

产权组织标准委员会（标准委）

第十三届会议

2025 年 11 月 10 日至 14 日，日内瓦

关于支持名称数据清理的产权组织新标准的提案

名称标准化工作队共同牵头人编拟的文件

摘 要

1. 名称标准化工作队在此提交关于支持名称数据清理的新产权组织标准的最终草案，供产权组织标准委员会（标准委）第十三届会议审议和通过。

背 景

2. 在 2023 年举行的第十一届会议上，标准委批准了经修订的第 55 号任务说明：

“就旨在实现知识产权文献中名称标准化的未来行动编写提案，以期制定一项产权组织标准，帮助知识产权局更好地从源头确保名称的质量。”

（见文件 CWS/11/28 第 75 至 78 段）。

3. 在 2023 年该同届会议上，标准委审议了名称标准化工作队提出的一套支持申请人名称清理的新指导原则。标准委商定在拟议的产权组织新标准名称中使用“建议”而非“指导原则”一词，因为这样更明确。标准委还注意到秘书处建议的名称：关于名称数据清理建议的“产权组织标准 ST.93”（见文件 CWS/11/28 第 135 段）。

4. 然而，标准委并未通过拟议的标准，而是将其发回工作队作进一步讨论和改进。标准委还注意到，秘书处将探讨在产权组织网站上公布音译表集的可能性。（见文件 CWS/11/28 第 136 段和第 137 段）。

5. 在 2024 年举行的第十二届会议上，标准委审查了名称标准化工作队提出的关于名称数据清理拟议产权组织标准的改进后的草案。多个代表团表示支持该拟议标准。然而，委员会未予通过，因为有一个代表团要求更多的时间来全面评估实施该拟议标准的潜在影响，并需开展内部磋商并与客户磋商。标准委要求工作队重新审议并在必要时继续改进该标准草案（见文件 CWS/12/29 第 85 至 91 段）。

6. 在 2024 年该同届会议上，标准委要求国际局于 2025 年举办名称数据清理讲习班，向所有相关方开放。标准委还鼓励其成员及观察员通过积极推广和参与讲习班，对国际局表示支持（见文件 CWS/12/29 第 92 段）。

7. 在第十二届会议上，标准委指出产权组织各项标准是基于最佳做法的建议。知识产权局和知识产权行业通常根据自身需求，以各自的节奏和方式实施产权组织各项标准。特殊情况下，知识产权局会同意同时按原样实施某些标准（例如产权组织标准 ST. 26），或以协调一致的方式实施（例如产权组织标准 ST. 92）。所有产权组织标准均可根据各知识产权局的实施经验或新需求进行后续改进，以确保其实用、有效且适应不断变化的需求。标准委还注意到，拟议的产权组织标准 ST. 93 具有产权组织标准的一般属性，该标准一旦通过，知识产权局可立即实施、逐步实施或完全不实施（若现有系统已足够完善）。该标准亦可如常根据知识产权局实施经验反馈，进行后续改进。

8. 根据标准委在第十二届会议上的决定，国际局于 2025 年 5 月 12 日组织召开了名称标准化讲习班。在随后于 2025 年 5 月 13 日举行的会议上，名称标准化工作队分析了讲习班成果，并进行了最终讨论，以编写“名称数据清理建议”最终草案。关于工作队历史及自上届标准委会议以来的进展详情，见文件 CWS/13/7。

9. 关于知识产权局使用的音译方案，工作队向标准委第十二届会议通报：已邀请各工作队成员局向秘书处提交其音译方案（如有）。此举将使客户及其他知识产权局能够查阅以不同语言和文字运作的主管局所采用的方案。这有利于知识产权局与其客户之间保持一致且有效的沟通，而不必修改现有数据库。为支持这一目标，鼓励各知识产权局共享其方案（如有）。

关于新标准的提案

目 标

10. 各知识产权局在识别同族专利内的成员时面临困难，因为同一同族专利内可能使用不同的申请人名称。此外，申请人名称可能包含拼写或排印错误。再者，为统计目的提供清洁申请人名称数据这一愿望已被广泛接受。

益 处

11. 在知识产权领域建立申请人名称数据清理标准，能带来提升运营效率与数据完整性的显著益处。这有助于有效地追踪和管理知识产权资产，即使跨越司法管辖区或随时间推移也可做到。经过清理的名称数据有助于实现不同数据集间的可靠关联，为组合管理、法律合规、所有权追踪及尽职调查等活动提供支持。干净的标准化数据可提升检索效率、减少冗余信息，并支持更精准的分析与竞争情报工作。此外，它能推动自动化进程，为知识产权分析领域的人工智能与机器学习模型开发提供支撑。尤为重要的是，该标准通过解决名称变体、缩写及多语言差异等问题，推

动全球统一。归根结底，申请人名称数据清理标准是提升决策质量、降低风险及实施战略性知识产权管理的基础性举措。

范 围

12. 拟议标准就清洁名称数据的接收、处理、清理和公布提出了一般性建议。该标准不提供与数据清理方法、名称本地化或转换（例如音译、转录和翻译）有关的具体方法。此外，它也不提供关于名称标准化方法（例如算法的选择、应用转换的位置和时间、频率或合并策略）的指导意见。

对上一版草案所作修改

13. 考虑到关于名称数据清理提案的讨论以及名称标准化讲习班的成果，工作队修订了拟议指导原则的初稿（见文件 CWS/12/16 Rev. 附件）。具体修改如下：

- 删除附件：基于名称标准化讲习班上提出的对附件不完整及可能存在无意偏见的担忧，拟议标准的附件已被删除。
- 对若干段落进行编辑性修改，以增强清晰度，反映对讲习班期间收到的评论意见的反馈及分析，包括：
 - 第 11 段：讲习班与会者建议，强化关于纳入申请人名称母语字符的指导原则，尤其涉及音译时。与会者指出，省略母语字符可能因不同音译体系导致同族专利成员间出现重大不一致。例如，“Чугаев”这一名称可被音译为以下拉丁字符：“Tschugaeff” “Tchugaev” “Tchougaevev” “Cugaev” 或 “Chugaev”，这将使准确关联相关记录或识别正确申请人变得复杂。第 11 段已据此修订。
 - 第 22 段：为避免潜在误解，删除了第二句。有建议指出，在公布阶段用系统生成的唯一编号替代既定的申请人识别码，可能使用户感到困惑并削弱申请人追踪的一致性。已更新第 22 段以反映此关切。

14. 该拟议标准作为本文件附件，以修订模式表示自上次草案以来的所有修改。删除线文本表示删除内容，下划线文本表示新增内容。

15. 建议这项新产权组织标准采用以下名称：

“产权组织标准 ST. 93——关于名称数据清理的建议”

16. 如果该新标准在标准委本届会议上获得通过，建议标准委请秘书处[在《产权组织手册》第 3 部分](#)公布这些建议。

关于在《产权组织手册》第 7 部分公布知识产权局使用的音译方案的提案

17. 在对音译方案的使用进行讨论后，建议收集知识产权局用于客户名称的音译方案，将其公布在产权组织网站上的《产权组织知识产权信息与文献手册》第 7 部分下。预计这一集中化资源将有助于在不同司法管辖区对客户名称进行准确交叉引用和核验。

18. 建议鼓励知识产权局向秘书处提供其音译方案网址，以便在《产权组织手册》第 7 部分中公布。

19. 秘书处提议在《产权组织手册》第7部分下增设新小节，用于公布各知识产权局使用的音译方案，标题为：“各局使用的音译方案”。

20. 请标准委：

(a) 注意本文件及其附件的内容；

(b) 审议并批准上文第15段所述新产权组织标准的名称；

(c) 审议并通过上文第10至14段所述并转录于本文件附件的新产权组织标准 ST. 93；

(d) 如上文第16段所述，请秘书处在《产权组织手册》第3部分发布新产权组织标准 ST. 93；并

(e) 如上文第17至19段所述，请秘书处发出通函，邀请各局提供其音译方案，并将所提供的音译方案公布在《产权组织手册》第7部分中。

[后接附件]

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

Introduction	2
DEFINITIONS	2
INTAKE	2
TRANSFORMATION OF NAMES	3
VALIDATION AND DISAMBIGUATION.....	3
MAINTENANCE	3
PUBLICATION AND DATA EXCHANGE	4
STATISTICAL PURPOSES	4
References	4
<u>ANNEX</u>	<u>1</u>
<u>Transliteration examples:.....</u>	<u>1</u>
<u>Transcription examples:.....</u>	<u>2</u>

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 ~~these~~ 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 representing the meaning of a word or concept in the source language with something that corresponds to ~~that~~ the 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~~ a form(s) to use to request the IPOs 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~~ of 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 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~~ 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 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.

involvement 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 pPreparation 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 EXAMPLES²:

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) and is not transliterated
جان /dʒɒn/	ج	ا		ن	
Arabic to English					
نجيب /nædʒiːb/	ن	ج	ي	ب	short vowel /æ/ on N is normally not written in Arabic script
Najib /nædʒiːb/	Na	j	i	b	
English to Japanese					
Bill /bi:l/	B	i	l	l	each syllable in Japanese is a consonant-vowel sequence
ビル [bi-ru]	\	/	\	/	
English to Hindi					
Adam /ædəm/	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}: sch³~~

~~{ru}: Ш → {en}: sh~~

~~{ko}: ㅣ → {de}: ja⁴~~

~~{ko}: ㅣ → {en}: ya~~

~~{gr}: Ω → latin: O⁵~~

~~{da}: /E → {de}: Ä or AE, {en}: AE⁶~~

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 Силовые машины → {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~~