**STUDY ON APPLICANT NAME STANDARDIZATION**

*Prepared by the International Bureau*

## Introduction and background

1. Steadily increasing volumes of industrial property (IP) information and documentation produced by Industrial Property Offices (IPOs) worldwide. This has been accompanied by rapid development of the technology behind IP information databases. Many tasks performed by the users of patent information – e.g. portfolio evaluation, statistical analysis of data – are rendered more difficult if the information is not sufficiently standardized.
2. Having standardized, accurate and reliable information about applicants and current owners of IP rights would significantly facilitate statistical analysis and help companies take strategic decisions, e.g., identifying competitors and potential business partners or evaluating the IP portfolio of a given company. Therefore, this issue is important for the whole IP community and, in particular, IP information users and applicants.
3. Under the umbrella of name standardization, stakeholders discuss a broad variety of issues:

* technical problems, for example, correcting misprints in published applications or issues related to translations and transliterations of names;
* procedural aspects, like the wish to have only one assignee per patent family or the need to have transfers of IP rights registered;
* legal considerations, for example, the use of “variations” of a company name due to certain provisions of the national legislation in the territories where the protection is sought.

1. Interested parties, in particular IPOs and IP information providers, invest significant resources trying to overcome some of the issues referred to above. For example, the Korean Intellectual Property Office (KIPO) has introduced applicant codes for natural and legal persons on the basis of social security number or the national registry of legal persons respectively; Thomson Reuters assigns unique codes to applicants submitting more than 500 applications. Another approach, taken by the European Patent Office (EPO) in collaboration with the Organisation for Economic Co-operation and Development (OECD), is to develop software able to spot possible discrepancies and suggest a “normalized” version of the name.
2. There are several fora, regional and international, where stakeholders try to streamline these efforts and make them more efficient. Applicant name standardization is on the agenda of IP5 Offices as well as other IPOs; it is discussed at IP information users’ meetings.
3. In WIPO, efforts to standardize applicant names and their presentation in IP documents were made within the framework of predecessor bodies of the Committee on WIPO Standards (CWS). As the result of these efforts, WIPO Member States adopted WIPO Standard ST.20, which promotes a uniform presentation of names occurring in name indexes as well as a uniform method of ordering the names in the index by IPOs. Despite the fact that the current version of ST.20 came into force in December 1993, some recommendations provided there are still relevant. However, as this Standard focuses mainly on printed name indexes, which were universal and widespread at that time but are less so now, some provisions and references are out of date and could be revised.
4. The International Bureau of WIPO has also implemented a process for standardizing applicants’ names in its statistical reports relating to PCT and Madrid Systems. The method adopted is based on calculation of similarity of names and address between applicants, combined with keyword searching for top applicants.
5. In September 2016, the International Bureau of WIPO organized a Workshop on applicant name standardization. The main objective of the workshop was to define the priorities for applicant name standardization and determine what role WIPO can play in assisting IPOs and other stakeholders in this task.

## Scope and objectives

1. This study summarizes the situation related to the standardization of applicant names in IP documents, in particular, it focuses on problems faced and solutions found so far; the aim is to investigate what role WIPO can play in this process.
2. This study is limited to names contained in patent applications and does not cover other types of IP rights. However, conclusions related to the presentation of names applicable to patent applications might be equally applicable to trademarks and industrial designs.

## References

1. WIPO Standard ST.20 “Recommendations for the preparation of name indexes to patent documents”
2. Materials of the WIPO Standards Workshop on applicant name standardization, Geneva, September 5, 2016
3. Materials of PDG IMPACT Task Force on applicant names
4. Data production methods for harmonized patent statistics: Patentee name harmonization, Eurostat, European Commission, 2006
5. Patent statistics at Eurostat: Methods for regionalization, Sector allocation and name harmonization, Eurostat, 2011
6. Report of discussion round 4 “Meaningful patent visualisation and name harmonisation – current practice and trends”, EPO Patent Information Conference, 2015
7. KIPO, “Applicant name standardization”, Third Global Dossier Task Force (GDTF) Meeting, 2016 [presentation]
8. Korean Intellectual Property Association (KINPA), “Standardized Applicant Name”, Third Global Dossier Task Force (GDTF) Meeting, 2016 [presentation]
9. American Intellectual Property Law Association and Intellectual Property Owners Association, “Summary of Industry-Only Third Global Dossier Task Force Meeting”, 2016 [presentation]
10. Andrey Sekretov, EAPO, “Specific aspects of names processing in EAPO”, East Meets West, 2016 [presentation]
11. Stephen Adams, Magister Ltd., “What’s in a name?”, IPI-MasterClass, Rome, 2015 [presentation]
12. Julie Callaert, ECOOM – KU Leuven, “Applicant name harmonisation for better search results”, EPO Patent Information Conference, 2015 [presentation]

## Stakeholders and difficulties they face

### Patent information users

1. Many tasks of IP departments in industry require analysis of massive amounts of data. The importance of applicant and ownership information for these processes is outlined below:

* Freedom to operate (FTO) search – Ownership information is relevant to determine the relationship with the company owning the “blocking” patent to start licensing negotiations, assess risks or prepare for litigation
* Company analysis – Having complete ownership information is crucial for analyzing the patent portfolio of a company (where the number of patents might be thousands).
* Portfolio benchmarking – Comparing patent portfolios (or parts thereof) with those of a company’s competitor possibly covering thousands of patent families. Ownership information is paramount to determine the scope of the portfolios
* Technology landscape – The purpose of technology landscaping is to identify how the patenting activity “looks like” and who is active in the given field of technology. The answer to the second question lies in the applicant information

1. Primary interest of patent information users is to identify the patent owner. Trying to determine the owner of a given patent at a given point in time, they face many difficulties, one of which is ambiguous applicant names. Industry would like to see applicant names in IP documents which are: uniquely identifiable, standardized and correctly written.
2. The following factors affect the outcome of patent analysis:

* Patents registered in the name of **subsidiaries** are often not included in the number of patents belonging to a parent company
* Patent documents may contain **different name variants** of the same applicant
* There might be **different assignee names in one patent family**.
* One international application may contain **multiple applicants** (for different designated states)
* Applicant name can be **misspelled**
* **Inventor name** may appear as the assignee name

### Patent offices

1. IPOs recognize the lack of harmonization in applicant names and try to find solutions, which will increase usability of data and connectivity with other data sources.
2. Many IPOs publish applicant name information “as submitted”, as provisions of national legislation often require an action by the applicant to introduce a correction to the bibliographic data, which might entail additional fees. This prevents IPOs from

* efficient correction of errors, in particular, typographical errors in applicant names
* using one spelling variation of the same applicant name (“IBM” vs “I.B.M.”)
* consistent indication of the legal forms and other “non-significant parts”[[1]](#footnote-1) of applicant names
* using acronyms in applicant names consistently

1. Another obstacle is a lack of recommendations with respect to presenting foreign names in the working language(s) of the IPO and providing translations (usually into English) of bibliographic data elements contained in applications published by the IPOs for the purpose of international data exchange. The situation becomes even more complicated when the application is submitted via a regional or international (e.g. PCT) system of IP protection.[[2]](#footnote-2)
2. There is also a range of technical difficulties related to applicant names faced by IPOs, for example:

* long applicant names, which do not fit in database fields or screen forms
* corrupted diacritic and other specific characters
* different codes for “same looking” characters, (e.g. UTF‑8hex code 0620 for Cyrillic “P” and UTF‑8hex code0050 for Latin “P”)

### Patent information providers

1. Commercial patent information providers devote significant resources to keeping up-to-date information on patent applicants and right holders. The main challenge for them is the low quality of “raw” applicant name information. Data corrections or review is required for around 20 per cent of patents processed. This corresponds to 11’000 patents per week, which require human intellectual value added. (See “Existing solutions/Thomson Reuters” for more information.)

### Summary and conclusions

1. Lack of standardization of applicant names has been a long-standing challenge for patent information users.
2. Commercial IP information providers invest significant recourses to include standardized applicant name information in their products; the task is extremely challenging, as the information coming from primary sources (published by IPOs) is not sufficiently standardized.
3. IPOs recognize the difficulties faced by patent information users, but they lack resources, guidance and legal power to accept and publish applicant names in a uniform manner providing “quality at source”.

## Existing practices and solutions

1. The process of applicant name standardization has the following four levels [[3]](#footnote-3):

* Normalization – correction of “trivial” errors (which leave open the possibility of multiple name variants for one applicant)
* Harmonization – using one name variant for an applicant (which might not be the ultimate owner, as IP rights can be registered in the name of a subsidiary when the beneficiary is the parent company)
* Bringing applicant name information in context by disclosing the corporate structure (at the time of filing the application)
* Keeping accurate dynamic ownership record

1. The question of which level of standardization should be aimed at is complex and depends on the objective of the exercise.
2. Different practices implemented by stakeholders are explained below.

### Cooperation in IP5 Offices

1. Applicant name standardization is one of the priorities in the IP5 “Global Dossier” project; its aim is better prior art searches and improved patent file administration through harmonization of applicant names across IP5 patent document collections.
2. The objective of this exercise is to unify multiple versions of an applicant name into single standardized entry and provide a mapping table between original names indicated in IP5 documents and standardized names.
3. Originally, the idea was to come up with single numerical code for a group of name variants belonging to one applicant. It was planned to group them using syntactic proximity, address and phone number information. The following factors prevented IP5 from developing this coding system:

* Some applicants had different names, which were not syntactically close
* Comparing addresses and phone numbers proved to be extremely labor intensive
* Introducing an applicant code system was not accepted by all IP5 Offices owing to legal constraints and opposition from industry

#### AIPLA’s position

1. Representatives of American Intellectual Property Law Association (AIPLA) requested that the IPOs consult with the applicant before changing the applicant’s name for its standardized version or assigning identifiers.

### Harmonized names

#### Customer data harmonization for the EP patent granting procedure (NACM Project)

1. The EPO aims at creating a single unique identification per legal entity or individual, creating a central Master Data repository for EP applicants with harmonized names and addresses (NACM project). To achieve this, the EPO is developing a “cleaning strategy” for name standardization and deduplication of names; it also develops quality standards (syntactic analysis of names) and data quality policies. It is planned that the results of the standardization will be validated by applicants.

#### Applicant name information in PATSTAT

1. PATSTAT is a tool developed by the EPO for conducting sophisticated statistical analyses of patent data. PATSTAT incorporates several solutions for standardization of applicant names and provides the following options for its users:

* Original names
* DOCDB standardized names (updated twice per year)
* HAN applicant names (OECD) (updated twice per year)
* PATSTAT harmonized applicant and inventor names (updated twice per year)

#### Standardization of applicant names for DOCDB (STAN)

1. DOCDB includes a field with standardized applicant names assigned to companies, universities and other legal entities. Assigning standardized names in DOCDB is a two-step process: automated analysis based on a specifically-developed algorithm and, if the conclusion is not certain, human check.
2. Currently DOCDB contains approximately 250’000 standardized names covering approximately 945’000 variations of applicant names “as filed”. Originally submitted applicant name information is also included in DOCDB entries.

#### Harmonized applicant names (HAN) developed by OECD

1. The Organisation for Economic Co-operation and Development (OECD) uses patent indicators to:

* assess technological performance of companies in a certain sector of economy,
* identify emerging technologies
* study knowledge diffusion and dynamics of technical change and geographical properties of the inventive process
* investigate innovation and company dynamics
* assess the economic value of inventions
* assess the role of universities in technological development
* study the process of globalization of R&D activities

1. To carry out these analyses, the OECD needs to link patent data with up-to-date information on the companies owning patents. To achieve that, OECD developed country-specific dictionaries of applicant names (OECD HAN Database), where patent applicants names are matched to company names listed in the ORBIS© database using a tailor-made algorithm – Imalinker (Idener Multi Algorithm Linker). The development has the following features:

* matching is performed on a **country by country basis**
* names are harmonized using **country specific dictionaries**
* **string matching** algorithms are used (token/string metric based)
* **thresholds of matching scores are very high**, in order to minimize, as much as possible, false-positive and false-negative matches
* **manual controls** to adjust the precision are applied[[4]](#footnote-4)

#### Harmonized name in PATSTAT

1. Applicant (and inventor) name harmonization in PATSTAT is a two-step process; the procedure was developed by KU Leuven and presented as a part of the Eurostat project conducted by the European Commission. [5]
2. The objective of the first step (Methodology layer 1 in Eurostat terminology) is to match names that appear to be similar, but differ because of spelling or language variations; this is achieved by comparing each applicant name with all others. This step is completely automated and targets accuracy of the exercise, if necessary at the expense of completeness. The process at this stage is as follows:

* Data pre-processing
* Character cleaning (converting to ASCII characters)
* Punctuation cleaning (pre-parsing)
* Name cleaning
  + Legal form indication treatment (legal form information is harmonized and moved to a special field)
  + Common company word removal (for example, “company” or “corporation”)
  + Spelling variation harmonization (variations and earlier identified misspellings of common words, like “system”, are replaced by a correct single spelling)
  + Condensing (removal of all non-alphanumeric characters)
  + Umlaut harmonization

1. The second step (Methodology layer 2) provides additional harmonization by closer inspection of different name variants. The inspection is based on approximate string searching, further manual validation and quality control. At this stage, the most important factors are:

* completeness
* high accuracy (“conservative” rules)
* high coverage in terms of patent volumes

1. It was shown that retrieved names with a patent count of more than 10 represent 99.6 per cent of total number of names included in patent documents; therefore the inspection efforts on the second step are limited to those names (“top 500 applicants”).

### Applicant identifiers

1. KIPO developed a system of applicant codes and has used it since 1987. When KIPOnet (KIPO IT automation system) was launched in 1999, the codes were assigned to existing applicants, and since then KIPO has been issuing codes for all new applicants based on registered social security number or corporate registration number.
2. KIPO applicant codes consist of 12 digits: DYYYYNNNNNNC, where D is a distinguisher (to identify domestic and foreign companies, institutions, domestic and foreign natural persons, attorneys, etc.), YYYY – year, NNNNNN – serial number, C – check digit.
3. In 2009, KIPO revised its regulation to include a requirement that applicants should use single applicant name and applicant code. Benefits of introducing applicant codes are as follows:

* applicant information is managed more effectively
* applicants avoid the necessity of repeatedly inputting information
* applicant names and other related information can be changed simultaneously in all applications filed by the same applicant
* accurate IP statistics.[[5]](#footnote-5)

#### Thomson Reuters

1. Thomson Reuters maintain an internal database of assignee names, which contains more than 1.8 million entries. Every assignee name processed in the Derwent World Patent Index (DWPI) is checked against this database, identified errors are corrected, and the name is abbreviated and standardized. For a recognized company, the Patent Assignee Code (PACO) is added to the entry. For a new company (with more than 500 applications), a new code is assigned and added to the internal database. These codes help to distinguish companies with similar names and identify subsidiaries, which do not carry the parent company name[[6]](#footnote-6).

### Summary and conclusions

1. The majority of solutions described address the standardization of patent applicant names on the first two levels: normalization and harmonization, leaving the issues of ownership aside.
2. There are three main approaches to standardization of applicant names:

* assigning applicant codes (KIPO)
* including an additional field in patent information databases for normalized or harmonized names (EPO PATSTAT)
* maintenance of a database of applicants which contains one single record per applicant with their harmonized name and address (EPO NACM)

1. Assigning applicant codes seems to be the most straightforward way for IPOs to keep accurate applicant records and provide “quality at source”, but in some national jurisdictions it is difficult to implement due to legal issues (applicant’s data should be published “as filed”) and resistance from the industry (see AIPLA’s position). Different national practices of registering legal entities and natural persons may cause more complications in implementing this approach at regional and international levels. Nevertheless, it seems useful to share this practice with IPOs so that they can identify a suitable and reasonable way of persuading applicants to indicate their names in the most unified manner possible, taking into account national legislation.
2. Including an additional field in patent information databases for normalized or harmonized names would require significant resources in IPOs and would be valuable for international data exchange only if different IPOs applied the same rules for applicant name normalization and harmonization. Some IPOs, like the EPO, have already gained experience and expertise in this area and it would be very beneficial for the IP community if they share it with other IPOs. There is also room for standardization of rules for normalization and recommendation for harmonization of applicant names.
3. Creation and maintenance of a database of applicants is similar to the first approach (assigning applicant codes), as the database ID, which is effectively a code, will be used to link the application with applicant information. However, the validation of the information contained in the database by applicants themselves might help to avoid some of the problems mentioned above. This approach would also require significant investments from IPOs. The alternative might be to create an international database of applicant names.
4. Thomson Reuters and other commercial IP information providers take steps to establish ultimate beneficiaries (identifying the corporate structure) of patent rights and to offer this information to their clients (third and fourth levels of standardization). It is extremely labor intensive and often falls outside the jurisdiction of IPOs.

## Role of WIPO in applicant name standardization

### Committee on WIPO Standards

1. One of the conclusions of the Workshop on applicant name standardization was the request from the participants to survey IPOs on how useful applicant identifiers can be, what might be the problems in introducing them and whether the IPOs were using a “dictionary” of applicant names. Once a new Task Force has been created within the framework of the CWS, it can address this task by developing a questionnaire to be distributed to IPOs by the Secretariat. The results of the survey will be published in the WIPO *Handbook on Industrial Property Information and Documentation* (WIPO Handbook).
2. In parallel, this Task Force can study whether the recommendations in the form of a WIPO Standard can be developed to assist in achieving better harmonization of applicant names in patent documents issued by IPOs. These recommendations can cover the following areas:

* normalization of applicant names
* translation/transliteration issues
* different name structure in different countries
* inclusion of additional fields for data exchange (XML Standards, INID codes, etc.)
* and others.

### Exchange of experience between IPOs

1. Many issues causing lack of harmonization of applicant names in patent applications fall outside the main scope of WIPO standardization activities due to the fact that they originate from internal procedures implemented in IPOs, national practices for example, of registering legal entities, different structure of names historically used in different regions, etc. Nevertheless, in order to increase awareness in IPOs of the problems faced, streamline existing practices and assist IPOs wishing to learn from other IPOs’ experience, it can be useful to exchange views, experience and to share practices and expertise related to applicant name harmonization among IPOs and other interested parties.
2. Considering that the following events or activities may contribute to facilitating the exchange of experience between IPOs, the International Bureau is ready to assist in organizing them upon request:

* **Workshops on specific aspects** of name standardization can help IPOs identify changes to national legislation necessary to allow “cleaning-up” raw data or to facilitate the enforcement of the requirement to report the change in ownership.
* **Training sessions** on name normalization algorithms, or other IT solutions related to applicant names will assist IPOs in developing the infrastructure necessary to improve their patent information products.
* **Exchange of practices (round tables)** related to business process will be useful to ensure “quality at source”. They can cover a broad range of issues, for example, validation of the front page by the applicant before publication (already with the corrected data), changes to application forms and the requirement to confirm the current owner when paying fees.

## Conclusions

1. Applicant name standardization is an important and complex task relevant to all stakeholders: IPOs (in developing countries as well as in developed countries), patent information users and applicants.
2. Solutions for applicant name standardization should be sought on the international level, as well as on regional and national levels. There is a room for developing a **WIPO standard** to cover certain aspects of applicant name standardization, but for many issues identified, the development of a WIPO standard seems premature or not appropriate. To assist its Members, the CWS can conduct a **survey on the use of identifiers and name “dictionaries”** by IPOs and publish the result in the WIPO Handbook. WIPO can also assist its member States in exchanging experience and practices by organizing **workshops, training sessions and round tables** on specific problems related to applicant name standardization.

[End of Annex and of document]

1. See paragraph 13 of WIPO Standard ST.20 (December 1993) [↑](#footnote-ref-1)
2. Sekretov, Andrey “Specific aspects of names processing in EAPO”, presentation at WIPO Standards Workshop on applicant name standardization [2] [↑](#footnote-ref-2)
3. Adams, Steven. “Origins of the problem - why is this happening?”, presentation at WIPO Standards Workshop on applicant name standardization [2] [↑](#footnote-ref-3)
4. Dernis, Helene “OECD HAN database: A solution on the harmonization of applicant names for patent statistics”, presentation at WIPO Standards Workshop on applicant name standardization [2] [↑](#footnote-ref-4)
5. Yoon, Jaewook “Difficulties related applicant names and current practices in KIPO”, ”, presentation at WIPO Standards Workshop on applicant name standardization [2] [↑](#footnote-ref-5)
6. Hajkowski, Steven “Assignee information in DWPI”, presentation at WIPO Standards Workshop on applicant name standardization [2] [↑](#footnote-ref-6)