

Patent families: when do different definitions really matter?

Catalina Martínez¹

Institute of Public Goods and Policies (IPP)

Consejo Superior de Investigaciones Científicas (CSIC), Madrid (Spain)

*This is a pre-copy-editing, author-produced PDF of an article accepted for publication in **Scientometrics** following peer review. The definitive publisher-authenticated version [Patent families: when do different definitions really matter? Scientometrics, 2011, vol. 86, issue 1, pages 39-63] is available online at: DOI: 10.1007/s11192-010-0251-3*

Abstract

Data on patent families is used in economic and statistical studies for many purposes, including the analysis of patenting strategies of applicants, the monitoring of the globalization of inventions and the comparison of the inventive performance and stock of technological knowledge of different countries. Most of these studies take family data as given, as a sort of black box, without going into the details of their underlying methodologies and patent linkages. However, different definitions of patent families may lead to different results. One of the purposes of this paper is to compare the most commonly used definitions of patent families and identify factors causing differences in family outcomes. Another objective is to shed light into the internal structure of patent families and see how it affects patent family outcomes based on different definitions. An automated characterization of the internal structures of all extended families with earliest priorities in the 1990s, as recorded in PATSTAT, found that family counts are not affected by the choice of patent family definitions in 75% of families. However, different definitions may really matter for the 25% of families with complex structures and lead to different family compositions, which might have an impact, for instance, on econometric studies using family size as a proxy of patent value.

Keywords: patent families; patent equivalents; priorities; patent value.

JEL classification: O31, O33, O34.

¹ Address for correspondence:

Catalina Martínez

CSIC Institute of Public Goods and Policies (CSIC-IPP)

Consejo Superior de Investigaciones Científicas (CSIC). Albasanz, 26-28 28037 Madrid (Spain)

Email: catalina.martinez@cchs.csic.es

1. Introduction

Given the territorial character of patent protection, when applicants want to protect their inventions in different countries, a patent application needs to be filed in each one of the patent offices where protection is sought. As a result, the first patent filing made to protect the invention (the priority filing) is followed by a series of subsequent filings and together they form a patent family.² Along these lines, the recently published OECD Patent Statistics Manual defines patent families as “the set of patents (or applications) filed in several countries which are related to each other by one or several common priority filings” (OECD, 2009).

However, as simple as it may appear, in practice it is not straightforward to identify all the members of a patent family, neither is it to assess whether they all protect the same invention. Not only are some families purely domestic, but also many of the publicly available statistics on patent families include a large share of “singletons” (*i.e.* single patent applications that form patent families on their own because they are not related to any other application). Moreover, the use of different patent family definitions and methodologies may lead to different patent family counts and different measures of patent family size. The first purpose of this paper is to provide insight into the latter by comparing the most commonly used definitions of patent families and by identifying the factors leading to differences in outcomes. The second purpose is to characterize the internal structure of patent families (the way in which patents come together to make up a family) in order to assess whether or not different definitions provide the same outcomes for the most common family structures.

It is now widely recognized that data on patent families can be used for many purposes. However, the first economic studies pointing at the advantages of using patent family data were published only three decades ago³ and patent family data had started to be compiled in statistical databases not long before.⁴ The increasing availability of data on patent families has gone hand in hand with the growing interest in patent families amongst researchers, statisticians and policy

² Subsequent filings have received multiple names in patenting studies, including external patents, external equivalents, equivalents, duplicated patents, multiple applications, secondary filings or patent family members. In this paper we use the term patent family member to refer to any patent included in a family, which comprises first filings (also called earliest priority filings) and subsequent filings. The term patent equivalent is used to characterise a specific type of family, as described in Section 3.

³ The pioneer in economic studies using family data was probably the German economist Konrad Faust (Faust and Schedl, 1982), although Grupp (1998) also cites studies on foreign patenting from the early 1980s by Russian authors Wassilëv and Adjubej, published in Russian.

⁴ To our knowledge, the very first efforts to construct patent families date back to the 1940s and were undertaken by Monty Hyams, the founder of private information provider Derwent. He started to publish data on patent families limited to the chemical sector and in the mid 1970s extended his analysis of patent families to all technologies and an increasing number of countries. The Derwent database was the only private source for international patent data for years. The Institut International des Brevets (IIB) in The Hague also began building patent families in the 1970s, before it became part of the European Patent Office in 1978.

makers. Several reasons may be put forward for the growing demand for patent family data, including a shift of focus from individual patents to patent portfolios in the management of intellectual property rights, the need for empirical evidence on patent strategies and patent value in economic studies. Multi-national filing strategies have also been largely facilitated by supranational procedures such as the European Patent Convention (EPC) and the Patent Cooperation Treaty (PCT), both set up at the end of the 1970s. Researchers wishing to build patent families based on their own definitions are also finding it easier to access raw data on patent linkages on a worldwide scale.⁵ However, most studies take family data as given, as a sort of black box, and do not go into the details of patent family building methodologies and underlying patent linkages. It is therefore the right moment to compare the different methodologies used to build the types of patent families most commonly used in the literature.

The paper is organized as follows. The next section presents some of the main economic and statistical uses of patent family data found in the literature. Section 3 presents the most frequently used definitions of patent families and discusses their differences. Section 4 provides a characterization of internal family structures for all extended patent families having earliest priorities in the 1990s and Section 5 concludes.

2. Why are patent families important? Economic and statistical uses

Patent family data are often used in economic and statistical studies. They may be used as the basis to build more refined patent indicators, to study different proxies for patent value or to investigate the motivations and strategies of patent applicants. The aim of this section is to briefly describe some of these uses.

Five uses of patent family data can be put forward: i) to prevent double counting; ii) to neutralize home advantage; iii) to forecast applications; iv) to analyze the internationalization of technology; and iv) to estimate patent value.

First, patent families prevent the double counting of single inventions when adding information from different patent offices, since applications filed in different countries to protect a given invention should be included in the same family (OECD, 2009).

⁵ Apart from commercial providers such as Derwent or Questel, the European Patent Office publishes information on patent families online on its website esp@cenet and produces since 2006 the Worldwide Patent Statistics Database PATSTAT (at the request of the Patent Statistics Task Force led by the OECD, which includes the EPO, USPTO, JPO, WIPO, NSF and the European Commission). Ever since it was first released, PATSTAT includes raw data on patent priority linkages and since 2008 it also produces ready-made tables on patent families. The OECD provides raw data on triadic patent families and publishes aggregate statistics online. The WIPO publishes aggregate family data annually since 2008.

Second, when the objective is to neutralize the home-country advantage of applicants, patent families are preferred to national patent statistics. Patent statistics based on counts from a single patent office would suffer from a home bias and overestimate the patent propensity of residents because applicants are more likely to file in their home country first and, eventually, later extend protection to other countries. Some studies have solved this problem by looking at patenting in third countries when comparing patent counts from different inventor countries, whilst others have done this by using patent families (Faust and Schedl, 1982; Grupp, 1998).

Third, patent family relations can be used to forecast patent application counts in order to plan future resource requirements at patent offices based on information on patent flows between countries with first filings and countries with subsequent filings. Hingley and Nicolas (1999, 2006) explore methods to provide simultaneous joint forecasts of patent applications at the three major patent offices, the US Patent and Trademark Office (USPTO), the Japan Patent Office (JPO) and the European Patent Office (EPO).

Fourth, patent family data are very well suited to analyze the internationalization of technology markets. Different indicators of internationalization and the geography of inventions based on patent family data can be found in the literature, most of which consist of imposing geographic filters to patent families. Debates on which filters are more suitable to capture different effects are found in the literature. Triadic families, the most widely used family indicator, require family members to be filed in all three major patent offices (USPTO, JPO and EPO) (Grupp et al., 1996; Grupp, 1998; Dernis, Guellec and van Pottelsberghe, 2001; Dernis and Khan and 2004; Guellec and van Pottelsberghe, 2004).⁶ More inclusive filters have also been used to identify high valued patents from small or developing countries not filed in the triad. Henderson and Cockburn (1993) regard patents as important if they have been applied for in only two of the three major economic regions, the United States, Japan and the European Community. Grupp (1998) proposed a similar definition based on filings made in only two triadic regions. More recently, Frietsch and Schmoch (2010) compared patent counts based on national office filings, EPO filings, PCT filings and triadic families. They recommend the use of “transnational patents”, defined as families using either EPO or PCT supranational filing procedures, to better capture recent globalization trends and expansion to emerging markets (e.g. China, India, South Korea). They conclude that in the future it would be enough to do country comparisons on the basis of PCT filings only – once the PCT system is well established in the patenting culture.⁷

⁶ In 2001, the OECD developed a methodology to produce the OECD triadic patent families defined as a set of patents taken at the EPO, the JPO and the USPTO to protect a same invention. Statistics of triadic patent families are published regularly at www.oecd.org/sti/ipr-statistics.

⁷ Other authors have proposed other alternatives. For example, Sternitzke (2009) proposes to change the definition of triadic families. His proposal consists of adding filings at the German and Chinese patent office to the triad in order to

Fifth, patent family data are also used for economic studies of patent value and applicant strategies. Four of these types of studies are described below.

First, patent family size, measured via the number of patent offices or number of patent applications belonging to a family, is frequently used as an indicator of patent value. Filing patents abroad is associated with higher costs for the applicant (in terms of patent office fees, patent attorneys bills and translation costs), which means that applicants would only extend protection to foreign countries if the time, effort and cost associated to it, is worth it. Therefore, the intuition goes, applicants would only seek international patent protection for their most valuable patents. The link between patent value and the size of patent families was first shown by Putnam (1996) in a cross-sectional econometric study using information on patent family size, which was an extension of the patent renewal model developed by Pakes and Schankerman (1984). Excluding inventions for which protection was only filed in the home country of the applicant, he estimated that the international component of annual capitalized patent returns of the 1974 patent cohort represented about 21% of annual private business R&D in the countries analyzed and that half the total value was captured by the top 5% of inventions, confirming a highly skewed distribution of patent value. He also found that the most valuable patent families were those including filings in major economies. This showed that not only does the sheer size of the family matter, but also the identity of the countries where protection is sought.⁸

Second, citations received by a patent in the search reports of other patents (forward citations) are also frequently used as an indicator of patent value⁹ and counts of forward citations can be improved by adding citations received by a patent to those received by its equivalents.¹⁰ A justification for this methodology is found in the EPO Guidelines for Examination (EPO, 2009, Part B, III-2), “The examiner should for reasons of economy exercise his judgment, based on his knowledge of the technology in question and of the available information retrieval systems, to omit sections of the documentation in which the likelihood of finding any documents relevant

take account of the importance of these two markets for specific technologies (reflected in the growing number of foreign filings registered in these offices), such as mechanical engineering for Germany, and telecommunications and mechanical engineering for China.

⁸ More recently, Deng (2007) examined the joint patent designation-renewal behaviour of EPO applicants finding that European patents granted through the EPO are substantially more valuable than those granted through the national route. She also found that the value distribution of patents is highly skewed (even more so for EPO patent families) and increases with the economic size of the country. In turn, van Pottelsberghe and van Zeebroeck (2008) propose a new indicator to measure the value of patents filed at the EPO (the scope-year index) that uses information about the countries where EPO grants are validated (scope) and on the number of renewals paid in each of those countries to maintain the patent alive (age). They stress the importance of considering both dimensions jointly given the dynamic character of patent families, and note that patent value measured at different points in time may provide completely different pictures.

⁹ The relation between forward citations and patent value was demonstrated by Harhoff et al. (1999), who showed that the higher the estimated economic value of US and German patents the more forward citations they received, by looking at USPTO patents citing USPTO patents and German patents citing German patents.

¹⁰ The EPO/OECD patent citations database was released in 2004 based on citations received by EPO patent applications and their PCT equivalents and was later extended to citations received by EPO equivalents in the national offices of EPC member states (Webb, Dernis, Harhoff and Hoisl, 2005).

to the search is negligible, for example documents falling within a period preceding the time when the area of technology in question began to develop. Similarly he need only consult one member of a patent family unless he has good reason to suppose that, in a particular case, there are relevant substantial differences in the content of different members of the same family.”¹¹

Third, litigation and procedural outcomes for patent equivalents filed in different jurisdictions can shed light on patent value and international patent application strategies. Graham and Harhoff (2006) find that EPO equivalents of litigated US patents are more likely to be granted and have higher opposition rates than the equivalents of unlitigated US patents. In their view, the owners of “EPO twins of US litigated patents” would be willing to compromise on the scope of their property rights in order to obtain a patent, which can be interpreted as a sign of the economic importance of the inventions they seek protection for.¹²

Finally, patent family linkages reflect patent applicant strategies and can thus be useful to investigate preferences of applicants and expected returns from patenting. Harhoff (2006) points out that some firms build patent portfolios by merging several priority filings or using divisional applications, a practice he called “patent constructionism”. Van Zeebroeck and van Pottelsberghe (2008) find that “constructionist” filing strategies at the European Patent Office, such as using the PCT route, filing divisionals or having more than one priority, are positively associated with patent value indicators.

3. Patent family definitions

The vast array of possible uses of patent family data is growing and being aware of the consequences of choosing one definition of a patent family over another is becoming all the more important. Narrow definitions would be preferred to broad definitions when comparing litigation outcomes in different jurisdictions, as this ensures that patents that protect exactly the same invention are being compared. In turn, more inclusive definitions of families may be preferred when exploring patent applicant filing strategies in order to ensure that all possible links between patent filings are included.

¹¹ This “economy of citations” at the EPO was already noted by Michel and Bettels (2001, p.189), “At the European Patent Office, examination experience has shown that most of the relevant information on the above criteria is obtained from 1–2 documents. According to EPO philosophy a good search report contains all the technically relevant information within a minimum number of citations. This is not necessarily the approach taken by other patent offices.”

¹² “The decision to pursue an application is the result of complex tradeoffs between patent scope and patent value. For an economically important invention, a patent holder may be willing to accept even a highly restricted patent (low scope) while for an economically unimportant one, the patent holder may simply withdraw the application. Applicants with valuable patents can be expected to put more effort into securing a patent grant (even if the claims are narrowed by the examiner during the give-and-take of the examination process).” (Graham and Harhoff 2006, p.13).

The problem is, as noted by Adams (2006, p.15), that “the definition of a family is not defined by law, but by each database producer for their own convenience”. In fact, each family definition may lead to a different aggregate count of patent families, but few comparative studies have been done so far to assess the causes of such differences. Most evidence available to date is based on compilations of examples. The aim here is to contribute to fill this gap by providing a comparative overview of some of the most popular definitions of patent families.¹³

3.1. Most widely used patent family definitions

This section presents four of the most widely used definitions of patent families. The first three correspond to families built solely using priority links recorded in patent databases.¹⁴ They are equivalents, extended families and first filing based families. The fourth definition refers to families whose composition is validated by experts. We will refer to this last type as expert-validated families based on novel technical content. All of them are briefly described below and their main features are summarized in Table 1.

**** Table 1 around here ****

First, *patent equivalents* are generally considered as the best way to identify patents that protect the same invention. They are defined as applications that have exactly the same priority or combination of priorities. EPO equivalents available in esp@cenet are an example of this.¹⁵

Second, *extended families* aim to capture all possible links between two given patent applications. However, they are sometimes criticized because the first filing may end up not protecting exactly the same invention as the last filing (and this can only be verified through expert control). Extended families are built by consolidating both direct and indirect priority links between patent applications within families. As a result it is possible to find two patent documents with no priority in common, but which are indirectly related because they both share at least one priority with a third application. The INPADOC extended priority patent families available in esp@cenet are an example of this type of family.¹⁶

13 One of the aims of the EPO/OECD workshop on patent families held in Vienna in October 2008 was to compare the outcomes of applying different family definitions to a sample of randomly-chosen patent applications. The experiment showed that differences, if they arise, tend to be related to cases where multiple priorities are claimed. Participants at the workshop distinguished between three types of families (extended; single-priority based; and equivalents) and some examples of filtered subsets (e.g. triadic).

¹⁴ These priority links can be of several types, including Paris Convention priorities, domestic priorities (continuations and divisionals), PCT relations and technical relations. The first three are recorded in patent documents. In contrast, technical relations (also called technical similarities) are additional links identified by patent examiners or database producers based on the similarity of inventions and other characteristics of patents (Martínez, 2010).

¹⁵ Data on EPO - Esp@cenet equivalents are available at www.espacenet.com. Inno-tec equivalents are another example available at www.inno-tec.bwl.uni-muenchen.de/personen/professoren/harhoff.

¹⁶ Data on INPADOC extended families are available at www.espacenet.com and PATSTAT.

Third, *single first filing based families (or single earliest priority families)* are families where each first filing forms a family together with the subsequent filings that claim it as priority. One important difference between this type of family and other types of families is that they are not mutually exclusive when, say, two first filings lead to a single subsequent filing. When this is the case, each subsequent filing will belong to two different families (each one defined by one of the two first filings claimed as priority) and would thus be counted twice in aggregate family counts. These types of families are used by the EPO to forecast future filings as well as to report statistics on first filings by country of priority and filing flows between different geographic blocs on the Trilateral website.¹⁷ This is also the methodology chosen by the World Intellectual Property Organisation (WIPO) to build the patent families it publishes annually in its patent statistics report.

Finally, *expert-validated families based on novel technical content* differ from the previous definitions in that expert control is used to validate the composition of patent families. They are based on the principle that a family must be formed by patent documents protecting the same technical content. Experts examine new applications and certify if their technical content matches that of existing families or not. Two examples of this last kind of family are DOCDB families and DWPI families. DOCDB families are primarily constructed by and for EPO patent examiners to optimize their work. They include patent documents that share identical “priority pictures”, understood as priorities adding new technical content after having excluded redundant priorities via expert control.¹⁸ In turn, the DWPI - Derwent World Patent Index (Nanu, 2003) is constructed based on the “novelty principle”, so that for a new family member to be added to a given family, it needs to have matching technical content with existing members. This implies that DWPI families take account of three pieces of information: i) priority linkages among applications, to identify potential family members; ii) the timing in which information on linkages becomes available in databases, to eventually start a new family when a new application is considered as a “basic” contribution, even if a priority link would suggest it should be attached to an existing family; and iii) expert control to validate matching technical content and family building.

The focus in the remainder of the paper is on the first three types of family definitions, that is, those which do not need expert validation and can thus enable researchers to replicate patent family data produced by other researchers by applying similar family building algorithms.

¹⁷ Data on single first filing based families are available at www.trilateral.net/tsr (based on the EPO - PRI system) and at www.wipo.int/ipstats/en/statistics/patents (based on the WIPO definition).

¹⁸ Based on the presentation by Fenny Versloot-Spoelstra from the EPO at the EPO/OECD workshop on patent families held in Vienna in November 2008. A more complete description is given in Martínez (2010).

3.2. Differing factors in non-expert validated family definitions

Equivalents, extended families and single first filing based families do not use expert control to validate the composition of families. They have advantages and disadvantages. On the negative side, they only rely on patent linkages reported in patent databases, which can be a problem when patent databases have incomplete information on patent linkages from certain offices.¹⁹ On the positive side, when methodologies are clear enough, other researchers can easily apply similar algorithms to reproduce and update the family data produced by other researchers. The aim of this section is to describe two main factors which may lead to differences in family counts and family size as a result of using equivalents, extended families or single first filing based families. These factors are: i) whether to include subsequent filings claiming multiple first filings as part of one or more families; and ii) whether or not to consolidate indirectly linked patent applications into single families.²⁰

As regards the first factor, Adams (2006, p.15) has already stated that the main source of family variations is a result of multiple priorities: “For many families there will be similar results; the main cause of variation is where one or more cases in the family claims multiple priorities. Some rules will put these members into a second distinct family, whilst others will group all cases into a single largest family”. Single first filing based families are an example of families that include applications in more than one family when they claim multiple earliest priorities.²¹ By contrast, in the case of extended patent families, subsequent filings claiming multiple priorities belong to the same family. Extended families are thus mutually exclusive: each patent document belongs to only one family.

Taking priority year 1999, a comparison between the number of INPADOC extended families and single first filing based families using the EPO-PRI and WIPO methodologies shows the impact of building a different family from each single earliest priority or not (Figure 1). Both EPO-PRI and WIPO families follow a similar approach: whenever two first filings lead to a single subsequent filing they consider that two families are formed.²² In contrast, the INPADOC

¹⁹ Additionally, Simmons (2009) cites typographical errors and changes in database standardization criteria as possible causes of misrepresentation of patent families and differences in family counts across different sources.

²⁰ Choosing different types of patent linkages (e.g. Paris Convention priorities, continuations, technical relations, PCT links) and patent documents (published v. unpublished) will also affect final patent counts of families based on similar definitions. For instance, INPADOC extended patent families include both published patent documents and unpublished priorities claimed in them as family members, whereas DOCDB patent families only include published patent documents as members. More information about the impact of these choices and a description of an algorithm to build extended patent families based on patent linkages included in PATSTAT can be found in Martínez (2010).

²¹ Statistics based on this kind of definition are regularly released by WIPO (WIPO, 2008) and EPO trilateral cooperation statistics (Hingley and Park, 2003; Hingley, 2009; www.trilateral.net)

²² Based on the presentations made by Peter Hingley from the EPO and Hao Zhou from the WIPO at the EPO/OECD workshop on patent families held in Vienna in November 2008, there might be other differences in their methodologies causing differences in patent counts, such as the kind of patent linkages used in EPO-PRI and WIPO single first filing based families, or the patent databases on which their algorithms are implemented. Exploring these possible additional differences is beyond the scope of this paper.

extended family concept would consider that only one family is formed. As shown in Figure 1, single first filing based families (EPO-PRI and WIPO) provide higher counts than extended families (INPADOC) with US, Japanese and German priorities, which is consistent with the fact that multiple priorities are more frequent in these three jurisdictions than in other countries.²³ It is also worth noting that a large proportion of these INPADOC extended families are singletons (36% for the US and Germany, 38% for France and as much as 85% for Japan and 56% for the UK).²⁴ The counts of EPO-PRI and WIPO families displayed in Figure 1 are based on published aggregate figures so we cannot provide information on how many singletons they include.

****Figure 1 about here ****

The second factor relates to whether or not indirect links between two given patents are consolidated to form broader families. In other words, it relates to the boundaries of patent families and where one patent family finishes and another family starts. Patent equivalents are an example of narrow families that only rely on direct links, whereas extended patent families are an example of broad families that consolidate both direct and indirect links. It might be easier to understand this with the help of an example. Figure 2 below presents the structure of a single extended family with six members, two earliest priorities [P1, P2] and four subsequent filings [D1, D2, D3, D4], where only D2 and D3 share exactly the same priorities [P1, P2] and are thus the only two equivalent subsequent filings within the broader family. Since equivalents are subsets of extended families, there will generally be fewer extended families than sets of patent equivalents.²⁵

**** Figure 2 about here ****

Based on the differences identified, an important question would be to know which definitions of families are best suited for what uses. First, equivalents are widely seen as the preferred definition when trying to identify the legal duplicates of an application in other offices. Nevertheless, further research on how to control statistically whether they are actually related to “the same invention” or not would be useful. Second, extended families would be the most useful definition for the analysis of patent applicant strategies, time lapse between first priority

²³ Out of the 165 763 INPADOC extended families (excluding singletons) with more than one earliest priority in 1991-1999, 41% have their origin in Japan, 27% in the United States and 7% in German. The remaining countries have lower shares.

²⁴ The shares of singletons in INPADOC extended families for Japan and the UK are high possibly due to lack of relations among applications or to lack of information about relations in the database. Indeed PATSTAT does not have complete information on divisional applications filed at the JPO (Martínez, 2010), but, to our knowledge, it includes all patent information from filings at the UKPO.

²⁵ As regards single first filing based families, it is clear that no consolidation is done at the level of the first filing, since each first filing is considered the “root” of a different family. Whether there is some degree of consolidation of indirect links between subsequent filings would depend on the rules chosen by each data provider.

and latest filing date for the protection of single or close inventions, the span of technology fields included in a family, or even possible changes of assignees across countries and over time within families. In practical terms, the extended families definition would also be the best suited to create new indicators by imposing filters ex-post, given that PATSTAT provides ready-made tables of INPADOC extended families. Finally, single first filing based families are a good tool to forecast filing flows within and across patent offices, but for applications in other areas it should be noted that, in face of multiple first filings, they can lead to double counting.

Another question would be to assess when differences across family definitions really matter, that is, how frequently equivalents, extended families and single first filing based families lead to different patent family outcomes. The next section aims to answer this question.

4. Internal structures of patent families

Most studies of patent families treat patent families as “closed sets” of patents, without questioning how patent linkages bring family members together. The aim of this section is to depart from such practice and look at the different “internal structures” of all extended INPADOC patent families filed in the 1990s. Family structures reflect the way family members are connected through patent links reported in databases.

4.1. Description of the algorithm to characterize internal family structures

Family structures are identified by applying a specifically developed algorithm that positions each family member in a “family tree” by counting the number of priorities, the number of different levels or generations within the family, and the number of family members within each generation.²⁶ The process is described more precisely below.

Let F be a family, defined as a set of n patent applications $F = \{a_1, a_2, \dots, a_n\}$.

For each a_i belonging to F :

- Let $P(a_i)$ be the number of parents of a_i (number of priorities claimed in a_i), where $P(a_i)$ can take values 0 to $n-1$.
- Let $C(a_i)$ be the number of children of a_i (number of applications claiming a_i as priority), where $C(a_i)$ can take values 0 to $n-1$.
- Let $L(a_i)$ be the level of a_i within the family, where level 1 corresponds to the root node (earliest priority of the family), so that if a_i is at level 1, $P(a_i)=0$.

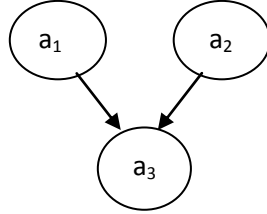
The internal structure of family F defined by priority relations can thus be represented by a matrix summarizing the position of all the applications belonging to the family:

$$S = [P \ C \ L \ N(P, C, L)]$$

²⁶ I am very grateful to Stéphane Maraut for having developed this algorithm.

where $N(P, C, L)$ is the number of a_i having the same values of P , C and L .

Consider the following example of a family F formed by three applications $F = \{a_1, a_2, a_3\}$:



Where the position of each application within it can be characterized as follows:

- For a_1 : $P(a_1)=0$; $C(a_1)=1$; $L(a_1)=1$
- For a_2 : $P(a_2)=0$; $C(a_2)=1$; $L(a_2)=1$
- For a_3 : $P(a_3)=2$; $C(a_3)=0$; $L(a_3)=2$

So that the internal structure of family F can be represented by the following matrix:

$$S = \begin{vmatrix} P & C & L & N(P,C,L) \end{vmatrix} = \begin{vmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 2 & 1 \end{vmatrix}$$

4.2. Most common internal structures of the families filed in the 1990s

Applying the algorithm just described to the 1 311 745 INPADOC extended patent families with earliest priorities in the 1990s, as reported in PATSTAT September 2008, and excluding families formed by one patent document only (singletons), a total of 47 437 different family structures are identified.²⁷ Among them, however, only a few structures are really popular: 86% of the structures represent just one family each, whereas only 10 structures represent 73% of all families and 25 represent 81%. In addition, more than half of the top 25 structures are made up of one earliest priority followed by several direct subsequent filings, what we will call “simple structures” from now onwards. The top structure (Structure ID 1), which characterizes 30% of

²⁷ Counts of INPADOC extended families excluding singletons reported here may differ when other versions of PATSTAT are used, given that new family members are added to existing families as they become published and new patent linkages added to the database. This means that new families can be formed, singletons may become families with multiple members and already existing families can be extended. As noted in the data catalog of PATSTAT September 2008, INPADOC extended priority families are formed by all the applications that “share a priority directly or indirectly via a third application. A ‘priority’ in this case means a link shown between applications as in tables TLS204_ APPLN_PRIOR (PARIS convention priorities), TLS205_TECH_REL (patents which have been technically linked by patent examiners on the basis of similar content) and table TLS216_ APPLN_CONTN (continuations, divisions etc).” Data on the INPADOC families themselves are available in table TLS219_INPADOC_FAM. The analysis in this paper is limited to patent families with earliest priorities after 1990 and before 2000. This ensured that data on patent linkages from PATSTAT were as complete as possible and that sufficient time was allowed for patent families to form. As regards the first issue, the following warning is included in the Data Catalog of PATSTAT September 2008: “Note that before 1991, the EPO did not record the so called ‘linkage type’ of priority numbers, that is the EPO did not record which kind of relation a given priority number has (Paris Union priority, continuation, division, etc.). Data in this element prior to 1991 is thus not reliable.”

all families consists of a single priority followed by only one subsequent filing. The second most common structure (Structure ID 2), which characterizes 12% of all families, is formed by one priority leading to two subsequent filings. Table 2 below provides more information on the 25 most common family structures and Table A1 in the Annex presents a graphical representation of each one of them.

**** Table 2 about here ****

A series of cross-tabulations and Pearson Chi-Square tests indicate that the distribution of families by structure (top 10) is significantly different across technology fields (five broad WIPO technology areas)²⁸ and countries of priority (top 5), as well as across different types of families, such as domestic and triadic (Annex Table A2). Many studies have shown the existence of a high correlation between triadic families and the value of patents, which have made the OECD triadic patent families a common indicator for researchers and statisticians aiming to analyze the most valuable patents of a country (e.g. van Zeebroeck and van Pottelsberghe, 2008). At the other extreme are the domestic families, which are those that remain within a single jurisdiction.²⁹ By choosing one of these two indicators (calculated as ex-post filters on the extended families) researchers can gain a first exploratory insight into the value of the invention protected by the family and the strategies of the applicant.

Interestingly, patent families in chemistry tend to be characterized more often than in other fields by Structure IDs 5, 6, 7 and 9, which are simple families formed by one single earliest priority followed by several subsequent filings. These are also the structures amongst the top 10 with the highest ratio of non-domestic *versus* domestic families, and of triadic *versus* non-triadic families.

4.3. Simple versus complex family structures

Family structures can be classified as simple or complex. As mentioned above, we consider all structures with a single first filing followed by one or several direct subsequent filings as *simple*. The rest of structures are considered as *complex*.

The application of the algorithm that classifies families depending on their structure shows that there are 54 different simple structures, where differences come from the number of subsequent filings claiming the single priority that is the root of the family, and as many as 47 383 different

²⁸ The correspondence is based on one of the IPC classes of the earliest priority of the family or of the subsequent filing with the oldest filing date, if there is no information on IPC for the earliest priority. For more information on the WIPO-IPC correspondence see www.wipo.int

²⁹ Domestic families are quite frequent in the US and Japan. Non-domestic families tend to be relatively more frequent in Europe than in Japan or the US, reflecting the fact that European countries have a regional office, the EPO, so that the first foreign filing for most European applicants, after a priority filing in their home countries, is made at the EPO (Frietsch and Schmoch, 2010).

complex structures. It also indicates that 75% of all extended families filed in the 1990s can be characterized as simple. The finding that such a large proportion of families have simple structures has important consequences for the comparison of outcomes across different family-building methodologies. It amounts to saying that 75% of all extended families filed in the 1990s also satisfy the definitions of single first filing based families and equivalents.

The logic behind simple family structures with many family members is not difficult to understand and corresponds to the most extended view of what patent families are. Simple structures would mostly reflect a filing strategy consisting of applying for protection for a single invention in one country first, probably the home country of the applicant, followed by a series of equivalent international filings aiming at obtaining protection for that same invention in other countries. Indeed, excluding Structure ID 1 (one parent, one child), 98% of the remaining simple families are non-domestic (see Annex Tables A2 and A3).

What may not be so easy to grasp are the patent procedures and applicant strategies leading to complex family structures. Let's take, for example, the complex structure that represents the highest percentage of families filed in the 1990s (1.9%) of all complex structures, characterized by two earliest priorities followed by a single first filing (Structure ID 8). Interestingly, 49% of the families filed in the 1990s with this structure have Japan as the priority country (of which 26% are domestic) and 21% have the United States (of which 92% are domestic).³⁰ This comparison already shows how a given complex structure can reflect quite different strategies; whereas in the United States this kind of structure would likely indicate the consolidation of USPTO provisionals, continuations or divisionals into a single purely domestic family, in Japan it might respond to the international extensions of the traditionally narrower JPO patents to other jurisdictions.³¹

A series of cross tabulations and Pearson Chi-Square tests indicate that the distribution of families by simple and complex structures are significantly different across technology fields and countries of priority, as well as across triadic and domestic families (see Table A3 in the Annex). They also show that whereas the shares of simple and complex families are roughly around 75% and 25% for all families³², the distribution changes when looking at the breakdown

³⁰ The priority date and country of a family with multiple first filings, as the one represented by Structure ID 8, are those of the first filing having the earliest filing date.

³¹ It is not uncommon to find patent applications at the EPO filed by Japanese applications that claim more than one Japanese priority. The difference in scope between applications filed at the JPO and in other offices is mainly due to a historical restriction on the number of claims allowed per patent that was annulled at the end of the 1980s (until 1988, the JPO allowed only one claim per patent).

³² The share of simple and complex families is different for the samples of families matched to technology fields and top priority countries in Annex Table A2, at roughly 74% and 26%, respectively. As regards technology fields, this is because the WIPO-IPC correspondence leaves out some IPC classes (which would mean leaving out of the analysis

by specific technology fields, countries of priority, or for families having domestic or triadic character. In particular, the share of families with complex structures exceeds the average of 25% in chemistry (31.2%); in families with priority in Japan (31.3%) and the United States (27.3%); in non-domestic families (27.6%); and in triadic families (37.6%). The highest shares of simple families are found in mechanical engineering (77.9%); in families with German (83.1%) and French priority (85.3%); in purely domestic families (84.5%); and in non-triadic ones (79%).

The above average share of complex structures in triadic families would be consistent with recent research by van Zeebroeck and van Pottelsberghe (2008), which points at a high correlation between patent value indicators (including triadic families) and the complexity of filing and drafting strategies of applicants at the European level. In turn, since the country of priority tends to be that of the applicant – as most first filings are domestic (OECD, 2009) – the above average share of complex families with Japanese and US priority indicates that Japanese and US applicants are more likely to file complex patent families than other applicants.

But what if the complexity of patent families is determined not only by applicant strategies but also by the specific characteristics of certain patent systems? If some countries have legal and procedural frameworks which are more likely to create additional patent linkages, their mere presence in the country composition of a family would increase the probability of their structure being complex. For instance, provisional and continuation filings are only found in the US. According to this view, patent families with at least one patent application made at the USPTO would be more likely to have a complex structure than patent families without US members. Indeed, whereas 25% of the whole population of families is complex, the share goes up to 31% when only families with US members are considered and down to only 13% when families with US members are excluded. The difference is also marked, but smaller, for Japan, where patents are traditionally narrower (fewer claims) than in other patent systems, which leads applicants to claim multiple Japanese priorities when extending protection to other countries: 30% of the families with Japanese members are complex, compared to 20% of the families with no Japanese members. Table 3 below presents this kind of comparison for the three largest patenting countries (US, Japan and Germany) and the European Patent Office, which are the four offices that attract the highest number of patents in the patent families with earliest priorities between 1991 and 1999, excluding singletons. Having a US member has by far the largest impact.

**** Table 3 about here ***

the patent families having those IPC classes), and in the case of top priority countries, it is because families with priorities in offices other than the top 5 are left out of the sample used for the analysis.

Further research should explore the impact of the characteristics of individual national patent systems on patent family structures, as well as the relation between complex family structures, patent value and applicant strategies. It should not be forgotten, however, that complex families are a very heterogeneous set of families, defined as complex solely on the basis of an internal structure which may take 47 383 different forms. Complex structures can be characterized by multiple priorities, multiple levels of subsequent filings, multiple links between applications or any other possible deviation from the simple structure of one single parent plus several children. Future research should also attempt to examine whether different complex structures can be grouped together based on what they have in common.

5. Conclusions

A patent family is generally understood as the set of patents filed in different countries with the objective to protect the same invention. However, some families are purely domestic and, in some cases, different family members might not seek protection for exactly the same inventions. Research “with” patent families seems to have a long future for an increasing range of economic and statistical studies, but, as argued here, it will benefit from research “on” patent families. A better understanding of the architecture of patent families would certainly be useful for the analysis of patent value and applicant strategies.

With this in mind, we have tried to shed light into the different methodologies underlying the most commonly used definitions of families. In addition, by using an algorithm that identifies and classifies internal family structures based on patent linkages using data from PATSTAT, we have found that 75% of all INPADOC extended priority patent families with earliest priorities between 1991 and 1999 have a simple structure, consisting of one single first filing and its direct subsequent filings. For them, different family definitions such as patent equivalents, extended families and single first filing based families will provide the same results. However, for the 25% of patent families that do not have a simple structure, different definitions may really matter and lead to different family compositions.

Based on this, some recommendations on the use of non-expert validated family definitions have been proposed. Equivalents are seen as the preferred option to identify legal duplicates of patent applications made in other offices, extended families are considered to be best suited for the analysis of broad patent applicant strategies and single first filing based families to forecast filing flows across patent offices. These recommendations would be relevant for families with complex structures, where the definitions of equivalents and single-first filings may lead to narrower family outcomes than extended families.

Moreover, although results should be treated with caution as they can only be as good as the raw data on which our algorithm is applied, we find that the distribution of families with different structures is not statistically independent from other patent dimensions. Complex families may favor specific technologies, countries or more valuable patents. In particular, complex structures are more likely to be found in the chemical sector and in families with Japanese priority. They are also more common in triadic than in non-triadic families and in families with US members than in families without US members. Different family definitions might thus have more impact for some family compositions than for others. For instance, the definition of equivalents may be more restrictive and result in smaller families than the definition of extended families for families with Japanese priorities, with a US member or for triadic families. This might be crucial in some contexts, such as in econometric studies using family size as a proxy of patent value. Researchers should be aware of these issues in order to choose the family definition that better serves their specific research purposes. Future research should further explore the relation between family structures and other dimensions of patent families.

Acknowledgements:

I would like to thank Stéphane Maraut for his comments and for developing the algorithm to characterize and classify different internal patent family structures for INPADOC extended families using raw data from PATSTAT. Hélène Dernis, Dominique Guellec, Peter Hingley and Rainer Frietsch provided helpful suggestions and comments on the OECD/STI Working Paper 2010/2 “Insight into different types of patent families” which introduces some of the ideas presented here, together with a more complete analysis of different types of patent families which benefitted from the discussions held at the EPO/OECD patent families workshop held in Vienna on 20–21 November 2008. Support from the OECD and the Spanish Government through the Ministry of Science and Innovation (CSO2009-10845) is acknowledged.

REFERENCES

- Adams, S. R. (2006), *Information sources in patents*, 2nd edition, K.G. Saur, Munich.
- Cohen, W., Nelson, R. and J. Walsh (2000), "Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not)", *NBER Working Paper* 7552.
- Deng, Y. (2007), "Private value of European patents", *European Economic Review*, 51, pp.1785-1812.
- Dernis, H., Guellec, D. and B. van Pottelsberghe (2001), "Using patent counts for cross-country comparisons of technology output", *STI Review*, 27, OECD, pp.129-146.
- Dernis, H. and M. Khan (2004), "Triadic patent families methodology", *OECD STI Working Paper* 2004/2.
- EPO (2009), Guidelines for examination in the European Patent Office, April 2009, available at www.epo.org
- Faust, K. and H. Schedl (1982), "International patent data: their utilisation for the analysis of technological developments", Workshop on patent and innovation statistics. OECD, Paris.
- Frietsch, R. and U. Schmoch (2010), "Transnational patents and international markets", *Scientometrics*, 82(1), 185-200.
- Graham, S. and D. Harhoff (2006), "Can Post-Grant Reviews Improve Patent System Design? A Twin Study of European and US Patents," *CEPR Discussion Paper* No. 5680, CEPR London.
- Grupp, H., Münt, G. and U. Schmoch (1996), "Assessing different types of patent data for describing high-technology export performance", in *Innovation, Patents and Technological Strategies*, pp.271-287, OECD, Paris.
- Grupp, H. (1998), *Foundations of the Economics of Innovation. Theory, measurement and practice*. Edward Elgar Publishing Ltd. Cheltenham, UK.
- Grupp, H. and U. Schmoch (1999), "Patent Statistics in the Age of Globalisation: New Legal Procedures, New Analytical Methods, New Economic Interpretation", *Research Policy* 28 (1999) 377-396.
- Guellec, D. And B. Van Pottelsberghe de la Potterie (2004), "Measuring the globalisation of technology. An approach based on patent data", *CEB Working Paper* 04-13.
- Harhoff, D., Narin, F., Scherer, F.M. and K. Vopel (1999), "Citation frequency and the value of patented inventions", *The Review of Economics and Statistics*, 81, 3, pp. 511-515.
- Harhoff, D. (2006), "Patent Constructionism: Exploring the Microstructure of Patent Portfolios", Presentation Prepared for the EPO/OECD Conference on Patent Statistics for Policy Decision Making, Vienna, October 23-24, 2006. Available at http://academy.epo.org/schedule/2006/ac04/_pdf/monday/Harhoff.pdf
- Henderson, R. and I. Cockburn (1993), "Scale, scope and spillovers: the determinants of research productivity in ethical drug discovery", Working Paper 3629-93, MIT and NBER.
- Hingley, P. and M. Nicolas (1999), "Improvements to methods for forecasting patent applications using information on patent families", unpublished paper presented at the International Forecasting Symposium, Washington DC

- Hingley, P. and W. G. Park (2003), "Patent Family Data and Statistics at the European Patent Office", paper presented at the WIPO-OECD Workshop on Statistics in the Patent Field, Geneva, available at <http://www.wipo.int/ipstats/en/resources/studies.html>
- Hingley, P. and M. Nicolas (2006), *Forecasting innovations. Methods for predicting numbers of patent filings*. Hingley and Nicolas eds., Springer.
- Hingley, P. (2009), "Patent families defined as priority forming filings and their descendents", Working Paper available at <http://forums.epo.org/patstat> under "work in progress".
- Lanjouw, J.O., Pakes A. and J. Putnam (1998), "How to count patents and value intellectual property: the uses of patent renewal and application data", *The Journal of Industrial Economics*, Vol. 46, No. 4, pp.405-432.
- Martínez. C. (2010), "Insight into different types of patent families", *OECD Science and Technology Working Paper* 2010/2.
- Michel, J. and B. Bettels (2001), "Patent citation analysis. A closer look at the basic input data from patent search reports", *Scientometrics*, 51 (1), pp. 185-201.
- Nanu, D. (2003), "The Derwent Patent Family and its application in Patent Statistical Analysis", presented at the WIPO-OECD workshop on Statistics in the Patent Field, Geneva.
- OECD (2009), *Patent Statistics Manual*, OECD, Paris.
- Pakes, A. and M. Schankerman (1984), "The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources", in Z. Griliches (ed.), *R&D, patents and productivity*, NBER Conference Series, Chicago, The University of Chicago Press.
- Putnam, J. (1996), "The value of international patent rights", PhD thesis, Yale University.
- Simmons, E.S. (2009), "'Black sheep' in the patent family", *World Patent Information*, 31, pp. 11-18.
- Sternitzke, C. (2009), "Defining triadic patent families as a measure of technological strength", *Scientometrics*, 81 (1), 91-109.
- Van Zeebroeck, N. and B. van Pottelsberghe de la Potterie (2008), "Filing strategies and patent value", *CEB working paper 08-016 and CEPR discussion paper 6821*.
- Van Pottelsberghe de la Potterie, B. and N. van Zeebroeck (2008), "A brief history of space and time: the scope-year index as a patent value indicator based on families and renewals", *Scientometrics*, 75 (2), 319-338.
- Webb, C. Dernis, H., Harhoff, D. and K. Hoisl (2005), "Analysing European and international patent citations: a set of EPO patent database building blocks", *OECD STI Working Paper* 2005/9, OECD, Paris.
- WIPO (2008), *World Patent Report: A Statistical Review*, WIPO, Geneva.

ANNEX

Table A1. Graphical representation of top 25 structures of INPADOC extended families
Excluding singletons. Earliest priorities 1991-1999


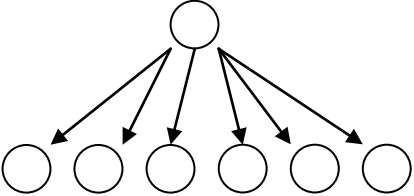
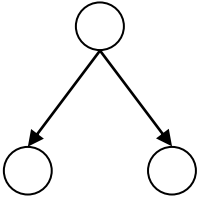
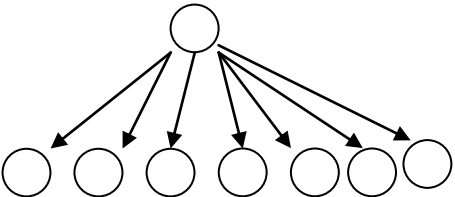
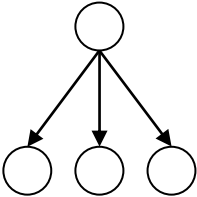
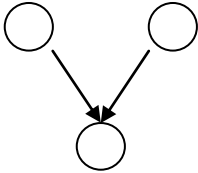
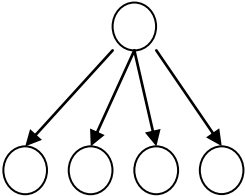
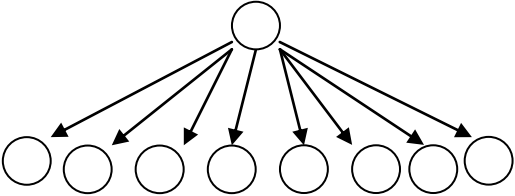
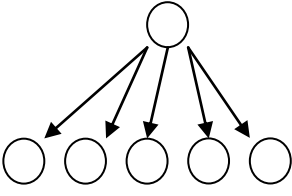
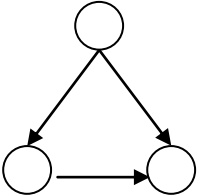
Structure ID	Family structure	Structure ID	Family structure
1		6	
2		7	
3		8	
4		9	
5		10	

Table A1 (cont.). Graphical representation of top 25 structures of INPADOC extended families

Excluding singletons. Earliest priorities 1991-1999

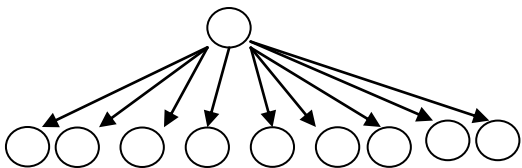
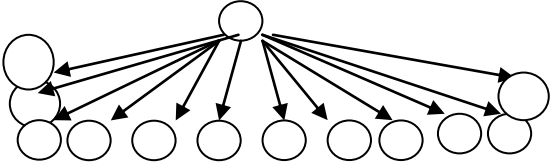
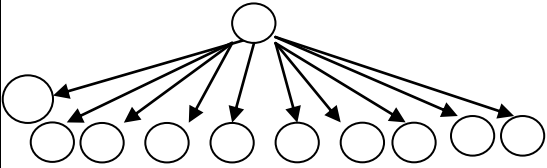
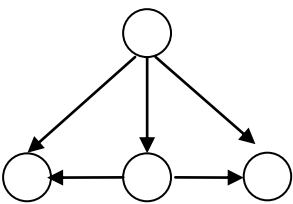
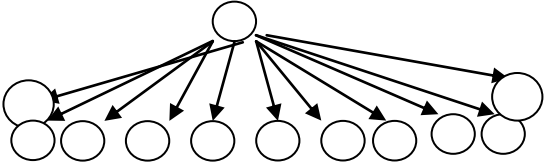
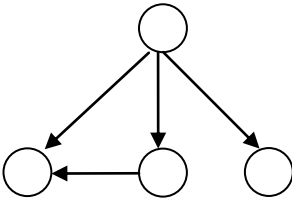
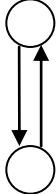
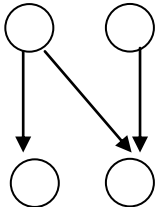
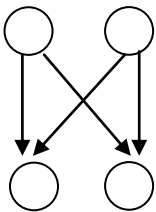
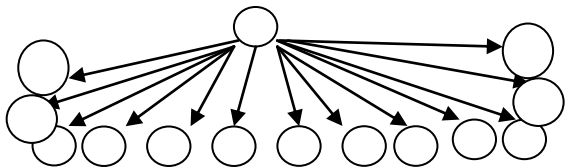
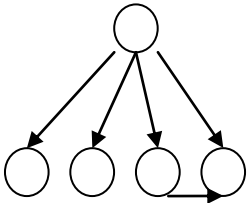
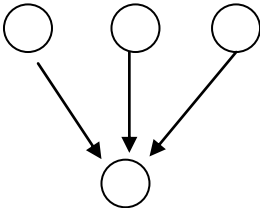
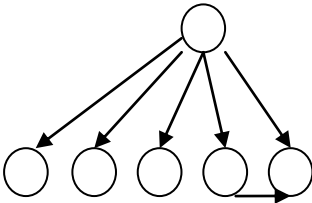
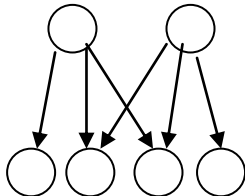
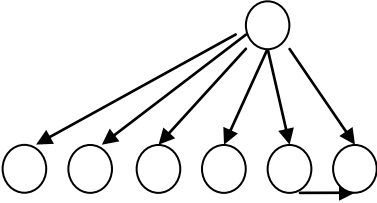
11		16	
12		17	
13		18	
14		19	
15		20	

Table A1 (cont.). Graphical representation of top 25 structures of INPADOC extended families

Excluding singletons. Earliest priorities 1991-1999

21		24	
22		25	
23			

Source: Own elaboration based on calculations using data from PATSTAT September 2008

Table A2. Top 10 internal family structures

Excluding singletons. INPADOC extended families with earliest priorities 1991-1999

Technology fields	top1	top2	top3	top4	top5	top6	top7	top8	top9	top10	Total
Electrical engineering	31813 40.1%	14578 18.4%	8879 11.2%	7615 9.6%	5261 6.6%	3184 4.0%	2001 2.5%	2309 2.9%	1227 1.5%	2541 3.2%	79408 100.0%
Instruments	55886 38.7%	24878 17.2%	15360 10.6%	14489 10.0%	10405 7.2%	7307 5.1%	4749 3.3%	4437 3.1%	3275 2.3%	3746 2.6%	144532 100.0%
Chemistry	66909 33.6%	32198 16.2%	21525 10.8%	19312 9.7%	17206 8.6%	13982 7.0%	11124 5.6%	4516 2.3%	8929 4.5%	3473 1.7%	199174 100.0%
Mechanical engineering	106224 39.3%	43256 16.0%	31308 11.6%	27929 10.3%	20626 7.6%	13825 5.1%	9187 3.4%	6465 2.4%	6152 2.3%	5048 1.9%	270020 100.0%
Other fields	48938 49.5%	14723 14.9%	9508 9.6%	6791 6.9%	4993 5.0%	3601 3.6%	2544 2.6%	3286 3.3%	1902 1.9%	2596 2.6%	98882 100.0%
Total	309770 39.1%	129633 16.4%	86580 10.9%	76136 9.6%	58491 7.4%	41899 5.3%	29605 3.7%	21013 2.7%	21485 2.7%	17404 2.2%	792016 100.0%

Pearson Chi-Square (36) = 16311.992

Asymp. Sig. (2-sided)= 0.000

Top 5 priority countries	top1	top2	top3	top4	top5	top6	top7	top8	top9	top10	Total
United States	101990 36.0%	53879 19.0%	30805 10.9%	26792 9.5%	21076 7.4%	15824 5.6%	11477 4.1%	5302 1.9%	8460 3.0%	7376 2.6%	282981 100.0%
Japan	129748 51.8%	35798 14.3%	18458 7.4%	20362 8.1%	12650 5.1%	6857 2.7%	3818 1.5%	12481 5.0%	2229 0.9%	7854 3.1%	250255 100.0%
Germany	52765 33.5%	23650 15.0%	23813 15.1%	18711 11.9%	14784 9.4%	10014 6.4%	6223 4.0%	1361 .9%	4079 2.6%	1956 1.2%	157356 100.0%
United Kingdom	15366 35.1%	7941 18.1%	4795 10.9%	3950 9.0%	3460 7.9%	2805 6.4%	2296 5.2%	760 1.7%	1740 4.0%	680 1.6%	43793 100.0%
France	9325 21.9%	5510 12.9%	5559 13.1%	6088 14.3%	5654 13.3%	4429 10.4%	3283 7.7%	137 0.3%	2456 5.8%	156 0.4%	42597 100.0%
Total	309194 39.8%	126778 16.3%	83430 10.7%	75903 9.8%	57624 7.4%	39929 5.1%	27097 3.5%	20041 2.6%	18964 2.4%	18022 2.3%	776982 100.0%

Pearson Chi-Square (36) = 58565.040

Asymp. Sig. (2-sided)= 0.000

Domestic	top1	top2	top3	top4	top5	top6	top7	top8	top9	top10	Total
Non-domestic	176816 24.6%	148893 20.7%	102730 14.3%	90696 12.6%	69328 9.6%	49159 6.8%	34278 4.8%	11990 1.7%	24476 3.4%	11486 1.6%	719852 100.0%
Domestic	215806 86.5%	8907 3.6%	900 0.4%	236 0.1%	87 0.0%	44 0.0%	21 0.0%	13538 5.4%	14 0.0%	10002 4.0%	249555 100.0%
Total	392622 40.5%	157800 16.3%	103630 10.7%	90932 9.4%	69415 7.2%	49203 5.1%	34299 3.5%	25528 2.6%	24490 2.5%	21488 2.2%	969407 100.0%

Pearson Chi-Square (9) = 349036.522

Asymp. Sig. (2-sided)= 0.000

Triadic	top1	top2	top3	top4	top5	top6	top7	top8	top9	top10	Total
Non-triadic	392622 48.3%	145983 18.0%	87969 10.8%	56056 6.9%	35937 4.4%	22631 2.8%	14613 1.8%	25527 3.1%	9516 1.2%	21479 2.6%	812333 100.0%
Triadic	0 0.0%	11817 7.5%	15661 10.0%	34876 22.2%	33478 21.3%	26572 16.9%	19686 12.5%	1 0.0%	14974 9.5%	9 0.0%	157074 100.0%
Total	392622 40.5%	157800 16.3%	103630 10.7%	90932 9.4%	69415 7.2%	49203 5.1%	34299 3.5%	25528 2.6%	24490 2.5%	21488 2.2%	969407 100.0%

Pearson Chi-Square (9) = 310271.850

Asymp. Sig. (2-sided)= 0.000

Transnational	top1	top2	top3	top4	top5	top6	top7	top8	top9	top10	Total
Non-transnational	325289 68.0%	62486 13.1%	27104 5.7%	11993 2.5%	5435 1.1%	2312 0.5%	1086 0.2%	23487 4.9%	560 0.1%	18962 4.0%	478714 100.0%
Transnational	67333 13.7%	95314 19.4%	76526 15.6%	78939 16.1%	63980 13.0%	46891 9.6%	33213 6.8%	2041 0.4%	23930 4.9%	2526 0.5%	490693 100.0%
Total	392622 40.5%	157800 16.3%	103630 10.7%	90932 9.4%	69415 7.2%	49203 5.1%	34299 3.5%	25528 2.6%	24490 2.5%	21488 2.2%	969407 100.0%

Pearson Chi-Square (9) = 421830.829

Asymp. Sig. (2-sided)= 0.000

Table A3. Simple and complex family structures

Excluding singletons. INPADOC extended families with earliest priorities 1991-1999

Technology fields	Simple	Complex	Total
Electrical engineering	76955 74.6%	26206 25.4%	103161 100.0%
Instruments	144277 73.2%	52855 26.8%	197132 100.0%
Chemistry	224092 68.8%	101448 31.2%	325540 100.0%
Mechanical engineering	274245 77.9%	77716 22.1%	351961 100.0%
Other fields	98059 77.3%	28878 22.7%	126937 100.0%
Total	817628 74.0%	287103 26.0%	1104731 100.0%
Pearson Chi-Square (4) = 8105.879			Asymp. Sig. (2-sided) = 0.000
Priority countries	Simple	Complex	Total
United States	294979 72.7%	110568 27.3%	405547 100.0%
Japan	233575 68.7%	106350 31.3%	339925 100.0%
Germany	164383 83.1%	33477 16.9%	197860 100.0%
United Kingdom	48760 74.3%	16880 25.7%	65640 100.0%
France	49342 85.3%	8536 14.7%	57878 100.0%
Total	791039 74.1%	275811 25.9%	1066850 100.0%
Pearson Chi-Square (4) = 17617.495			Asymp. Sig. (2-sided) = 0.000
Domestic	Simple	Complex	Total
Non-domestic	766714 72.4%	292961 27.6%	1059675 100.0%
Domestic	226048 84.5%	41502 15.5%	267550 100.0%
Total	992762 74.8%	334463 25.2%	1327225 100.0%
Pearson Chi-Square – continuity correction (1) = 16685.978			Asymp. Sig. (2-sided) = 0.000
Triadic	Simple	Complex	Total
Non-triadic	785436 79.0%	209292 21.0%	994728 100.0%
Triadic	207326 62.4%	125171 37.6%	332497 100.0%
Total	992762 74.8%	334463 25.2%	1327225 100.0%
Pearson Chi-Square – continuity correction (1) = 36453.867			Asymp. Sig. (2-sided) = 0.000

TABLES AND FIGURES TO BE INSERTED IN THE MAIN TEXT

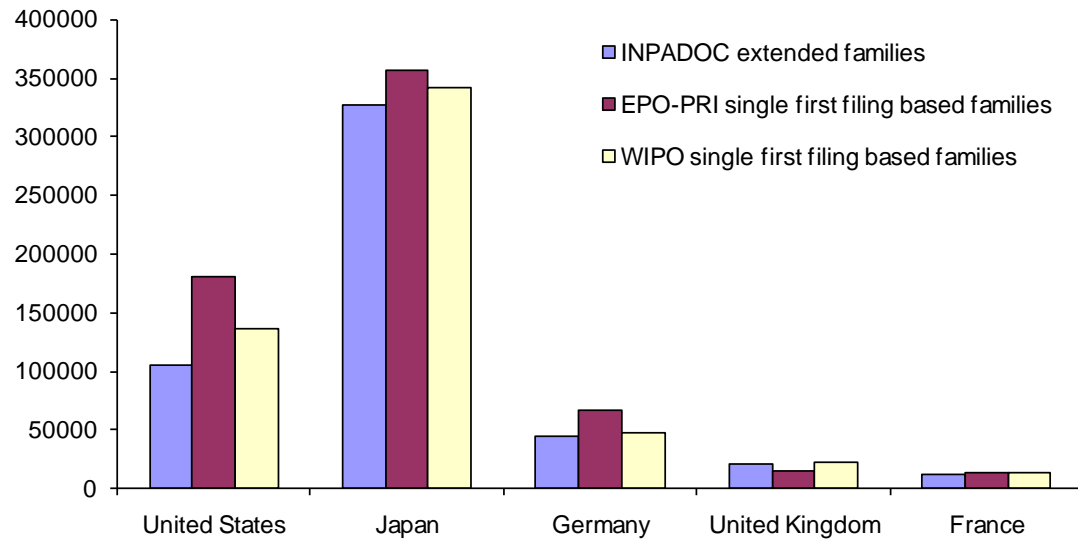
Table 1. Most widely used patent family definitions

Type	Definition	Interpretation	Uses	Expert validation	Multiple earliest priorities	Consolidation of indirect links	Examples
Equivalents	Applications having exactly the same priority or combination of priorities.	Patents protecting same inventions.	Analysis of citations received, procedural history and legal differences of patent documents protecting the same inventions in different jurisdictions.	NO	YES	NO	EPOesp@cenet equivalents Inno-tech equivalents
Extended families	Applications directly or indirectly linked through priorities.	Patents protecting same or related inventions.	Analysis of applicant strategies to extend patent protection over time and in different countries as well as cumulativeness of inventions and patent thickets. Basis for OECD triadic families which are calculated by imposing the restriction that extended families need to have USPTO, EPO and JPO applications to be triadic.	NO	YES	YES	INPADOC extended families
Single first-filing based families	Applications originating from a single first filing (earliest priority). Each first filing is considered as the root of a different family. One subsequent filing claiming priority in two different first filings will appear in two different families.	Patent activity following a priority forming first-filing.	Statistical analysis of patent filing flows between priority countries and offices of subsequent filings used in patent offices statistical reports and to forecast patent office workloads.	NO	NO	--	EPO-PRI families WIPO families
Expert-validated families based on novel technical content	Applications adding new technical content to the state of the art are considered as the root of new families. Subsequent filings with matching content are added to the family.	Patents protecting new technical content.	DOCDB simple families: Primarily constructed by and for patent examiners to optimize their work. DWPI: commercial patent database to provide information on patent strategies and internationalization of technology.	YES	YES	YES	DOCDB simple families Derwent World Patent Index (DWPI)

Note: Data on **EPO - Esp@cenet equivalents** are available at www.espacenet.com and on **Inno-tec equivalents** at www.inno-tec.bwl.uni-muenchen.de/personen/professoren/harhoff. Data on **INPADOC extended families** are available at www.espacenet.com and PATSTAT. Data on single first filing based families are available at www.trilateral.net/tsr (based on the **EPO - PRI system**) and at www.wipo.int/ipstats/en/statistics/patents (based on the **WIPO definition**). Data on families based on expert-validated families based on novel technical content are available in **DOCDB simple patent family** as reported in the PATSTAT database since 2008 and in the **Derwent World Patent Index (DWPI)** available at www.thomsonreuters.com/products_services/scientific/DWPI

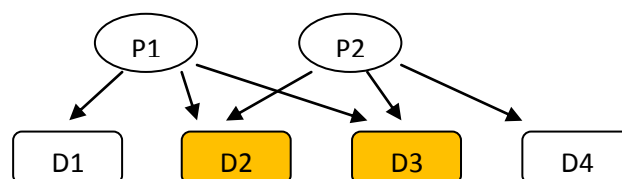
Source: Own elaboration, partly based on discussions held at the EPO/OECD workshop on patent families October 2008. More information in Martínez (2010).

Figure 1. Extended families versus single first filing based families
Top priority offices. Earliest priority year 1999. Including singletons.



Source: Counts of EPO-INPADOC extended families from PATSTAT data on EPO-PRI families correspond to updated figures from Table B1 in Hingley and Park (2003) provided by P. Hingley and data on WIPO families come from www.wipo.int/ipstats/en/statistics/patents patent families by country of origin (1990-2005) as of 3 June 2009.

Figure 2. Example of extended family and equivalents



Source: Own elaboration

Table 2. Top 25 family structures of INPADOC extended families
Excluding singletons. Earliest priorities 1991-1999

Structure ID	“Simple structure”	Number of earliest priorities	Number of applications in the family structure	% all families
1	yes	1	2	29.7
2	yes	1	3	11.9
3	yes	1	4	7.8
4	yes	1	5	6.9
5	yes	1	6	5.2
6	yes	1	7	3.7
7	yes	1	8	2.6
8	no	2	3	1.9
9	yes	1	9	1.8
10	no	1	3	1.6
11	yes	1	10	1.3
12	yes	1	11	1.0
13	yes	1	12	0.7
14	no	-	2	0.6
15	no	2	4	0.6
16	yes	1	13	0.5
17	no	1	4	0.5
18	no	1	4	0.5
19	no	2	4	0.5
20	yes	1	14	0.4
21	no	1	5	0.4
22	no	1	6	0.4
23	no	1	7	0.3
24	no	3	4	0.3
25	no	2	6	0.3
Top 10	-	-	-	73
Top 25	-	-	-	81

Notes: i) A “simple structure” is characterized by one first filing plus one or several direct subsequent filings; ii) Family Structure ID 14 is cyclical meaning that no application can be considered as the “earliest priority” because each one claims the other as priority.

Source: Own calculations based on PATSTAT September 2008.

Table 3. Simple and complex family structures in subsets of INPADOC extended families
Excluding singletons. Earliest priorities 1991-1999

	Total	Share of Simple	Share of Complex
With US member	886665	69%	31%
Without US members	425080	87%	13%
With Japanese member	685958	70%	30%
Without Japanese members	625787	80%	20%
With German member	552267	72%	28%
Without German members	759478	77%	23%
With EPO member	634513	71%	29%
Without EPO members	677232	79%	21%
All families	1311745	75%	25%

Source: Own calculations based on PATSTAT September 2008.