on top applicants. Name changes were not tracked over time and the system was not fully automated. Ongoing monitoring and updates were therefore necessary to maintain data integrity. The system had several limitations; it was time-consuming and required experts to provide ongoing manual refinement.

Canadian Intellectual Property Office

Khadar Hosh, a data analyst at the Canadian Intellectual Property Office (CIPO), said that managing applicant names across the Office's diverse business areas (patents, trademarks, industrial designs and copyright) was complex. The law mandated the collection of applicant names but offered minimal guidance on formatting or standardization. The resulting inconsistencies had an adverse impact on data quality, client service and interoperability with other IP offices.

The intake and validation of names differed across business areas. For instance, patent names must be intelligible in either English or French and were verified manually against legal documents. Names for trademark and industrial designs were recorded as submitted, with greater reliance on dropdown searches. The lack of a unified name management system, the disconnect between databases and the use of manual processes all led to further fragmentation and inconsistency.

CIPO needed to improve data quality through clearer guidelines on name formatting and standardization and by adopting a unified, Office-wide name management strategy.

AI-powered tools for name data normalization

Yann Cherel, a consultant with the WIPO Data Management and Governance Section, said that WIPO used a range of AI-powered techniques, including Python-based tools, active learning and custom language models, to support name harmonization. The effectiveness of those tools varied considerably, depending on the purpose: an accuracy rate of 70 per cent might be acceptable for marketing purposes but transactional data required accuracy as high as 99 per cent.

Data profiling and segmentation were key to tailoring algorithms appropriately and thorough data cleansing at source was needed to ensure reliable input for AI models. Achieving the required level of accuracy demanded a blend of AI, rule-based systems and human validation. The process was time-consuming and required ongoing oversight and refinement. There was a need to reinforce data governance frameworks, adopt effective policies and combine AI with human oversight for better results in name harmonization.

Applicant Name Standardization initiative of the five intellectual property offices (IP5)

Jumi Lee, Senior Deputy Director, KIPO, presented updates on the IP5 Applicant Name Standardization (ANS) initiative. IP5 offices had been working to standardize applicant names as part of the Global Dossier initiative to improve consistency and efficiency in patent filings. A key goal was to build a global mapping table of applicant names and addresses across IP5 regions. Initially, 59 companies had joined the initiative but participation had stalled since 2019, mainly owing to a lack of awareness of the initiative.

In 2022, the IP5 industry had requested that name standardization be aligned with the Global Assignment project, the aim of which was to simplify the transfer of IP rights (IPRs) across jurisdictions. A notable challenge was illustrated by a Korean company's global name change, which had necessitated the hiring of local agents in each country. That highlighted the need for centralized processes.

KIPO proposed splitting the task into two parts:

1. Data cleaning of names.
2. Issuing a global identifier with user consent for those using IP5 collaborative services.

A survey of 47 industry participants had revealed a lack of awareness of the mapping table but strong support for standardized IDs, especially for global assignment purposes. They had expressed concern about the accuracy of automated name cleaning and the need for applicant consent.

The industry now favored a targeted approach: issuing WIPO global identifiers to users who needed them, rather than overhauling the full name dataset. At the 2024 IP5 roundtable discussion, industry stakeholders had supported that revised approach.

KIPO recommended shifting the project's focus from maintaining the global mapping table to enabling each IP5 office to issue and manage a unique ID for applicants. Those IDs would help to ensure name consistency, reduce errors and facilitate such services as global assignment.

Those findings were to be presented at the May 2025 IP5 Heads and Industry meeting and the task closure conditions would be revised accordingly.

Session 2: Cooperation on cleaning and normalizing name data

Collaborative initiatives among intellectual property offices

Nicolas Hauw, a senior project manager at WIPO, said that the aim of the Organization's Global Identifier Project was to address widespread issues relating to identity verification and name standardization in the IP ecosystem. They included typographical errors, inconsistent name records across jurisdictions and inefficient IP portfolio management, all of which led to increased costs, limited transparency with regard to IP ownership and even cases of identity fraud.

The WIPO Global Identifier Project had been born of a 2019 survey that had highlighted the need for a stable, globally accepted digital identifier for individuals and legal entities involved in IP processes. The identifier, which could be issued as an alphanumeric code or digital certificate, was designed to be used by inventors, applicants, law firms and other stakeholders. It facilitated centralized, user-controlled data management and simplified transactions by enabling automatic updates across all participating IP offices.

The three-phase WIPO Global Identifier Project involved: (i) feasibility studies and technical documentation; (ii) the building and testing of the global identifier system with partner IP offices and end users, covering technical infrastructure and real-world use cases; and (iii) the system's gradual rollout.

Mr. Hauw noted that the project was currently in its second phase: building and testing the global identifier system with partner IP offices and end users. This phase was focused on developing the technical infrastructure and validating real-world use cases. The system was being designed around a trust model, in which WIPO serves as the platform provider, while IP offices act as global identifier issuers and verifiers. A central registry was expected to maintain the global identifiers, ensuring global validity. The issuance of global identifiers required a secure process for identity proofing. Verification, on the other hand, required only

minimal data (the global identifier number and an email address), making it user-friendly and secure.

Stakeholder involvement remained key to ensuring that the system is co-created with end users for practical adoption. Once implemented, global identifiers were expected to allow users to auto-fill applications, reduce errors and manage IPRs more effectively, including ownership transfers and renewals, all of which would improve IP transaction workflows.

Global agreement was needed on identity standards, data privacy concerns and the integration of global identifiers with existing legacy systems. The WIPO Global Identifier Project could transform the management of IPRs through secure, interoperable digital identification.

Global agreement was needed on identity standards, data privacy concerns and the integration of global identifiers with existing legacy systems. The WIPO Global Identifier Project could transform the management of IPRs through secure, interoperable digital identification.

Engagement of IP user groups

Arndt Mecke, Deputy Chair of the Patent Documentation Group (PDG) IMPACT Working Group, said that name standardization problems in the IP sector had an impact on data accuracy and legal clarity. Variations in spelling, even across Latin-based languages, could lead to identification issues. Modern IT systems made it possible to record accurately original name spellings, which should be preserved alongside any localized or transliterated versions.

The DIN 91379 standard in Germany was a model for preserving character integrity across European languages. ST.93 would have a further positive impact by mandating the use of a universal character set (such as UTF-8) and the retention of original applicant data. It should align with or build upon WIPO Standard ST.20.

Global identifier systems had potential benefits but identifiers based on Latinized names could lose meaning across jurisdictions. There was, therefore, a need for integration with authoritative national databases to ensure legal reliability. True identification required preserving and sharing accurate, original data across systems.

Guido Moradei, representing the Confederacy of European Patent Information User Groups (CEPIUG), highlighted that problems with inconsistent patent applicants and assignee names across national and international databases persisted. Variations in spelling, punctuation, abbreviations and company name changes often produced incomplete or fragmented search results, leading to a risk of legal and business errors. A standardized global identifier system was needed to ensure accurate and up-to-date data. Progress was being made with digital filing and database improvements but legacy errors and data inconsistencies remained a challenge for patent information professionals.

Session 3: Round table discussions and conclusions

Mr. Yun, summarizing discussions and noting consensus on the critical importance of data quality in name standardization and normalization efforts across IP offices, said that the approach to name normalization depended largely on its purpose. Two main purposes had emerged: (i) authentication and protection within IP registries, and (ii) analytical applications such as statistics, marketing and customer relations management. Those different uses required varying levels of accuracy and called for different methods and solutions.

Many offices used AI-powered tools and algorithms mainly for analytical tasks and some were already publishing normalized names to support better data analytics. There was broad agreement that unique identifiers would be ideal for IP protection and registry functions, although no universally accepted, unique global identifier yet existed. The WIPO Global Identifier Project, which involved working with various offices and industry stakeholders, was an important step on the road to addressing that gap.

Managing legacy data, often scattered across different databases and IPRs, continued to pose a challenge. Several offices, including WIPO, had introduced mandatory identifier fields to improve new applicant data quality. Electronic filing systems with pre-filled applicant information were helping to improve accuracy at the source, allowing applicants to verify and correct their details. Cooperation between offices, such as through the sharing of regional data, would be vital for progress.

Speakers had stressed that recommendations in the draft ST.93 should be flexible in order to avoid unnecessary restrictions while promoting shared AI training datasets to improve normalization tools globally. Its scope, whether focusing on companies or individuals, must be defined clearly, given that individual names were particularly difficult to harmonize.

In order to achieve high accuracy in name normalization, AI would need to be combined with a rule-based approach and human oversight through iterative testing. The process would take time and require cooperation among offices. Policy and data governance were equally critical, especially as some stakeholders were reluctant to change names in official registries, even if normalized names were acceptable for analytics.

Participants saw promise in the WIPO Global Identifier Project but its full adoption would take time and effort. The main aim of the draft ST.93 was to improve data quality at the source and not just to enhance analytical capabilities.

Given the international nature of IPRs, the need to work together globally on name standardization was clear. Sharing normalized names and best practices between offices, supported by unique IDs where possible, would be key. The Name Standardization Task Force would take into account the insights gained at the Workshop when they finalized the draft ST.93, which would be submitted for examination by the Committee at its 13th session in November 2025. Further collaboration among IP offices, data scientists and legal experts would be essential for reaching a consensus.

Key takeaways:

1. Adoption of a standard for name data cleaning

Participants emphasized the importance of a standardized approach to the cleaning of customer name data. IP offices needed a practical guidance document to help them with the consistent collection, cleaning, management and publication of name data. At the same time, structured processes for name intake to ensure reliability, reduce duplication and improve data usability across systems and jurisdictions were also needed.

2. Global identifiers and interoperability

A recurring theme at the Workshop was the need for a unique global identifier to support consistent identity management across offices. While some offices have implemented identifiers locally, a global standard has yet to be adopted universally. Participants acknowledged the WIPO Global Identifier Project as a promising step in that direction.

3. Purpose-driven normalization

Different use cases, such as IPRs management versus analytics (statistics, marketing

and customer relations management) require different levels of accuracy and methods for name normalization, and therefore tailored solutions.

4. Emerging role of AI

Many offices are exploring or already using AI-powered tools for normalization. However, the quality of AI outcomes is closely tied to the training datasets and local context, such as language and naming conventions. Participants suggested that collaboration on shared datasets would help to improve AI training quality across jurisdictions.

5. Legacy data versus new data

Offices are tackling the complexity of legacy data, which often resides in disparate systems and formats, while improvements in electronic filing systems are helping to ensure better data quality at source for new applications.

6. National versus foreign applicants

Some offices have already put validation tools in place for national applicants, but validation with the applicant ID for foreigner applicants was not yet possible.

7. Cross-office collaboration

Several speakers highlighted the value of inter-office collaboration, such as through the use of pre-validated data from regional offices or national business registries. However, there are legal and technical constraints, in particular with regard to access and real-time updates from external registers.

8. Transformation of customer names

Normalization also requires sensitivity to transliteration, translation and cultural naming variations. It was reiterated at the Workshop that names do not inherently have an “English” version and that blind transliteration can introduce errors.

9. Incremental and iterative progress

Achieving clean, harmonized name data, especially globally, will take time and should proceed through phased collaborative, iterative improvements.

10. Learning from each other

The sharing of best practices, tools and datasets among IP offices and regions can accelerate progress.

11. Role of applicants and attorneys

Improving data quality at the source, including through better tools for applicants and greater responsibility for accuracy, is essential.

Conclusions and next steps

The Workshop showed that the challenge of name standardization is nearly universal but that solutions vary widely depending on national legal frameworks, technological capabilities and language systems. Participants recognized that they all have to meet the twin challenges of maintaining legal integrity and ensuring data cleanliness for research, transparency and operational efficiency.

There was interest in efforts by WIPO to develop a globally recognized identifier system. Several presenters noted that integrating such a system would be feasible within their current frameworks, particularly if stakeholders consented to the harmonization and if the identifier remained distinct from the official legal name record.

The updated draft ST.93 will be discussed further during the upcoming meeting of the Name Standardization Task Force. The final version is scheduled to be submitted for review and possibly adoption at the Committee's 13th session in November 2025.

Topics to be addressed by the Name Standardization Task Force

1. Finalization of draft ST.93

The draft ST.93 received broad support during the Workshop and the subsequent Task Force meeting. The Task Force is encouraged to complete its revision and prepare the draft for submission to the Committee at its 13th session in November 2025.

2. Potential revision of WIPO Standard ST.20

Two participants, a patent information consultant based in the United States and the PDG, raised the question of whether ST.20 should be revised. The Task Force acknowledged that, although ST.20 remains relevant for certain indexing principles, its practical use has diminished with the shift to online search systems. A short survey will be conducted among IP offices to assess current usage and the need for revision.

3. Concerns regarding Annex 2 of the draft ST.93

Annex 2 of the draft ST.93 contains inconsistencies in the transcription of the Russian letter “Ш” (e.g., 'sh' in English vs. 'sch' in German). Such discrepancies, it was noted, could lead to inaccuracies and a proliferation of Romanized name variants (e.g., “Чыраев” appearing as “Tschugaeff”, “Tchugaev” and “Tchougaev”). The Task Force agreed to consult with the Russian delegation to verify and possibly revise Annex 2.

4. Proposal for a new INID code for original/native names

It was suggested during the Workshop that a new INID (Internationally agreed Numbers for the Identification of Data) code could be introduced to capture the original or native form of applicant names, especially when transliteration is involved. Such a change would require broader consultation across all areas of IP (patents, trademarks and designs) and does not fall within the immediate scope of the Task Force's work. The idea will be documented for future consideration.