C

CWS/14/24
原文：英文
日期：2026 年 10 月 6 日

产权组织标准委员会（标准委）

第十四届会议
2026 年 11 月 23 日至 27 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

概　要

1.　名称标准化工作队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）第十四届会议审议和通过。

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标题中使用“建议”而非“指导原则”一词。标准委还注意到秘书处建议的标题：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，而是将其发回名称标准化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该标准草案并在必要时继续改进（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，知识产权局会同意按通过的版本，在同一时间框架内实施某项标准，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要求。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，或逐步实施，或完全不实施（若现有系统已足以满足需要）。与其他产权组织标准一样，该标准亦可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作出的决定，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，并举行了最终讨论，以编写“名称数据清理建议”最终草案。文件 CWS/13/7 提供了关于工作队沿革的更多详情和自上届标准委会议以来的进展。

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，从而便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）。

10. 在第十三届会议上，标准委审议了关于名称数据清理的拟议产权组织标准的改进稿，但由于未能达成共识，未通过拟议的产权组织标准 ST.93。标准委要求名称标准化工作队重新审查并完善产权组织标准 ST.93 草案，向委员会下届会议提交最终提案供审议（见文件 CWS/13/33 第 105 和 106 段）。

关于新标准的提案

<u>目　标</u>

11. 该拟议标准的目标是为知识产权局清理申请人名称数据的做法确立一般性建议，以提高此类数据的一致性、准确性和可靠性。拟议标准旨在支持同族专利成员的识别、统计分析以及知识产权信息的其他用途。

<u>益　处</u>

12. 拟议标准将为知识产权局提供一个共同框架，以提高申请人名称数据的质量、一致性和可用性。实施该标准将有利于对跨数据集、跨司法管辖区和跨时期的记录进行识别和关联，从而支持更高效的检索、组合管理、统计分析以及知识产权信息交换。该标准还有助于减少重复和不一

致，支持数据处理活动自动化和数据驱动工具的开发，并促进以更加统一协调的方式处理因多种语言和文字的使用而产生的名称变体、缩写和差异。

范　围

13. 拟议标准就知识产权局对申请人名称数据的接收、处理、清理和公布提出了一般性建议。其范围不延及与数据清理、名称本地化或名称转换（包括音译、转写和翻译）有关的具体方法或技术解决方案，也不延及名称标准化的具体方法，包括算法的选择、应用转换的阶段、应用频率或记录合并策略。

对上一版草案所作修改

14. 根据标准委在第十三届会议上作出的决定，名称标准化工作队成员受邀审查在该届会议上提交的产权组织标准 ST.93 的草案并发表评论意见。根据收到的意见，对草案作了以下修改：

- 在第 3 段中增加“Validation”（验证）和“Disambiguation”（消歧）的定义；以及
- 修订第 11 段，以澄清语言与字符集之间的区别。

15. 工作队还审议了一项提案，即作为例外，在标准中加入一项说明，明确其自愿性质。考虑到工作队的讨论情况，并为确保与其他产权组织标准保持一致，秘书处提议，作为例外，在国际局的编辑性说明中加入以下文字：

“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.”（“本标准的实施是自愿的，不具约束力。各局可根据自身情况，决定是否实施以及何时实施，可全部实施、部分实施，也可不予实施。”）

16. 拟议标准转录于本文件附件。对标准委第十三届会议上提交的草案所作的全部修改均以修订方式显示，删除线文本表示删除内容，下划线文本表示新增内容。

17. 建议这项新产权组织标准采用以下标题：

“产权组织标准 ST.93——关于名称数据清理的建议”

18. 在标准委本届会议通过该新标准的前提下，建议由标准委请秘书处在《产权组织知识产权信息与文献手册》第 3 部分公布该标准。

关于修订第 55 号任务说明的提议

19. 如果拟议的新产权组织标准 ST.93 获标准委通过，第 55 号任务现行说明项下的工作将被视为完成。建议工作队继续开展工作，以支持各知识产权局实施该标准，并根据实施过程中取得的经验，维护和进一步完善该标准。继续就客户名称数据清理方面的做法收集和交换信息将有助于这项工作。因此，工作队共同牵头人提议将第 55 号任务的说明修订如下：

“分享清理客户名称数据方面的做法，包括所采用的算法以及清理后数据或标准化数据的用途；为知识产权局制定实用指南；并视需要维护和更新产权组织标准 ST.93。”

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

20. 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。

21. 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。

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

23. 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 段所述的对拟议新产权组织标准增加的编辑性说明；

(c) 审议并批准上文第 17 段所述的新产权组织标准的名称；

(d) 审议并通过上文第 11 至 16 段所述并转录于本文件附件的新产权组织标准 ST.93；

(e) 如上文第 18 段所述，请秘书处在《产权组织手册》第 3 部分公布新产权组织标准 ST.93；

(f) 审议并批准上文第 19 段所述的第 55 号任务的拟议新说明；并

(g) 如上文第 20 至 22 段所述，请秘书处发出通函，邀请各局提供其音译方案，并将收到的音译方案公布在《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 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-8 [1]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]

[附件和文件完]