

Assignee names reconciling the irreconcilable?



The world's most complete IP data collection



LIGHTHOUSE IP

Contents

Origins of the problem – why is this happening?	3
What do we mean by 'standardised information'?	5
Level 1: Normalisation - errors at source and in transmission	7
Level 2: Harmonisation - ambiguity at source and in transmission	10
Level 3: Contextualisation - implicit information and user-defined clustering	13
Level 4: Augmentation - keeping dynamic information current	15
Who can - or should - contribute to standardized information?	17
About the author	19

Origins of the problem - why is this happening?

This White Paper considers the implications of the beguilingly simple goal of 'assignee standardisation' in patent information, what that term means to different users, and the challenges in achieving this goal. For the sake of simplicity, the discussion will be limited to considering only corporate names rather than personal ones.

When an intellectual property right such as a patent is granted, it immediately places all affected third parties into the position where – assuming that they intend to engage in commerce legally and ethically – they need to be able to discover what they can or cannot do (classic Freedom-to-Operate information) and to whom they may be answerable in the event of accidental or deliberate infringement (legal status information, specifically ownership of the right). If either of these types of disclosure is inadequate, it seems that the balance of benefits (the so-called 'patent bargain' between inventor and society) is no longer even but has become biased in favour of rights owners and against third parties.

To restore the balance, the quality (fitness-for-purpose) of patent information needs to improve, specifically to ensure that ownership records for patents, which are typically captured at publication in the form of assignee or proprietor names, are complete, accurate and up to date. It is no longer adequate simply to record these data at the point of grant. They must be accurately collected, reviewed and updated at all points throughout the life of the application, from filing to the end of its active term and beyond.

In addition to this basic information need, further demands arise as additional uses for patent information are discovered, and as patent holders become more sophisticated in their management of these assets during their lifetime. The particular viewpoints of patent owners, industry competitors and neutral third parties may

raise demands for the enrichment of basic published data with additional meta-data from other sources, in order to meet new needs for retrieval, analysis and (most importantly) accurate comprehension of the status of patent assets.

Multiple parties – applicants, patent offices, commercial database producers, licensees, proprietors, other national agencies such as courts – can all have an influence upon, and contribute to, accurate and standardised assignee data. However, there is no global agreement as to which of these groups can – or should – take on the responsibility for the tasks involved. There have been a few international initiatives by some major patent offices and WIPO, but no complete solution. Is it even a realistic target? Firstly, all parties need to agree on the objective, since 'standardisation' has a multitude of nuanced meanings in this field.

What do we mean by 'standardised information'?

The term 'standardised' can be broken down into several different steps, as different parties in the business of patent information can contribute to – or defeat – efforts to achieve better quality data. For the purposes of this White Paper, we will consider four basic levels:

- 1) Normalisation;** this can be achieved by ensuring a basic level of error correction, which includes rationalising trivial spelling errors as well as common variations, the use of abbreviations or different corporate designations (such as 'Limited' or 'Ltd.' etc.). The objective is to ensure that the database record displays the same form of assignee name for each application in their portfolio. In a global environment like patent information, normalisation efforts should also extend to the challenges of systematic transcription, transliteration and/or translation of assignee names from one language to another.
- 2) Harmonisation;** unlike normalisation, this step accepts that there may be more than one equally authoritative form of an assignee name but allocates priority to the use of only one form. This form of data manipulation is well known in other areas of information science, and can be achieved through the construction of thesauri or simple "authority lists" of approved terms. The challenge is then to manage such lists, and to decide who has control over their development.
- 3) Contextualisation;** this form of standardisation attempts to place an individual record within its proper wider context. Typically, when dealing with organisations, this process creates links between different subsidiaries of the same corporate family, and is an attempt to understand the ultimate beneficiary of any intellectual property rights. As such, contextual information is of most interest to patent information analysts in their examination of national and international trends.

4) Augmentation; a further step in improving assignee standardisation may be needed for certain user groups. Commercial awareness that patents can be tradeable assets has increased over the last decades. Consequently, both databases and tools which are being applied to large-scale portfolio analysis must be able to recognise that changes in both form and ownership may – indeed, probably - have happened over the lifetime of the patent. However, the legal mechanisms for capturing these changes, and the information industry’s processes for making them available, have struggled to adapt to the idea that the information linked to a patent record needs to be dynamic. This means going beyond the details first published at grant, as a static bibliographic record, and paying equal attention to the recording of reassignments, licensing information, name changes of existing owners, and so on.

Level 1: Normalisation - errors at source and in transmission

One of the major improvements which has been achieved in name standardisation is the elimination of a large proportion of transcription errors occurring in manually filed applications. In the years when electronic databases were produced by re-keyboarding of paper records, the conversion process could itself introduce errors; similarly, an accurate rendering of data which are incorrect at source results merely in the faster and more widespread propagation of the original error.

As an example, consider **Table 1**, which shows a number of entries from the UK official gazette of the mid-1990s, disclosing a sequence of applications in the UK, claiming Italian priority. It is clear to the human eye that all five cases are filed by the same company, particularly given the sequential priority application numbers (all filed on the same day at the same receiving office). Yet the applicant's name – which will typically enter an electronic database without further correction – appears with no fewer than four variations of punctuation and/or expansion of abbreviations. These errors have the potential to ruin any assignee name search or statistical analysis.

Patent application	Applicant name as filed
IT 93 BO A 0254	G.D SpA
IT 93 BO A 0255	G.D. S.p.A.
IT 93 BO A 0256	G D Società per Azioni
IT 93 BO A 0257	G.D. S.p.A.
IT 93 BO A 0261	G.D Società per Azioni

Table 1: Trivial variation in assignee name

As many more patent offices are geared up for electronic filing, conversion errors are being reduced but they cannot eliminate errors at source. The names of some applicants (particularly SMEs) may be inconsistently recorded at the moment of filing an application, especially if the company deals with multiple external professional representatives and/or is filing a low number of applications per year. Most patent offices accept (and reproduce) an applicant or assignee name exactly as supplied, and may be reluctant to correct even 'obvious' mistakes on their own initiative. Even if the error is detected, applicants may be deterred from submitting an amendment if the patent office charges an extra fee to do so. If the application is then published 'as filed' at 18 months after priority, these inconsistencies become propagated through primary bibliographic databases and are rarely corrected.

Normalisation of assignee names which have originated in one language or script but need to be re-published in one or more other languages raises additional problems. For example, if a Japanese company files a Convention application in South Korea and the subsequent Korean publication enters an English-language database, the final rendering of the assignee is likely to look radically different from that of the Japanese priority document in the same database. Even if only one step is involved (such as Cyrillic to Latin script), there are different standards which may be used depending upon the target language, resulting in multiple forms of the 'same' name when publications are grouped into a family record; the surname Чайковський in Cyrillic may be transcribed into Chaikowskii for the purposes of a German-language publication but into the form Tchaikovsky for an English-language publication. If both target documents are members of the same patent family, it may appear that the family has multiple different assignees.

Resolving these 'trivial' differences is not easy. When dealing with corporate assignees, the priorities and opinions of applicants, patent offices and database users often differ as to what should be the 'preferred' or 'authentic' form of name, and how to control a listing of authorised variations; to some, the inclusion of "& Co., Inc." or the change from "GmbH" (or even G.m.b.H.!) to "A.G." is significant, whereas

other users are less concerned. Likewise, the rendering of diacritics or language-specific characters from one language into another may not seem of prime importance in some circumstances, but to other users they may be legally significant. Western database producers have traditionally struggled with anything other than a basic Latin character set, but the challenge of capturing data with characters like ð, þ (Icelandic), ø, å (Danish), ş, ı (Turkish), ł, ś (Polish) or д, л (Bulgarian) is likely to arise even if the database producer is only handling EPO information.

Even the treatment of spaces and punctuation can be varied at the point of data entry, and cause issues in later retrieval or analysis; for example, a British packaging company with the legal name of “The Metal Box Company Limited” was in business for many years, but by the 1980s had created the simpler wording “Metal Box” as a trading identity, followed later by the trendy lower case version “metalbox”. These variations are very difficult to capture, let alone normalise to an accepted form in databases which, until the recent adoption of XML data standards, could not distinguish upper- or lower-case letters for display, and certainly not at the point of search.

Level 2: Harmonisation - ambiguity at source and in transmission

Even if applicant data are collected in an agreed legal format at the point of filing, this may not be helpful to all potential users of patent information. Company analysts tend to be more immersed in industry structures 'on the ground' rather than legal company registration records, and often start to search using a well-accepted form of corporate name; hence, 'everybody knows' that the patents applied for by 'International Business Machines' appear in the index as 'IBM', or that automotive patents belonging to Porsche will be listed under P, and not under D ('Dr. Ing. h.c. F. Porsche Aktiengesellschaft!'). If patent information is to become truly accessible and useful to all user communities, the authorities which create or capture patent data need to be aware of different user viewpoints and to ensure that the databases which they create can accommodate these varying needs, as much as it lies within their power to do so.

One possible way forward would be multi-national efforts to create company name thesauri, reconciling variant name forms at the national level (at least), and preferably across national and language boundaries as well. In an age of 'big data', it seems reasonable for national patent offices to leverage existing national company registration data, and incorporating these into online filing tools, such that an applicant can be offered a drop-down list of properly formatted names at the point of filing. Applicants will certainly wish to retain a right to over-ride or replace such suggestions with their own input, but some degree of variation could be avoided if the applicant has some pre-selected options in front of them. Some work has already been carried out on the use of company identifiers such as registration numbers or VAT numbers, as a proxy for an approved text form. There is a need for further research in this area, including how to extend this principle to non-profit institutions such as academic or NGO applicants.

At each stage in 'standardisation', the solution to one level must always be backwards-compatible with the lower levels. Therefore, attempts to establish *harmonised* names will also need to consider more basic *normalisation* issues such as name spelling. Consider the impact of the 'minor' error on the front page of the US Design Patent shown at *Figure 1*.

(12) United States Design Patent Blomberg et al.	(10) Patent No.: US D741,475 S
	(45) Date of Patent: ** Oct. 20, 2015
(54) RESPIRATOR MASK HAVING A COMMUNICATION GRILLE	D277,520 S * 2/1985 Gregory D24/110.1 4,562,837 A 1/1986 Schlobohm D286,213 S 10/1986 Maryanek 4,688,567 A * 8/1987 Kikuchi et al. 128/206.15 (Continued)
(71) Applicant: 3M INNOVATIVE PROPERTIES COMPANY, St. Paul, MN (US)	
(72) Inventors: David M. Blomberg , Lino Lakes, MN (US); Mark W. Schulz , Minneapolis, MN (US); Charlie Wood , Minneapolis, MN (US)	FOREIGN PATENT DOCUMENTS AU 359131 12/2014 AU 359132 12/2014 (Continued)
(73) Assignee: 3M Innovation Properties Company, St. Paul, MN (US)	OTHER PUBLICATIONS

Figure 1: US Design Patent with "incorrect" assignee name

The difficulty in determining an authoritative form is that sometimes variation can arise simply through inconsistent translation, rather than any actual change in company identity. Without conducting further research, it is not always clear whether from patent bibliographic records whether the applicant is a different legal entity or not: see *Table 2* for some examples.

Typical search term	Common database entries
3M	Minnesota Mining & Mfg [JP publications] 3M Innovative Properties Co [US publications] Three M Innovative Properties [electronic record only; not on front pages] 3M Innovation Co Ltd [CN publications]
Sandoz	Sandoz Ltd Sandoz-Patent GmbH Sandoz-Erfindungen Verwaltungsgesellschaft m.b.H.
City University	City University (London) City University of Hong Kong Osaka City University Dublin City University [etc.]

Table 2: A simplistic search for a complex subject

Data collection authorities will have to decide how to represent characters in an extended Latin set (is ü always to be replaced by ue? Should 'Hoechst' really be amended to 'Höchst'?) and, looking beyond Latin characters, there are choices to be made concerning transliteration, transcription or translation. These can become confused if a company deliberately chooses a form of name which makes more sense in the spoken rather than the written form. One striking example is the Belarus company name *ЮНАТЭК*; this literally transliterates to *YUNATEKS* but to capture the data in this form would ignore the fact that the name was deliberately adopted to render as the phoenetic form *UNATEX* and create an identity which would be more familiar to the English-speaking world. The company's domain name is *unatex.net* (*).

*Example courtesy of Andrey Sekretov, EAPO.

Level 3: Contextualisation - implicit information and user-defined clustering

Thus far, our consideration of how to standardise names has only addressed the quality of the *explicit* information provided by the applicant and/or published by the patent offices. However, many industrial applicants are national or international groupings of individual companies which belong to larger corporate structures, and some users of patent information would like to be able to set an individual patent cluster into its larger context. This means being able to use the implicit information derived from an understanding of company structures, but which is rarely made available from the original patent publications.

Proprietors are entitled to assign their patents to specific entities such as a group profit centre or asset management department, rather than the ultimate beneficiary. However, this can make it difficult for third parties to reconcile company or group financial results with the apparent structure of the patent portfolio. This lack of transparency is not only unhelpful to many patent information users, but is also of concern to national tax authorities concerned with the preservation of their own tax base (preventing so-called 'base erosion' or 'profit shifting') and the management of 'Patent Box'-style incentives. Furthermore, potential purchasers of patent assets or potential licensees both need to be able to establish who has a controlling interest in the assets, which may not be obvious from the original published proprietor names. Consider the examples of Air Liquide and the Volkswagen Group – in the one case, many subsidiaries can be readily identified from the form of the company name, but in the other it takes much more research to establish corporate relationships:

Parent company	Related companies, subsidiaries or legal identities
Air Liquide	L'Air Liquide, Société Anonyme pour l'Etude et l'Exploitation des Procédés Georges Claude [FR publications] American Air Liquide, Inc. [US publications] Air Liquide Medical systems Srl [IT publications] Air Liquide Deutschland GmbH [DE publications]
Volkswagen	VW, SEAT, Skoda, Scania, Audi, Bugatti, Lamborghini [etc.]

Table 3: Some indication of complex corporate trees

It is clear that, in order for some users to do effective analyses of patent data, some degree of reformatting will be required. It is tempting to suggest that databases which were originally created with one purpose in mind (e.g. subject-based searching to establish patentability) should be modified or adapted to fit the demands of other use cases (e.g. IP portfolio assessment as part of due diligence searching). However, this may not be a trivial exercise in itself, and large-scale modifications may result in compromising the expectations of one user group in order to satisfy the demands of another. The perfect 'one-stop shop' for patent information, in the form of a single data source which fulfils the needs of every user (technical and legal, beginner or expert...), may never be realisable.

To date, one way of trying to retain a proper balance of usefulness has been to enrich the original published records with supplementary metadata from elsewhere (such as new fields on 'parent company of proprietor' based on existing corporate trees) whilst leaving the original published data in place. An alternative approach is to use in-house tools and expertise during the search phase and after extraction of relevant records, in order to cluster the information into a more usable form. In the former case, the onus is on the database producer to obtain and load the metadata; in the latter, the burden is transferred to the database user, to impose their own understanding of company structure (or indeed, any other customised facets) upon the original data set at the point of analysis. This latter approach has the benefit that the data may be explored in exactly the way which the user requires, rather than according to preset criteria. At the present time, there seems to be little universal agreement on the best mechanism; it is likely that both methods will be used for the foreseeable future.

Level 4: Augmentation - keeping dynamic information current

The final, and most difficult, challenge in assignee names is the issue of maintaining the information current. Unlike other tangible assets such as cars or houses, there is rarely any legal requirement in Western countries for changes in ownership to be registered with a national authority. Since patent rights may be maintained for up to 20 years, it is not uncommon for the intellectual property to change hands during its life. However, most simple bibliographic patent databases only capture the name of the assignee *at the time of publication or grant*. In order to discover the current owner, it is usually necessary to consult additional registers or to monitor for other official announcements. If more than one change has happened, it is extremely laborious to reconstruct a timeline of who owned the rights, and when – which is just the information which may be needed in the event of infringement litigation. This highlights the fact that, even for level 3 standardisation and much more for level 4, the user community needs not only “who owns whom” information, but also “who used to own whom, and when” over a minimum 20-year period.

There are many factors which complicate the gathering of comprehensive ownership information in a timely manner. The laws in many countries with long-established IP regimes rarely make it a legal requirement to report assignments post-grant back to the patent authorities, and the only penalty for an out-of-date ownership record may be an inability for the new owner to commence litigation or to claim damages for infringement until the ownership record is corrected. Large corporate patent owners complain about the administrative burden of recording bulk changes across large portfolios when company merger or divestment takes place. They also tend to be less concerned about out-of-date records, since a due diligence exercise carried out before an acquisition should include access to accurate IP registers of the target company. However, this privilege does not extend to other patent information users, who may wish to reconstruct an accurate record of a company’s IP independently of making a formal bid, relying upon publicly accessible resources.

Some have argued that, given an accurate starting record and a public timeline of company merger activity, it should be possible to infer the correct ownership of any given patent today. However, in practice company mergers and acquisitions are not as simple as a wholesale transfer of an entire portfolio from one owner to another, so updating of a database of ownership records is not something which can be achieved *en bloc*, by algorithm. The terms of merger may include a requirement for only part of the IP to be reassigned to the new owner and other parts to be divested or otherwise sold on; the simple formula of "patents which used to belong to X now belong to Y" does not hold. Taken over several generations within a 20-year period, and it becomes virtually impossible to establish true ownership of an individual patent, beyond reasonable doubt.

There have been some proposals to make changes in ownership more transparent, by requiring regular reports from entities at the time of paying maintenance fees. For some countries, this would involve an annual reconfirmation of the current owner of each granted patent, and some long-pending applications as well. In 2014, the USPTO put forward a Notice of Proposed Rulemaking which would have required additional reporting at the time of other significant events in the life of a patent, including when a patent became involved in supplemental examination, ex-parte reexamination, or a trial proceeding before the PTAB. However, this proposal was not implemented.

In addition to the inconvenience of not being able to establish current ownership, out of date records can confuse the ability to monitor for litigation surrounding patents of interest. If a research company is inadequately briefed to check whether "Patent 1234567 has been the subject of legal proceedings", they may waste time looking for actions in which the owner at grant is a party, but overlook cases where the new, unregistered, owner is the actual party. The infamous Dyson -v- Hoover case in the English courts centred around an application filed in the name of Rotork, subsequently granted and reassigned to Notetry and not finally owned by Dyson (according to the public record) until just over a year before the High Court judgement. The patent expired 8 months afterwards.

Who can - or should - contribute to standardisation?

Whilst it is generally agreed that at least some of the steps outlined above *could* be undertaken, there remains the question of whether they should be. We live in a time when there is increased demand for the interoperability of 'big data' but also an increased sensitivity towards data protection principles. Business users of patent information tend to prefer that their source data is easy to 'clean up' or consolidate for effective analysis, whereas legal users may want to retain the granularity which assists their internal management of the portfolio.

Even if agreement can be reached as what changes are desirable, it is not commonly accepted which parties in the patent information chain should take responsibility for implementing the changes. Patent offices commonly take the view that it is not their place to make unilateral changes to data submitted by applicants, except in certain easily-recognised circumstances. They are even less willing to enter the murky waters of data enrichment, leaving it to the private sector or the user community to sort out the best mechanisms. Other players in the arena can be understandably reluctant to take on the burden of ensuring compliance with new, more rigorous, documentation standards if the workload or objectives are at odds with their business priorities. By breaking the processes for standardisation in the above 4 levels, it may be easier to perceive some logical structure as to who should be do what, to improve the present situation. The *Table 4* summarises this approach.

Level of standardisation	Applicant	Patent Office	Third party*	Proprietor(s)
1. Normalisation	Y	Y	Y	Y
2. Harmonisation		Y	Y	
3. Contextualisation	Y		Y	
4. Augmentation		Y		Y

Table 4: Summary of parties contributing to standardisation
 (*) e.g. commercial database producers, national government data

Some plans for action are suggested by this breakdown of the problem. At level 1, it may become possible to empower patent offices to maintain authority files of corporate applicant names at their own national level, and to begin to insist upon the use of unique identifiers at the point of application.

For level 2, there is already scope for patent offices and third parties to work together to develop existing national identifiers into an international thesaurus of harmonised corporate names, applied across worldwide data. Standardisation at level 3 is generally accepted to be beyond the remit of patenting authorities, other perhaps than encouraging applicants to disclose ultimate parent / beneficiary as well as legal applicant, but there is scope for third parties to work with government agencies to create and maintain accurate corporate trees, including proper archiving policies for records of historical restructuring. Finally, at level 4, it would be helpful if future patent law standards should include a means for offices to collect timely, frequent and complete reporting of changes in legal ownership and ultimate beneficiary, which could be achieved without undue burden on the proprietor if it is integrated with other administrative actions which they are already accustomed to perform.

The issue of assignee name standardisation is as relevant today as it has been for the last 30-40 years, and progress towards a workable solution has been very slow. It is important not to underestimate the issues; quality databases require substantial investment and allocation of adequate resources, but overcoming the problems would remove some major barriers to more widespread and informed use of patent information.

About the author

Stephen Adams is the managing director of Magister Ltd., a UK-based consultancy specialising in patents information. Mr. Adams is a Qualified Patent Information Professional (number 20190044100092) and holds a B.Sc. in chemistry from the University of Bristol and an M.Sc. in Information Science from City University, London, as well as professional memberships of the Royal Society of Chemistry (RSC) and the UK's Chartered Institute of Library and Information Professionals (CILIP). He is the author of three editions of "Information Sources in Patents", the latest published in 2020 by Walter de Gruyter KG, contributed several book chapters and written numerous articles in the field of patent information, including over 25 refereed papers for the Elsevier journal "World Patent Information". His professional service includes the Editorial Advisory Board of "World Patent Information" between 2006-2020 and three terms on the Board of PIUG Inc., the International Society for Patent Information, as Director-at-Large (2002-2006) and Vice-Chair (2014-2016 and 2016-2018), as well as service on the management committee of the UK's Patent and Trade Mark Group over many years. He received the PIUG's Special Recognition Award in 2008 and the IPI Award in 2012 for outstanding contribution to patent information.



Tupolevlaan 81, 1119 PA Schiphol-Rijk
The Netherlands
Telephone +31 85 800 0024
info@lighthouseip.com

lighthouseip.com

The world's most complete IP data collection



LIGHTHOUSE IP