



HAL
open science

The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents?

Patricia Laurens, Lionel Villard, Antoine Schoen, Philippe Larédo

► To cite this version:

Patricia Laurens, Lionel Villard, Antoine Schoen, Philippe Larédo. The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents?. *Scientometrics*, 2018, 114, pp.91-112. 10.1007/s11192-017-2578-5 . hal-01671905

HAL Id: hal-01671905

<https://hal.science/hal-01671905v1>

Submitted on 22 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Authors: Patricia Laurens, Lionel Villard, Antoine Schoen, Philippe Larédo

Title: The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents?

Authors' addresses:

Patricia Laurens: Université Paris-Est, CNRS – LISIS – IFRIS, 2, bd Blaise Pascal, 93160 NOISY LE GRAND (France)

Lionel Villard: Université Paris-Est, ESIEE – LISIS – IFRIS, 2, bd Blaise Pascal, 93160 NOISY LE GRAND (France)

Antoine Schoen: Université Paris-Est, ESIEE – LISIS – IFRIS, 2, bd Blaise Pascal, 93160 NOISY LE GRAND (France)

Philippe Larédo: 1-Université Paris-Est, LISIS – IFRIS, 2, bd Blaise Pascal, 93160 NOISY LE GRAND (France), 2- University of Manchester, Manchester Institute of Innovation Research, Booth Street East, Manchester M13 9SS (United Kingdom)

Corresponding author: Patricia Laurens

Email: patricia.laurens@esiee.fr

Tel: + 33 1 45 92 66 34

Abstract:

This paper proposes to broaden by more than 10% - compared with the current practice - the set of applications for priority patents, which is used to compute worldwide patent indicators. This extension is made possible thanks to the inclusion in the corpus of documents used for the calculation of indicators of the first filing patent applications that are designated as “artificial priority patents” in the PATSTAT database and currently discarded for the production of indicators. This research aims to show how adding these “artificial” patent applications can modify the value of the worldwide patent indicators. Artificial patent applications have never been used before because they contain very scarce information in their original state. We present a methodology we have developed to, first, replenish the artificial patents with information retrieved from close patents belonging to the same INPADOC family. Then, we study in details a range of indicators characterising the trends in the internationalisation of corporate R&D inventive activities. We provide evidence that the internationalisation pattern can be modified when including replenished artificial corporate patents in the indicator calculation. At the world level, incorporating artificial priority patents does not affect the trends over time, nor introduce any significant changes in the values of the indicators. However, analyses performed at a smaller scale, such as the firms’ continent level or the firms’ sector, show significant changes of the level of the intercontinental internationalisation in particular for the US firms.

Keywords: R&D; technology; patents; internationalisation; indicators; firms

MSC: 00A99

JEL: O34

Acknowledgements: This work was supported by RISIS, a project of the Infrastructure EU FP7 programme. The authors would like to thank Jos Winnink from CWTS (Leiden university) for his careful reading and valuable comments.

Introduction

Patents have become a growing source of information about knowledge dynamics. For a long time patents have been studied using the US patent office (USPTO) as a major source, mixing thus all patents first applied for in the US and patents extended from other offices because their grantees wanted to protect their rights on the most important market of the world (Griliches, 1990; Jaffe et al., 1993; Patel and Pavitt, 1997; Pavitt, 1985; Patel and Pavitt, 1991; Scherer and Harhoff, 2000; Harhoff et al., 2003; Harhoff et al., 1999). For this, owners used the long established international agreements on patents (the first one dating back 1883). Taking into account the internationalisation, the progressive deployment of the European Patent Office (EPO) and the growing role of Japan, the OECD promoted the idea that critical patents are those ‘extended’ to the key world markets; i.e. the US, Europe and Japan (the so-called triadic patents). Recently, Korea and China were included among the key world markets and new types of patents (IP5 patents or IP5 patent families) including IP protection in the Triad, plus China and Korea have appeared (IP5 Statistics Report, 2011; Dernis et al., 2015). This has two advantages: one is to avoid the home bias (i.e. all the patents deposited in one office); the other is that, for OECD, these patents represent the most valuable technology-based inventions. We retain two elements from these analyses that are further developed in section 2. First, in these approaches the first patent, the priority patent is key for capturing the origin of new knowledge; second the value of this new knowledge is linked to the strategy of its owners: the more markets covered, the more valuable is the invention. The focus on ‘priority patents’ is thus more and more considered as central and authors like de Rassenfosse (de Rassenfosse, 2013) have put it at the core of analysis of knowledge dynamics. But many patents remain orphan (i.e. never extended by their owners), so that it is important to consider patents that have been extended in other countries (even if it is not in all countries or markets, especially when analysts want to be more attentive to smaller firms). This is why, without discarding triadic and IP5 approaches, transnational patents - patents that have been extended in at least another patent office (or a selection of patent offices) - have become also more used.

These analytical developments have been made possible thanks to the development of a new data source, the PATSTAT database, produced by EPO, which includes all published patent applications and granted patents from all databases produced by national, regional and worldwide patent offices. PATSTAT links patents through priorities and builds different types of ‘patent families’ that correspond to priority patents and extended patents sharing priorities (extensions of patents in other offices, modification of patents like continuation, division, ...) (Martinez, 2011). But in quite a number of cases, patents are missing and PATSTAT is bound to create what are called ‘artificial priority patents’. This happens when in PATSTAT a published patent, not considered as a priority patent, mentions an earlier priority document that has not been delivered to the EPO and is thus unknown to the office. Such a priority document may be missing if the office where it has been filed has not published it (for any reason) or if the priority document is not a patent of invention (see further explanations in section 3). Artificial priority patents contain initially only scarce information. We are only provided with the patent office where it has been applied for, the date and the type of the applied document. In particular, the names and addresses of both the applicants and the inventors are missing as well as the IPC codes. Because of this lack of information we may miss a quite significant number of patents when counting patents considering priority patents. This can thus introduce a bias in overall analyses of knowledge dynamics, may this be at country, technology, sectoral and even more at firm or organisation levels. Section 3.1 further elaborates on this issue and evaluates the relative importance of these artificial patents: their importance (13% of total priority patents) requires that we further inquire their potential impact. We are not aware of any article that has conducted such an analysis before and we guess this has driven many analysts simply not to consider them when computing patent indicators.

This article thus focuses on reintroducing these artificial patents in the core PATSTAT database, and on evaluating their impact on indicators based on priority patent statistics. To analyse their role and impact on knowledge production, a first step is required: to retrieve information about them. Section 3 explains the methodology we have developed for doing so, using subsequent patent documents in patent families. Section 4 conducts an ‘impact’ analysis based upon previous work published on the patent portfolios of the largest industrial firms (Laurens et al., 2015) at firm, country and sector (selecting healthcare and ICT). It shows that this addition does not change the overall figures at country level but that it has a clear impact at sectoral level and, even more, at firm level for quite a significant number of major global firms.

II Priority patent indicators

The worldwide count of priority patents as an indicator of inventive activity has been introduced by de Rassenfosse (de Rassenfosse et al. 2013). The priority patent is the first filing, i.e. the initial patent application introduced for protecting a new invention. It is most often filed in the country of the inventor but it can also be applied for elsewhere in national, regional or worldwide patent offices when the applicant is targeting foreign markets¹. The building of indicators based on worldwide count of priority patents is now possible thanks to the regular update of the PATSTAT database. This well known and widely used database collects patents published in more than 180 patent offices and displays the same information on all patents following the structure of the EPO internal database DOCDB.

2-1 Limitations of the most widely used indicators

The worldwide indicator “counts priority patent applications filed by inventors from a given country regardless of the patent office of application” (de Rassenfosse et al. 2013). It has been designed in order to avoid the biases associated with the other widely used patent indicators that rely either on patents from a single patent office (USPTO or EPO), on PCT patent applications or on triadic patents - i.e. patents filed in all three USPTO, EPO and JPO (and more recently extended to the Korean and Chinese patent offices, the so-called IP5).

Building indicators based on patents applied for in a single patent office is known to suffer from home bias. Such indicators overestimate the weight of the domestic inventive activity compared to foreign inventions. As EPO is a regional patent office ensuring a simplified way to protect inventions in the countries that are member states of the European Patent Organisation, patent applications filed at EPO are considered as suffering less from home bias compared to national patent offices. However, EPO is only one possible way for firms to file patents, and, in particular for cost reasons, quite a number of large firms, when questioned, have told that they still patent mostly in their domestic office.

To avoid home bias, OECD created triadic patent indicators based on triadic and PCT patent applications to give a better picture of the global inventive activity. Triadic patents are designed to monitor inventions protected in large geographic and important economic zones. They thus concern inventions with large potential markets and inventions with a significant economic value. For analysts, they allow fair comparisons of the production of inventions across countries and over time. However, as it has been designed 15 years ago, one could reasonably ask whether an indicator based on triadic patents covering the US, Europe and Japan is still relevant. This limitation explains why OECD and the IP5 offices have proposed to enlarge the initial triadic indicators in order to cover additionally both China and Korea (IP5 Statistics Reports 2011 Edition; Dernis et al. 2015). It seems all the more

¹ According to de Rassenfosse (2013), priority patents are filed in the home country in most of the large countries. By contrast, in small countries (Switzerland, the Netherlands or Belgium), first filings at EPO dominate. In Canada or Israel, priority patents are massively filed at the USPTO.

relevant as SIPO, the Chinese Patent Office, has become since 2011 the first patent office in terms of the number of applications with a fast growing share of foreign applications.

Another approach analyses the PCT route, which has been established in order to simplify the procedures for international patent applications through a unified procedure for each of the 150 states that have signed the Patent Cooperation Treaty. The PCT route, operated by WIPO (World Intellectual Property Organisation), provides a period of 18 months for a possible extension, which adds to the 12 month priority period defined by the Paris convention, offering thus the applicants an extended period of time to assess whether or not the invention is worth to be protected in other markets. Besides, the applicants also benefit of the results of a mandatory “International Search” (IS)² for deciding what to do. Such reports are very appreciated as they provide a useful insight into the prior art that the applicant will have to deal with³. The delivery of the report can lead the applicant to an early withdrawal before entering into the expensive regional (EPO) or national phases (Guellec and van Pottelsberghe de la Potterie, 2000). These particular features of the PCT route make it attractive. However, if a patent protection is looked for only in a few countries, the direct filing of applications in the countries of interest may be less expensive and the time required to obtain the effective protection provided by the granting may also be shorter. Even if the PCT route has received a growing interest over time, it represented only 48% of the non-resident patent filings worldwide in 2005 (Wipo Patent Report, 2007). In 2013, the non-resident direct applications exceeded by large the non-resident PCT national phase entries in the patent offices in the UK, Germany and the United States, where they represent between 65% to 75% of the non-resident applications. In the Chinese and Japanese patent offices and at EPO, this share ranges between 30% and 45% (World Intellectual Property Indicators, 2014). These figures show that indicators relying only on PCT patent applications cannot give a full account of the overall inventive activities.

2.2 Interests and limitations of the priority patent indicators

Indicators based on worldwide priority patents rely on the priority applications filed worldwide under the Paris convention or the PCT route. They offer several advantages over the indicators examined above. First, they include all the priority patent applications wherever the applications are filed and do not suffer thus from home bias. Moreover, being the first filing for a new invention, the application date is the closest known date to the real date of the invention. It is thus used in statistics as a proxy to date the novelty of an invention (OECD, 2009). The indicators based on the worldwide priority patents allow to have a complete view of the inventive activity wherever the technical novelty has been created and whatever its value. They encompass the inventions created by entrepreneurs or those originating from emerging countries and are well adapted to study emerging technologies. Worldwide priority patent indicators have been introduced by de Rassenfosse (de Rassenfosse, 2013) to compare the inventive activity of inventors in different countries using the counting of families (see below for further explanations). They have been used by Picci to study internationalisation (Picci and Savorelli, 2012). In our research group, we have analysed the trends of the internationalisation of corporate R&D activities over time designing indicators of internationalisation based on the count of worldwide priority patents (Laurens et al., 2015).

Being exhaustive, the selection of priority patents suffers from an institutional bias: it includes patents of uneven values, as they are filed in patent offices where rules and application cost differ widely. Filtering the priority patents is an option to get rid of the very low value patents or to partially mitigate this institutional bias. This can be achieved by selecting, for example, only transnational priority patents, i.e. the priority patents that have been extended in at least a second patent office (Alkemade et al., 2015; Frietsch and Schmoch, 2010). Quite similarly, OECD now proposes to consider ‘IP5 patent

² Applicants can also ask for an optional “Supplementary International Search” and an “International preliminary examination”.

³ These reports are produced by one of the fifteen “International Search Authorities” (ISA) that have received the agreement by the International Bureau of WIPO. The choice of the patent office and of the ISA when filing a PCT application is strategic for many reasons: the price, the quality and the speed of production of the IS report vary.

families', that is to select transnational patents for which the extended family includes a patent filed in one of the IP5 patent offices (Dernis et al. 2015).

However, one of the main difficulties encountered when using worldwide priority patent indicators is the heterogeneity in the quality of information included in the patents. Missing items are frequent, such as the absence of the country of residence of the applicants and even more the inventors, the lack of detailed geographical information regarding inventors' and applicants' location. The absence of IPC code (International Patent Classification)⁴ is also a common situation. The following figures help to understand the problems faced. 58% of the priority patents filed between 2000 and 2005 lack information on the country of residence of the inventors (PATSTAT edition 2011) and this is more often found in patents applied for in national offices such as Japan, France or Brazil (de Rassenfosse, 2013). According to our own estimation (PATSTAT edition 2011), almost 23% of the priority patents filed between 1995 and 2010 do not provide information on the patent applicants and inventors. This share reaches 32% for the patents filed at the USPTO, 33% at EPO and 12% at JPO. An even higher share of patents includes only partial information on applicants' or inventors' addresses (missing cities or countries). Concerning the IPC classification, 16% of the priority patents do not include any IPC code.

Multiple methods have been used to recover missing information. Methodological approaches have been developed exploiting, when it exists, information found in "close" patents, i.e. patents related to the same invention and thus belonging to the same patent family. There are two approaches for doing so. Data can be retrieved from patents belonging to the same DOCDB family, made of patents citing exactly the same priority or group of priorities. They can also be retrieved using patents from the same INPADOC family. An INPADOC family is an extended group of patents that, two by two, share at least one priority: this means that all the members do not necessarily share all the priorities included in the family (Martinez, 2011). OECD (with REGPAT) and some offices (in our case, the French office, INPI) provide freely their own additions. For our research, which investigates internationalisation of R&D in large and R&D intensive firms, we have used the INPADOC approach to replenish the priority patents with all the key elements for analysis (Alkemade et al., 2015).

Quite a number of recent analyses have focused on the location of inventing activities, using inventor addresses as a source for identifying firm's research locations. When no information is present, a simpler approach consists in using the country of the applicant or the patent office where the patent has been filed as a proxy for the country of the inventor. An algorithm to fill the missing countries of residence of inventors was developed by de Rassenfosse using the retrieval of information in direct equivalent or second filing applications (de Rassenfosse, 2013).

2.3 Artificial priority patents

Most of the worldwide priority patent applications included in the PATSTAT database are the published patents originating from the EPO internal database, DOCDB, that constitutes the core PATSTAT database. However, there is in PATSTAT an additional group of priority patents, classified as "artificial patents" that are identified by a selected range of identifiers⁵. PATSTAT database designers have been faced - in approximately one case out of ten - with patents that are not considered as priority patents as they claim as a priority patent a prior document in the same office or in another office although this claimed priority patent cannot be retrieved⁶. PATSTAT designers have thus created an "artificial PATSTAT application due to priorities" that just has the basic information mentioned in the subsequent patent: the date of filing, the office and the type of application of the

⁴ IPC codes define the technological classes to which the patent corresponds. They can be very detailed since, at the lowest level, there are around 75000 categories.

⁵ Rules for identifying the artificial applications are given in the PATSTAT database manual.

⁶ "If we can not identify the application claimed as a priority with 100% certainty, we have created artificial entries in the table TLS201_APPLN for these prior applications assuming the priority data in the document is 100% accurate" (European Patent Office, 2011).

priority document⁷. All artificial priority applications thus correspond to inventions that have been mobilised as claimed priorities in further subsequent applications. They thus correspond to valuable knowledge since their owners have judged important to further protect them. Counting patent applications by patent office or application date in the PATSTAT database automatically includes both artificial patents⁸. Conversely, counting patent applications by applicants' or inventors' countries will not include the artificial patents as the corresponding information is missing.

We thus do not share the approach by Hall and Helmers (2013), who, after having identified the large number of patents with missing priorities in PATSTAT, have nevertheless simply proposed to forget them. The authors "simply allowed the patent to serve as its own priority".

This paper focuses on ways in which we can mobilise these artificial patents in the study of priority patents and it tries, using our work previously done on large firms, to evaluate their impact.

III Methodology

In this section we present the method set up to retrieve information in order to replenish artificial priority patents of the PATSTAT database. We start by a characterisation of artificial patents, which helps to further understand the reasons and strategies which can explain such status.

3.1 Artificial priority patents in the PATSTAT database

The overall PATSTAT database (version October 2011) contains 61 570 794 applications which have a related publication stored in the DOCDB database and 6 561 807 applications claimed as 'artificial priority applications'⁹. The latter are not present as applications in the DOCDB database but, for the sake of database integrity, they are clearly mentioned in INPADOC families. Between 1995 and 2010, PATSTAT database contains 1 705 454 artificial priority patents of invention¹⁰. They account for 13.6 % of all the priority patent applications during this period of time. This share remains stable across the different versions of the PATSTAT database (e.g. 13.15% in the PATSTAT edition spring 2014) but it slightly increases over the patent application filing years in a given edition.

More than 50% of the artificial priority patents in the PATSTAT database are applications filed in the US and 25% in Japan (Figure 1). The rest originates from patent offices in Germany (4.7%), EPO (3.9%), in Korea (2.9%), Australia (2%) or China (1.6%).

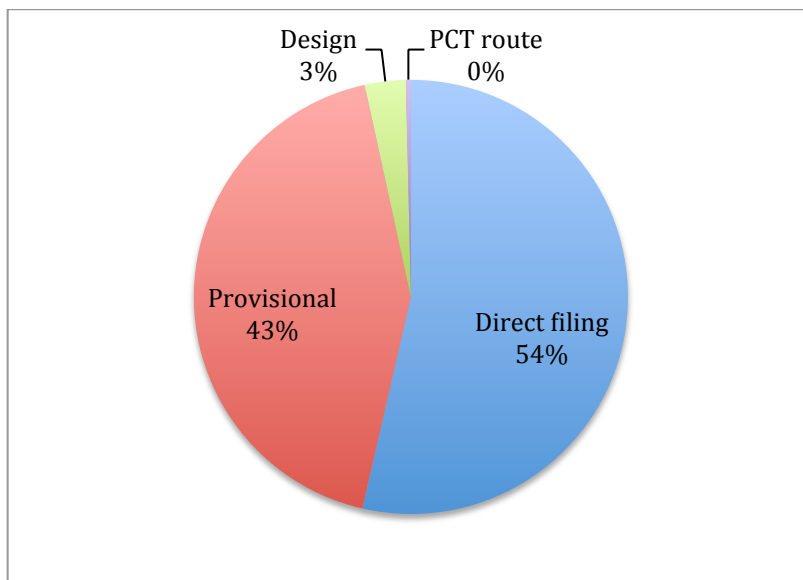
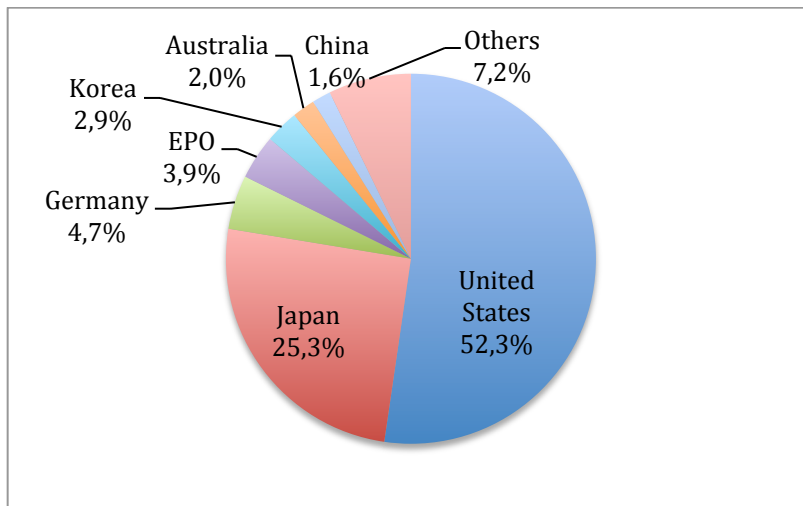
Figure 1: Distribution of the artificial priority patents according to the patent office and patent type.

⁷ To be exhaustive, there is a second source of artificial patents: those mentioned in publications cited by patents but that cannot be retrieved, so called "artificial PATSTAT applications due to unknown cited publications". This paper does not cover this second source of artificial patents.

⁸ Such a count is obtained using a simple query in table TLS201_APPLN.

⁹ According to the figures given in the EPO Worldwide Patent Statistical Database 2016, the overall share of artificial priority patents has decreased from 10,7% in the PATSTAT 2011 to 9,8 % in the PATSTAT 2016 (European Patent Office, 2016).

¹⁰ Patents of invention are patents with `ipr_type = PI` in PATSTAT.



Among the largest patent offices in terms of number of filed priority patents, the share of artificial priority patent applications varies significantly: the artificial priority patents represent approximately 1/3 of the priority patents applied for in the US and at EPO, around 10% in Germany and Japan, between 1% and 4% in Korea, France and China and below 1% in Italy, Russia and the UK. In small European countries (Sweden, Switzerland, Denmark), as well as in India or South Africa, this share exceeds 50% ¹¹. These numbers and percentages of artificial patents in PATSTAT show that excluding the artificial priority patents in the calculation of worldwide priority patent indicators leads to ignore a significant share of the inventive activity originating from two of the five largest patent offices - the EPO and USPTO - and of many countries.

The distribution of worldwide artificial patent applications shows that artificial priorities come mostly from two routes: direct filings (54%) and provisional patents (43%).

A provisional patent application is a low-cost national first patent filing allowed in countries like the United States, Australia, Canada and India. It does not require the owners to include formal patent claims nor prior art statement. It provides the means to establish an early effective filing date to be used for a subsequent non-provisional patent application within the following year and is also useful to

¹¹ The share of artificial priority patents tends to decrease over time in most countries, except in Japan.

extend the duration of a patent to twenty-one years from the first filing. A provisional patent application also delays the eventual higher cost of filing a non-provisional application (Crouch, 2014; Chen, 2016). When a subsequent non-provisional patent is filed, it refers to the provisional patent as a priority, the latter appearing as an artificial priority patent in PATSTAT. This path is extensively used at the USPTO with 27% of all priority patents filed between 1995 and 2010 (731 792 provisional priorities, which acquired a subsequent status of artificial patents). This represents 85% of the artificial priority patents originating from this office¹².

Direct filings following the Paris convention represent 54% of total artificial priorities. They correspond to the following situation. A patent has been filed in one office and the applicant has 12 months for filing applications in other patent offices: these extensions are called 'Paris convention applications'. The priority patent becomes an artificial priority when it has not been published in the office where it was filed¹³. This happens when the applicant withdraws it before publication, while keeping the 'Paris convention application' active. When a direct first filing is further extended following the PCT route, it also often appears as an artificial priority. This strategy gives the applicant 30 months to file the same patent in the selected countries. The PCT procedure includes a search, which is often highly rated by applicants (see above); it offers applicants the possibility to buy time (2,5 years against 1 in the Paris convention) but it has a high fee.

Let us take the European patent office as an illustration. 65 286 artificial priority patents, among the overall 196 971 priority patents, have been filed at the EPO between 1995 and 2010: 99.7% are Paris convention priorities. These artificial priority patents applied for at the EPO belong to 58 375 different INPADOC families and 82% of these INPADOC families include a PCT patent application as a subsequent filing¹⁴. These basic statistics show that a high share of artificial EP priority patents are related to a subsequent PCT patent application filed either at the EPO (almost 2/3 of them) or at the WIPO office in Geneva (1/3 of them). According to our understanding, such an artificial EP priority patent is a first filing, which has been applied for at the EPO, then further extended using the PCT route, and eventually withdrawn by the applicant before its publication. Such a strategic intellectual property (IP) management, consisting in the withdrawal of a priority patent prior to its publication but after its extension through a PCT route, offers several advantages: besides allowing to claim the priority date of the EP priority in the PCT extension, it delays the choice of the countries where to proceed in and the fees and enables to benefit from the EPO search. The EPO as an International Search Authorities seems to be one of the best options since the office offers a high quality search at a reasonable speed (Inovia, 2014)¹⁵.

A similar situation is observed for applicants filing a priority patent at JPO. Like EPO, 99.5% (420 000 patents) of the artificial priorities are Paris convention priorities. Similarly, the share of artificial priority patents filed at JPO and related to a subsequent PCT patent application (through INPADOC family links) exceeds by far the share of non artificial priority patents linked with PCT applications (50% vs. 2%). In Germany, 71% of the artificial priorities are Paris convention priorities, but this corresponds to another strategy since 66% are internal patents, i.e. national priority filings applied for in Germany followed by a subsequent filing also in Germany claiming the priority of the first application.

To sum up, the statistics show that the majority of the artificial priority patents present in the PATSTAT database are domestic priorities further extended in their own countries (provisional patents and continuation patents in the United States, internal patents in Germany). The rest is a set of priority patents further extended in a worldwide patent protection via the PCT route. In one word,

¹² 12,5% of non provisional artificial patents filed in the US are priority parent patents followed by continuation patents and continuation in parts. Continuation patents permit to add new claims to the invention disclosed in the original application, while continuation in parts disclose new subject matter (Martinez, 2011).

¹³ Or published in the office but the information related to this filing was not transmitted by the office to the EPO.

¹⁴ The ratio is only 44% when the EP priority is not artificial.

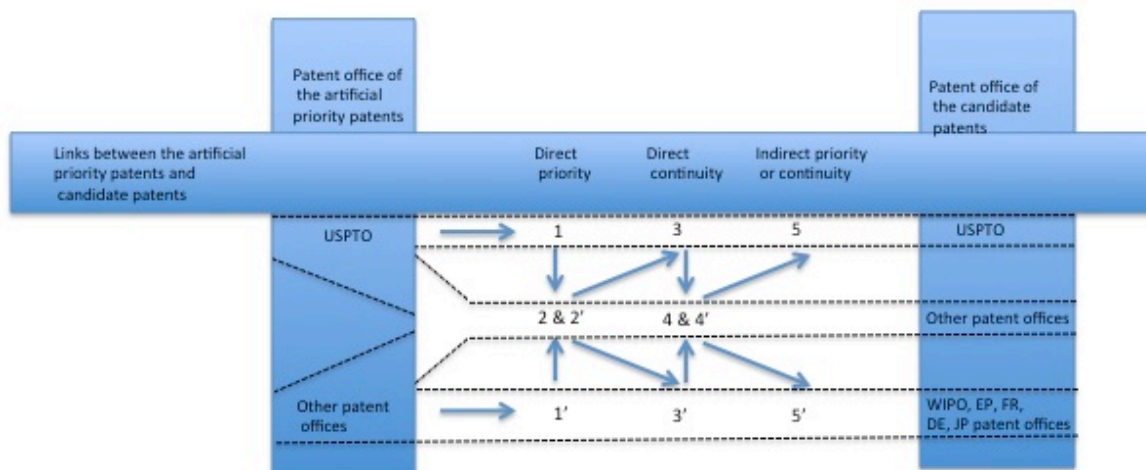
¹⁵ This is why numerous US applicants often file both a US application and a PCT application choosing EPO as the ISA in order to benefit from a second opinion to complement the one originating from the US examiner.

artificial priority patents combine two dimensions: first, they are linked to national IP laws, which aim at improving or facilitating locally the protection of inventions. They depend, second, on corporate strategies developed by applicants aiming at a worldwide protection of their inventions via the PCT route. Artificial patents reflect therefore the skills developed by the applicants for optimising IP protection. The artificial priority patents are thus all the more important that they appear as pointing to inventions that the applicants have taken a particular care to protect.

3.2 Replenishing information in artificial priority patents

With information that only relates to the application date and the patent office, the artificial patents in Patstat are incompletely filled. In order to include this set of patents when calculating indicators, it can become crucial to access to the names and addresses of the applicants and inventors of the patents or to its area of technology. It is thus important to find a way to recover lacking information. The filling of both the applicants' and the inventors' names and addresses but also of IPC codes in the artificial priority patents is based on the identification and use of pieces of information retrieved from non artificial patents belonging to the same INPADOC family. To select the candidate patent, which the information is retrieved from, a set of sequential rules has been defined in order to select a candidate patent as close as possible to the artificial one. The criteria relate to the type of information available in the candidate patent, the nature of the priority link between the artificial patent and the selected candidate patent, the time lag between the two applications and the patent offices. In order to be considered as a potential candidate, the patent must contain information concerning addresses, either the inventors' or the applicants' addresses or both. A candidate with a direct priority (Paris convention) link with the artificial patents is selected first (from table t1s 204 of the PATSTAT database). In absence of such a link, a candidate with a direct continuity link is selected (continuation, division/continuation in parts/internal continuities in table t1s 216). At last, candidate patents with indirect links are selected. Besides the type of links, we also consider the time lag between the artificial and the candidate patents. For two candidates with the same link, we select the candidate with the date of filing closest to the filing date of the artificial patent (the maximum authorised time lag has been fixed to five years). Finally, we give a higher priority to US candidates to fill US artificial patents. For non-US artificial patent applications, candidates filed as PCT, or EP, JP, FR, DE received the highest priority. A schematic view of the rules of selection of candidates chosen for replenishing artificial patents is presented in figure 2. As soon as a candidate patent is selected, the information it contains regarding the names, addresses and IPC codes is used to fill the artificial patent.

Figure 2: Schematic description of the different steps to fill inventors' and applicants' addresses in artificial priority patents.



This procedure worked very well as we were able to propagate information from a candidate patent in 99.3% of the artificial priority patents filed between 1995 and 2010, i.e. 1 728 234 documents over a total of 1 740 763 documents. 92.6% of the propagated information originates from a candidate with a direct priority link and 6.2% from a candidate with a direct continuity. The rest (less than 1%) originates from a candidate with an indirect link with the artificial patent (Table 1). For most patent offices, the replenishing of artificial patents heavily relies on candidates with a direct link. However, in Germany, but also in the US and Australia, the artificial patents also benefit from information available in continuity direct links.

Table 1: Replenishing information in artificial priority patents: type of links between the artificial patents and the selected candidate patents in the offices with the largest number of artificial patents.

Country of the patent office of artificial patents	Direct priority (%)	Direct continuity (%)	Indirect priority (%)
United States	90.8	8.5	0.7
Japan	99.7	0.0	0.0
Germany	68.5	30.8	0.2
EPO	96.2	0.2	0.2
Korea	97.2	0.1	0.7
Australia	92.2	5.8	1.3
China	98.7	0.0	0.3
All offices	92.6	6.2	0.5

Information on applicants was recovered for 87% of artificial patents and the information on inventors for 95% artificial documents. For 86% of artificial patents, both the applicants' and inventors' data have been propagated. In 85% of the cases, it also gives the country of location for the applicants' and the inventors' addresses. This corresponds to the retrieval of 2 849 211 additional applicants' names and 4 560 439 additional inventors' names. It expands the initial list of priority patent applicants by 25% and the list of inventors by 20%.

These high retrieval percentages show that the filling step of artificial patents was very effective. It thus provides a better coverage of the overall inventive activities when using the priority patents, in particular in patent offices where the share of artificial patents is high (EPO or USPTO).

IV Measuring the impact of artificial priority patents: the case of the R&D internationalisation of large firms

After a brief description of the Corporate Invention Board (CIB) patent database, we show to which extent the replenishment of information in the artificial patents of PATSTAT has allowed to enlarge this corporate priority patents database. We then investigate if changes of the values of the priority patent indicators can be detected in the case of the internationalisation of the inventive activities of the set of large firms with the highest R&D investments that compose this database.

We use the Corporate Invention Board (CIB) to measure the impact of the addition of artificial patents. The CIB gathers the patents of the 2000 largest firms identified by the IPTS scoreboard (Industrial R&D Investment Scoreboard Report, 2009) complemented by top private applicants from WIPO, EPO and USPTO rankings. The CIB considers groups, including all majority owned subsidiaries (in total 316 000 firms' names). The firms' patents have been retrieved in the PATSTAT

database following a procedure described in Laurens et al. (2015). The propagation of information in artificial priority patents has enlarged the list of applicants in PATSTAT to be matched with the firms' names by 25%. The enlarged overall database of corporate patents includes 5 659 074 priority applications, from which 13% (736 727 documents) are artificial patent applications. This adds 1,4 million applicants, i.e. more than 20%, to the final list of applicants and 2,3 million additional inventors, i.e. 17%, to the final list of inventors.

For the sake of comparisons with previous published work (Laurens P. et al., 2015, Alkemade et al., 2015) we then only retrieve patents of invention (type A and W)¹⁶ and, in order to discard inventions only protected in the local market, we solely consider applications of transnational priority patents (see section 2 above). We also only consider the subset of firms with a sustained inventive activity over the period¹⁷. We thus end up with a subset of 1048 firms (see annex 1 for the countries of origin and the industrial sectors). They own almost 1,3 million transnational priority patents applied for from 1994 to 2009; 20% are artificial patents¹⁸. Artificial patents originate first from Japanese firms and inventors in Japan (57%-58%), then from US firms (9.3%) and inventors in the US (7.3%) or German firms (8.4%) and inventors in Germany (10.1%). On average ¼ of transnational priority patents of Japanese firms (or involving Japanese inventors) are artificial patents. In US firms, this share is 10.7%. It reaches 13.5% for German firms (Table 2).

Table 2: Share and distribution of artificial patents by country of residence of the inventors and country of firms (top 9 countries) for transnational priority patents applied for from 1994 to 2009.

Inventors' country	Number of transnational priority patents	Share of artificial (%)	Distribution of artificial (%)	Firms' country	Number of transnational priority patents	Share of artificial (%)	Distribution of artificial (%)
Japan	573 781	26.3	57.6	Japan	583 107	25.9	57.5
US	211 654	9.1	7.3	US	228 653	10.7	9.3
Germany	188 999	14.0	10.0	Germany	163 114	13.5	8.4
Korea	95 498	10.4	3.8	Korea	97 148	10.5	3.9
France	59 769	8.4	1.9	France	58 913	6.0	1.4
Netherlands	25 239	75.0	7.2	Netherlands	45 970	52.9	9.3
UK	18 800	21.2	1.5	Switzerland	23 447	34.3	3.1
Sweden	16 870	22.9	1.5	Sweden	20 814	13.2	1.1

¹⁶ Type A refers to patent and W to PCT patents. We exclude all other type of documents such as utility models, design patents and any other type of documents.

¹⁷ i.e. which have applied for at least 5 patents during each of the two periods of time, 1994 to 1996 and 2003 to 2005.

¹⁸ The share of artificial patents decreases over time (from 24% in 1994 to 17% in 2009).

Italy	14 416	9.0	0.5	Finland	14 926	4.0	0.2
Total	1 296 792	20.2	100.0	Total	1 296 792	20.2	100.0

We now analyse the impacts of artificial priority patents on internationalisation indicators, first on the whole set of firms, then at a sectoral level considering two research intensive industrial sectors (ICT and healthcare) and finally on individual firms. For this, we compare sets of indicators calculated including artificial patents to those calculated when excluding them.

4.1 Impact on overall internationalisation indicators

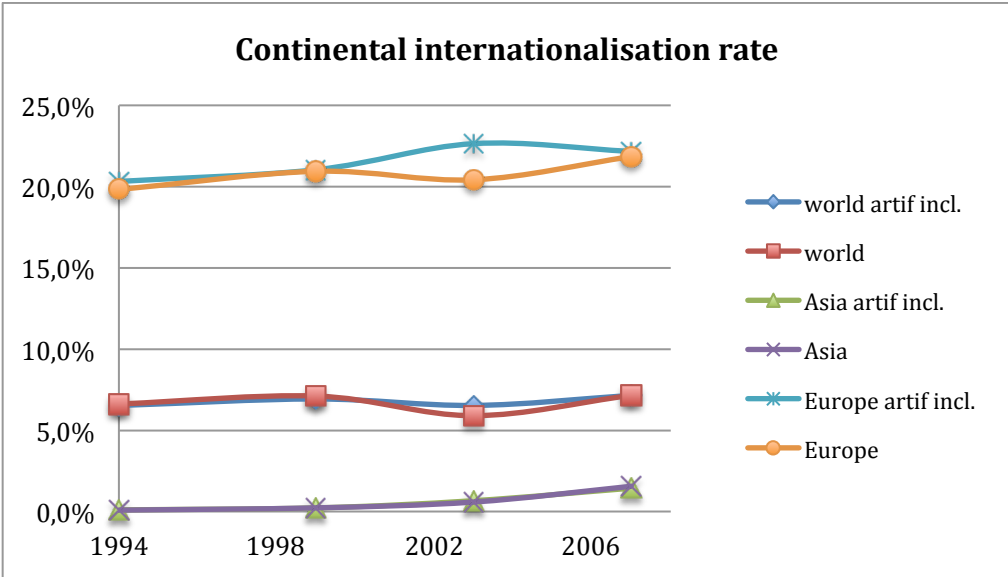
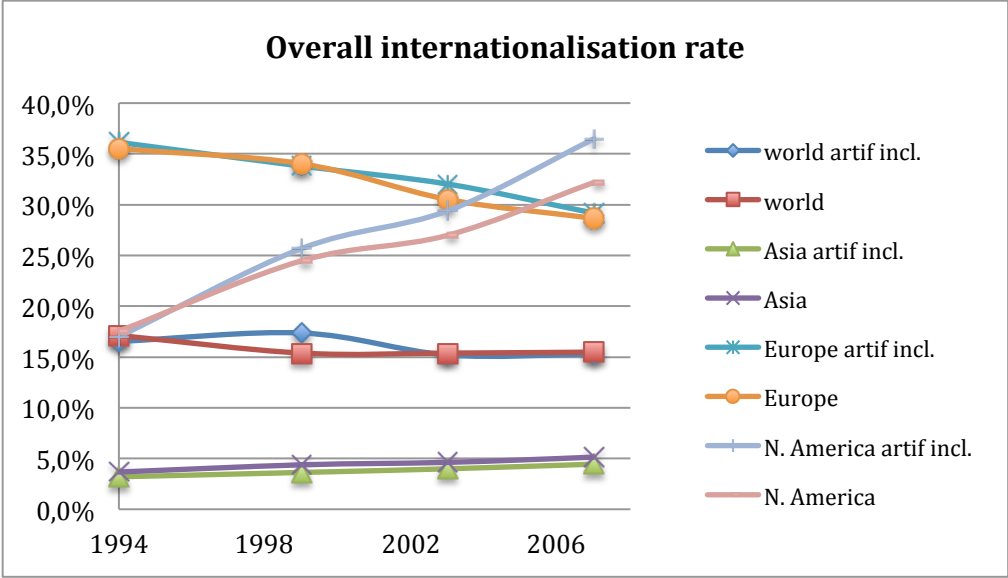
We use three indicators to characterise internationalisation: first overall internationalisation i.e. the share of inventions where the country of residence of the inventors differs from the location of the firm's headquarters; second continental internationalisation, very important for analysing europeanisation dynamics, i.e. the share of inventions where the country of residence of the inventors differs from the location of the firm's headquarters but both remain located in the same continent; third the share of intercontinental internationalisation, i.e. the share of inventions where the continent of the inventors differs from that of the firm's headquarters. Figures 3 a, b and c present results.

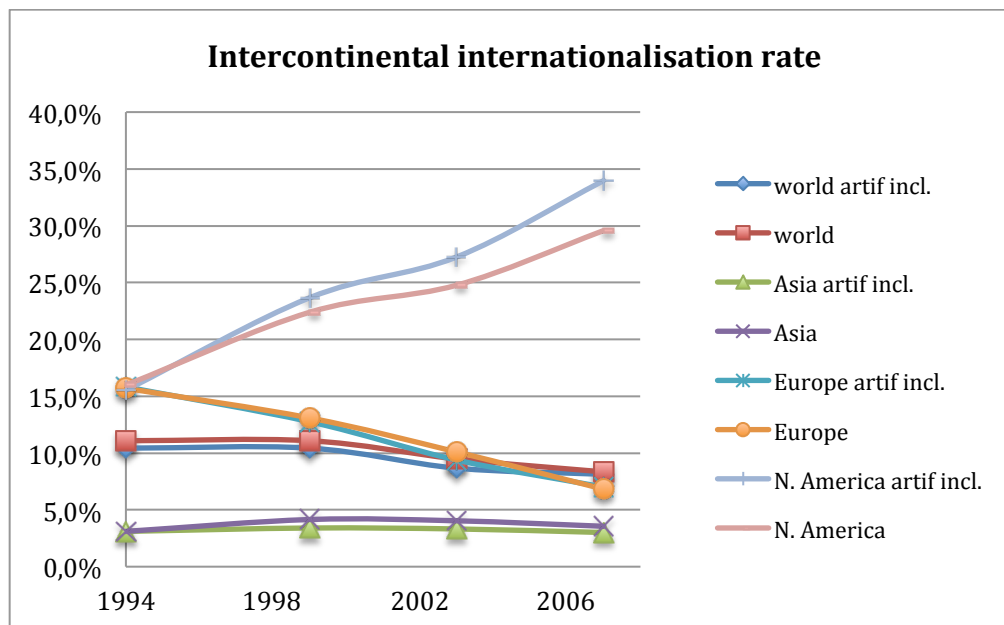
At the world level, the overall internationalisation has lost approximately 1 point over 13 years to stabilize around 15% in the late 2000 (Figure 3-a). This stable internationalisation level results from the combination of a slight increase of the intercontinental internationalisation rate (+ 0.5 point to reach 7% in 2007) (Figure 3-b) and a moderate decrease of the continental internationalisation rate (-1 point to reach 8%) (Figure 3-c). The inclusion of artificial patents in the calculation of the internationalisation indicators does not change these trends, nor the values of the three indicators.

However the situation differs slightly when considering trends at continental level. Nothing changes for Asia and Europe. Including or not the artificial patents, we still monitor for Asia, a weak increase of the overall internationalisation to 5% (Figure 3-a) due to a nascent continental internationalisation (1.5% in 2007) (Figure 3-b). And for Europe, adding artificial patents does not change the overall decrease from 35% to 29%, due to a drop by a factor of two of the intercontinental internationalisation (from 16% in 1994 to 7% in 2007).

This is not the case for North America, where we witness a growing overseas internationalisation over time. But the trend is more pronounced when including the artificial priority patents: in 2007, the overseas internationalisation was 29% without the artificial patents; it exceeded 34% including them (Figure 3-c). The addition of the artificial patents emphasizes the growth of the intercontinental internationalisation of US firm R&D activities. Furthermore, the gap between the level of the overseas internationalisation including or not artificial patents widens over time.

Figures 3 a,b,c: Internationalisation indicators over time





4.2 Impact at industrial sector level

Annex 2 shows that the share of artificial patents is unevenly distributed among the technology areas of the patents. The average share is 19% but it varies from 9% to 77% depending on the 35 technology domains (WIPO classification) considered. For instance it is around 10% for computer technology, digital communication or telecommunications, 16.9% for optics, but it rises to 20.7% for chemical engineering, 38% for materials and 39% for macromolecular chemistry. And it is even far higher for biotechnology (55.7%), pharmaceuticals (69.4%) and food chemistry (76.8%).

To have a better idea of the impact of these differentiated rates of inclusion, we decided to move to an analysis of the effects at the level of the industrial sectors of firms. Industrial sectors are built following the ICB classification and all firms are allocated (as is done in the IPTS scoreboard) to one industrial sector only. For better illustrating the effects, we further selected two research intensive industrial sectors: the ICT sector (which gathers both hardware and software services) and the healthcare sector (which includes pharmaceuticals and biotechnology). We restrict our analysis to the set of priority patents filed between 2003 and 2005.

Adding the artificial patents to the overall database including all industrial sectors enlarges the volume of the 2003-2005 patent dataset by 20% but it only slightly changes the overall rate of internationalisation: from 12.3% to 15.2%. It has different effects depending upon continents: it slightly reduces the overall level of internationalisation for Asian firms (from 4.6% to 4%) while it increases the rate for European firms by 1 point (from 30.5% to 32%) and of US firms by 2 points (from 27 to 29.4%). Firms from the ICT sectors by and large share these trends (from 13.9% to 15.3%). In ICT, the share of artificial patents is below the average (13.3%). The situation is quite different for the healthcare sector where the share of artificial patents is higher (39.2%). The overall internationalisation increases from 30.1% to 33.3% but it changes radically both for European and US firms: in European firms, the rate increases by 4 points (from 61.0% to 65.0%) and in US firms the internationalisation rate jumps from 27.9% to 35.6%.

4.3 Impact at firm level

The share of artificial transnational priority patents for the 20 largest firms' patent portfolios (most of them being located in Japan) varies strongly from 95.8% (Philips) to 5.6% (Robert Bosch) (see annex 3). It is above the average (20%) for eight firms, most often Japanese firms (6 firms). Amongst the largest firms (more than 5 000 transnational patents filed between 2003 and 2005), Panasonic and Philips have a share of artificial patents that is close or exceeds 50%. It is also the case for large pharmaceutical firms like Takeda, Daichi Sanyo or Eisai in Japan, Astra Zeneca or Novartis in Europe and Schering Plough or Merck & co in the United States. As a whole, the firms from the pharmaceutical sector show a high share of artificial patents. This is correlated to the observation of a higher share of artificial patents in the technology fields related to chemistry and pharmaceutical products or processes (« Pharmaceuticals », « Biotechnology », « Analyses of biological materials », « Organic fine chemistry » ...). (see Annex 2). Crouch has also found that the field of new drug invention was a field where the share of provisional applications was the highest. According to him, the pharmaceutical sector uses artificial applications first to provide a potential extra year of protection that can be highly valuable (Crouch, 2014).

Moreover, The distribution of the firms by country of residence and sector are given in annex 1.

To conclude, the propagation of information on applicants in the artificial patents of the PATSTAT database prior to the matching with firms' names has impacted the building of the corporate patent database in two ways. It has increased the number of corporate priority patents (+13%) and has also modified the distribution of the corporate patents according to the country of residence of inventors, location of firms, industry sectors and technology fields.

Conclusion

When looking at knowledge dynamics, it is important to know the initial dates at which owners have protected the new knowledge they have produced. PATSTAT, the database that gathers the databases of all individual patent offices, has difficulty to retrieve the original priority patent of more than one patent out of ten (13% for the global database). PATSTAT thus includes 'artificial patents' that are labelled with a special range of identifier and contains nearly no information, apart from its date and office of filing and the extended family it belongs to. This may well explain why little attention has been given to these artificial patents, while many authors have mentioned the importance of patents without an identified priority.

These artificial patents are linked to legal aspects both at national (especially the US and Japan) and international level (the differences between Paris convention and PCT). Firms 'optimise' their IP protection for extending the duration of the patent by one year (using when possible 'provisional' patents), or by delaying for up to 2.5 years the choice of the countries where to extend the priority patent.

To address this issue, we have developed a method to replenish the artificial priority patents by propagating information from close patent documents (belonging to the same families). This has made possible to fill information for 99% of the artificial priority patents.

In order to assess the impact of such replenishment, we have used previous work done on the internationalisation of large firms. The results show that when looking at global indicators (covering the whole database), this inclusion does not change global figures. But, as soon as we zoom at technological or at sectoral levels impacts differ widely between technologies (from 6 to 62%) and at sectoral level: the case of healthcare (including pharmacy and biotechnology) is illustrative of the extend of impact. It is even more pronounced when zooming down at the 'micro' level of firms, where it seems difficult to analyse robustly knowledge portfolio or strategy without taking into

consideration the optimisation practices; it strongly impacts any comparative approach that could be developed.

Based on the results of this work, we strongly recommend assessing in depth the changes that the inclusion of artificial patents might introduce in priority patent-based indicators rather than discarding them without further consideration.

References

Alkemade, F., Heimeriks, G., Schoen, A., Villard, L., Laurens, P. (2015). National and sectoral characteristics of the internationalization of the inventive activity of multinational corporations. *Research Policy* 44, 1763-1772.

Hall, B.H., Helmers, C. (2013). Innovation and diffusion of clean/green technology: Can patent commons help?. *Journal of Environmental Economics and Management* 66, 33-51.

Chen, C.T., Chen, D.Z. (2016). Who files provisional applications in the United States?. *Scientometrics* 107, 555-568.

Crouch, D. (2014). A first look at who files provisional patent applications (downloaded at <http://patentlyo.com/patent/2008/06/a-first-look-at.html>).

Dernis, H., Dosso, M., Hervás, F., Millot, V., Squicciarini, M., Vezzani, A. (2015). World corporate top R&D investors: Innovation and IP bundles. A JRC and OECD common report. Luxembourg: Publications Office of the European Union.

European Patent Office, Data Catalog for the EPO Worldwide Patent Statistical Database, 2016 distributed to EPO PATSTAT customers.

Frietsch, R., Schmoch, U. (2010). Transnational patents and international markets. *Scientometrics* 82(1), 185-200.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature* 28 (4), 1661-1707.

Guellec, D., van Pottelsberghe de la Potterie, B. (2000). Applications, grants and the value of patent. *Economics Letters* 69, 109-114.

Harhoff, D., Narin, F., Scherer, F.M. (1999). Citation frequency and the value of patented inventions. *Review of Economics and Statistics* 81, 511-515.

Harhoff, D., Scherer, F.M., Vopel K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy* 32, 1343-1363.

Industrial R&D Investment Scoreboard Report, 2009. Institute for Prospective Technological Studies of the European Commission IPTS, (downloaded at <http://iri.jrc.ec.europa.eu/scoreboard12.html>).

The 2014 US Global Patent & IP Trends Indicator. 2014. Inovia. (downloaded at www.inovia.com).

IP5 Statistics Report 2011 Edition,
(downloaded at www.fiveipoffices.org/statistics/statisticaldata.html).

Jaffe, A.B., Trajtenberg, M., Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108(3), 577-598.

Laurens, P., Le Bas, C., Schoen, A., Villard, L. Larédo, P. (2015). The rate and motives of the internationalisation of large firm R&D (1994-2005): Towards a turning point?. *Research Policy* 44, 765-776.

Martinez, C. (2011) Patent families: When do different definitions really matter?. *Scientometrics* 86(39), 39-63.

OECD, 2009. *OECD Patent Statistics Manual*. OECD Publishing, pp. 29–33.

Patel, P., Pavitt, K. (1991). Large firms in the production of the world's technology: an important case of 'Non-Globalisation'. *Journal of International Business Studies* 22(1), 1-21.

Patel, P., Pavitt, K. (1997). The technological competencies of the world's largest firms: complex and path-dependent, but not much variety. *Research Policy* 26(2), 141-156.

Pavitt, K. (1985). Patent statistics as indicators of innovative activities: possibilities and problems. *Scientometrics* 7(1-2), 77-99.

Picci, L., Savorelli, L. (2012). Internationalized R&D activities and technological specialization: an analysis of patent data. OECD paper (downloaded at <https://www1.oecd.org/site/stipatents/6-3-Picci-Savorelli.pdf>).

de Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., van Pottelsberghe de la Potterie, B. (2013). The worldwide count of priority patents: a new indicator of inventive activity. *Research Policy* 42, 720-737.

Scherer, F.M., Harhoff, D. (2000). Technology policy for a world of skew distributed Outcomes. *Research Policy* 29, 559-566.

World Intellectual Property Indicators 2014, Economics and Statistics Series published by the World Intellectual Property Organisation.

The WIPO Patent Report, 2007 Edition, edited by the World Intellectual Property Organisation (downloaded at www.wipo.int/ipstats)

Annex 1

Distribution of firms according to their headquarters' country

Location of firms	Number of firms	Share of firms (%)
North America	365	34.8
United States	350	33.4
Europe	401	38.3
Asia	274	26.1
Other	8	0.8
Total	1048	100.0

Distribution of firms according to their industry

Sector of industry	Number of firms	Share of firms (%)
Basic Materials	27	2.6
Consumer Goods	126	12.0
Consumer Services	325	31.0
Financials	178	17.0
Health Care	139	13.3
Industrials	23	2.2
Oil & Gas	12	1.1
ICT	24	2.3
Telecommunications	14	1.3
Utilities	180	17.2
Total	1048	100.0

Annex 2

Share of artificial transnational priority patents applied between 2003 to 2005 by technology fields in the CIB database (35 technology fields according to the WIPO classification)

Technology fields	Distribution of patents over technology fields (%)	Share of artificial (%)
Food chemistry	4.8	76.9
Pharmaceuticals	8.3	69.4
Organic fine chemistry	7.1	60.2
Biotechnology	8.5	55.7
Basic materials chemistry	3.4	47.9
Analysis of biological materials	19.4	44.4
Macromolecular chemistry. polymers	0.9	39.2
Materials. metallurgy	14.6	38.2
Environmental technology	7.5	21.9
Electrical machinery. apparatus. energy	3.6	20.9
Chemical engineering	0.5	20.7
Textile and paper machines	1.2	20.3
Other special machines	3.8	19.2
Micro-structural and nano-technology	0.9	18.7
Medical technology	0.7	17.7
Optics	3.5	16.9
Surface technology. coating	0.3	16.4
Measurement	0.1	16.0
Handling	0.6	15.3
Mechanical elements	0.4	15.2
Transport	1.2	14.9
Audio-visual technology	0.2	14.8
Engines. pumps. turbines	0.6	14.4
Semiconductors	0.2	14.2
Basic communication processes	1.2	12.9
IT methods for management	0.8	12.7
Civil engineering	0.8	12.5
Machine tools	3.3	12.4
Thermal processes and apparatus	0.4	11.8
Telecommunications	0.3	11.6
Control	0.3	11.4
Digital communication	0.4	10.3
Furniture, games	0.2	10.3
Computer technology	0.4	9.3
Other consumer goods	0.1	9.0
Total	100.0	19.0

Annex 3

Share of artificial transnational priority patents in firms (priority applications: 2003-2005)

Share of artificial patents in the 20 largest firms

Name of firms	Country	Sector	Number of transnational priority patents	Share of artificial (%)	Weight of the firm's patents in CIB (%)
SAMSUNG ELECTRONICS	Korea	Industrials	17582	8.6	5.6
PANASONIC	Japan	Consumer Goods	10464	48.9	3.3
HITACH	Japan	ICT	9501	13.4	3.0
LG	Korea	Basic Materials	9344	8.8	3.0
TOSHIBA	Japan	ICT	8195	12.2	2.6
SIEMENS	Germany	Industrials	7675	12.9	2.5
CANON	Japan	Industrials	7100	26.0	2.3
FUJIFILM	Japan	Consumer Goods	6771	25.4	2.2
FUJITSU	Japan	ICT	6387	6.0	2.0
PHILIPS ELECTRONICS	Netherlands	Consumer Goods	6163	95.8	2.0
SONY	Japan	Consumer Goods	6109	13.8	2.0
ROBERT BOSCH	Germany	Consumer Goods	5319	5.9	1.7
SEIKO EPSON	Japan	ICT	4865	25.6	1.6
NEC	Japan	ICT	4444	26.8	1.4
IBM	US	ICT	4427	13.5	1.4
DENSO	Japan	Consumer Goods	3872	19.4	1.2
HONDA MOTOR	Japan	Consumer Goods	3559	10.2	1.1
TOYOTA MOTOR	Japan	Consumer Goods	3500	13.0	1.1
SHARP	Japan	Industrials	3273	27.9	1.0
RICOH	Japan	ICT	2993	35.1	1.0

Firms with the largest share of artificial patents in Asia, Europe and United States (for firms with more than 50 transnational priority patents)

Name of firms	Country	Sector	Number of transnational priority patents	Share of artificial (%)	Weight of the firm's patents in CIB (%)
ONO PHARMACEUTICAL	Japan	Health Care	97	99.0	0.03
TAKEDA PHARMACEUTICAL	Japan	Health Care	193	96.9	0.06
DR REDDY'S LABORATORIES	India	Health Care	73	95.9	0.02
ASTELLAS PHARMA	Japan	Health Care	114	95.6	0.04
EISAI	Japan	Health Care	151	95.4	0.05
DAIICHI SANKYO	Japan	Health Care	430	94.7	0.14
SHIONOGI & CO.	Japan	Health Care	68	92.6	0.02
KIRIN	Japan	Consumer Goods	251	92.0	0.08
SEMICONDUCTOR ENERGY LABORATORY	Japan	ICT	1019	91.4	0.32
KANEKA	Japan	Basic Materials	391	91.3	0.12
NOVOZYMES	Denmark	Health Care	146	98.6	0.05
NEUROSEARCH	Denmark	Health Care	75	97.3	0.02
ASTRAZENECA	United Kingdom	Health Care	348	96.3	0.11
PHILIPS ELECTRONICS	Netherlands	Consumer Goods	6163	95.7	1.96
NOVO NORDISK	Denmark	Health Care	229	94.3	0.07
SCHINDLER	Switzerland	Industrials	192	93.8	0.06
ROYAL DUTCH SHELL	UK	Oil & Gas	344	89.5	0.11
SYNGENTA	Switzerland	Basic Materials	86	86.0	0.03
H. LUNDBECK	Denmark	Health Care	55	85.5	0.02
NOVARTIS	Switzerland	Health Care	166	80.7	0.05
SCHERING PLOUGH	United States	Health Care	59	78.0	0.02
MERCK & CO	United States	Health Care	493	47.3	0.16
ROCKWOOD	United States	Basic Materials	80	42.5	0.03
PFIZER	United States	Health Care	71	35.2	0.02
DOW CHEMICAL COMPANY	United States	Basic Materials	274	21.2	0.09
ABBOTT LABORATORIES	United States	Industrials	139	20.9	0.04
ORACLE	United States	ICT	76	19.7	0.02
ALTRIA GROUP	United States	Consumer Goods	58	19.0	0.02
JOHNSON & JOHNSON	United States	Health Care	851	16.9	0.27
SEALED AIR	United States	Industrials	84	14.3	0.03