Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY

INSIGHT INTO DIFFERENT TYPES OF PATENT FAMILIES
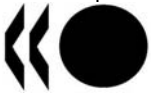
STI WORKING PAPER 2010/2
Statistical Analysis of Science, Technology and Industry

By Catalina Martinez
CSIC – Institute of Public Goods and Policies, Madrid

JT03278654

STI Working Paper Series

The Working Paper series of the OECD Directorate for Science, Technology and Industry is designed to make available to a wider readership selected studies prepared by staff in the Directorate or by outside consultants working on OECD projects. The papers included in the series cover a broad range of issues, of both a technical and policy-analytical nature, in the areas of work of the DSTI. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

Comments on the papers are invited, and should be sent to the Directorate for Science, Technology and Industry, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its member countries.

www.oecd.org/sti/working-papers

INSIGHT INTO DIFFERENT TYPES OF PATENT FAMILIES[1]

Catalina Martínez[2]

CSIC-IPP, *Institute of Public Goods and Policies-Consejo Superior de Investigaciones Científicas*, Madrid

Abstract

What are patent families? What is the impact of adopting one definition or another? Are some definitions of patent families better suited than others for certain uses in statistical and economic analysis? The aim of this paper is to provide some answers to these questions, compare the methodologies and outcomes of the most commonly used patent family definitions and provide guidance on how to build families based on raw data from the EPO Worldwide Patent Statistics database (PATSTAT). One of our findings, based on a characterisation of family structures, is that extended patent families and other family definitions, such as equivalents and single-priority families, provide identical outcomes for about 75% of the families with earliest priority dates in the 1990s because they have quite simple structures. Differences across definitions only become apparent for the families with more complex structures, which represent 25% of the families of that period.

[2] *Institute of Public Goods and Policies (*IPP), Consejo Superior de Investigaciones Científicas (CSIC). Centro de Ciencias Humanas y Sociales Albasanz, 26-28 28037 Madrid. Email: catalina.martinez@cchs.csic.es.

ÉCLAIRAGE SUR DIFFÉRENTS TYPES DE FAMILLES DE BREVETS[1]

Catalina Martínez[2]
CSIC-IPP, *Institute of Public Goods and Policies – Consejo Superior de Investigaciones Científicas*, Madrid

Résumé

Qu'est-ce qu'une famille de brevets ? Quelles conséquences l'adoption de telle ou telle définition peut-elle avoir ? Certaines définitions des familles de brevets sont-elles mieux adaptées que d'autres à certains usages en analyse statistique et économique ? Le présent document a pour objet d'apporter des réponses à ces questions, de comparer les méthodologies et les résultats des définitions de familles de brevets les plus courantes et de donner des indications sur la marche à suivre pour construire des familles de brevets à partir des données brutes de la base de données mondiale de l'OEB sur les brevets (PATSTAT). L'une de nos conclusions, fondée sur une caractérisation des structures des familles de brevets, est que des familles de brevets étendues et d'autres types de familles de brevets, tels que les équivalents et les familles de brevets partageant la même priorité, fournissent des résultats identiques pour 75 % environ des familles dont les premières dates de priorité se situent dans les années 90, car elles présentent des structures relativement simples. Les définitions ne commencent à diverger que pour les familles offrant des structures plus complexes, lesquelles représentent 25 % de l'ensemble pour cette période.

TABLE OF CONTENTS

1. Introduction

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). Given the territorial character of patent protection, when an applicant wants to protect an invention internationally, a patent application has to be filed in each of the countries where protection is sought (either one by one or collectively through supranational filing procedures). As a result, the first patent filing made to protect the invention, the so-called priority filing, which is usually made in the home country of the applicant, is followed by a series of subsequent filings and forms, together with them, a patent family.[3] In practice, however, it is not straightforward to identify all the members of a patent family, at a certain point in time, and differences across information sources arise from the use of diverse methodologies. Our purpose in this paper is to provide insight on the most commonly used definitions of patent families in order to better understand the implications of their differences and the advantages and disadvantages of using one over another in certain cases.

Economists, statisticians and policy makers have for a long time requested information on how single inventions are protected in different countries, but the cost and effort required to obtain it has always been a challenge. 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. Derwent's database was the only private source for international patent data for years. On the public side, the *Institut International des Brevets* (IIB) in The Hague began to build patent families in the 1970s, before it became part of the European Patent Office (EPO) in 1978, which continued to produce and release patent family data from then onwards. International organisations like OECD and WIPO also started to publish patent family data a few years ago, based on different definitions, but the breakthrough for research using patent family data occurred in 2006, when the EPO released for the first time the worldwide patent statistics database PATSTAT, at the request of the Patent Statistics Task Force led by the OECD, which includes EPO, USPTO, JPO, WIPO, NSF and the European Commission.

Patent data was at first mainly the domain of patent practitioners, and economists and statisticians started to pay attention to it at the end of the 1970s. The first economic studies pointing at the advantages of using patent family data instead of individual patent filings were probably published in the beginning of the 1980s.[4] Since then, the increasing availability of data on patent families has gone hand-in-hand with growing interest from researchers, statisticians and policy 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 IPR management and the need for empirical evidence on patent strategies and patent value in economic studies, as well as the fact that multi-national filing strategies have 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.

It is now widely recognised that patent families can be used for many purposes, such as to analyse patenting strategies of applicants and countries, monitor the globalisation of inventions and study the inventive performance and stock of technological knowledge of different countries. Moreover, raw data on

---

[3] 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.

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

patent linkages at worldwide scale is becoming more and more accessible to researchers (mainly thanks to PATSTAT) who may be increasingly willing to build their own patent families based on their own definitions. The number of studies using patent family data is probably going to grow substantially in the years to come. However, most studies to date take family data as given, as a sort of black box, without getting into the obscure details of patent family building methodologies and underlying patent linkages. It is therefore the right moment to recapitulate and document the history of patent family data and its uses, and investigate the differences and commonalities among the different types of patent families, which is the main objective of this study. Another objective is to characterise the internal structure of patent families in order to assess to what extent they are simple or complex and how different family definitions may affect family outcomes.

The paper is organised as follows. The next section, Section 2, presents briefly some of the most relevant economic and statistical uses and interpretations of patent family data. Section 3 presents the most commonly used definitions of patent families, indicating sources where information is available about them. Section 4 compares family counts based on different definitions. Section 5 is devoted to PATSTAT as a source of patent linkages (family relations) and provides guidance on how to build families from raw patent data using algorithms. In Section 6, we identify different internal family structures and classify families into those with simple and those with complex structures. In section 7 we propose a set of rules to identify patent equivalents within families with complex internal structures, and Section 8 concludes.

2.   Economic and statistical uses of patent families

Patent family data has been used in economic and statistical studies with many different objectives, such as to eliminate the home bias, to avoid double counting, to set an economic threshold in patent statistics, to estimate patent value, to monitor globalisation, to compare different patent systems, to analyse applicant filing strategies and to estimate workload at specific patent offices and filing flows across different patent systems. Some of these studies only require data at the macro level, aggregate counts of families by country, some others need family data at the micro level, as they need information on each member of the patent family. Table 1 below provides a brief description of the most common uses of patent family data and some of the first references to economic and statistical studies available in the literature that relate to each approach. This review does not aim to be exhaustive, it is presented for illustrative purposes. The literature in this field is growing rapidly and researchers are increasingly proposing new uses of patent family data.

Table 1. Different uses and interpretations of patent family data

Type of analysis	Objective	Use	Some references
Macro level (*e.g.* patent family counts by country of inventor)	Eliminate double counting in international comparisons of patent statistics	Group each country´s total patenting into patent families so that only the priority patents are counted.	Faust and Schedl (1982) Grupp (1988)
	Set an economic threshold in patent statistics	Exclude domestic applications with no foreign extension, which are supposedly of lower value than those in international patent families. Select the highest value patents among those with the highest number of foreign equivalents (family size), with international extensions in specific countries (*e.g.* triadic family) or going through supranational procedures (*e.g.* transnational patents, foreign-oriented patent family).	Faust and Schedl (1982) Grupp (1988) Henderson and Cockburn (1993) Grupp and Schmoch (1999) Dernis, Guellec and van Pottelsberghe (2001) Hingley and Park (2003) Dernis and Khan (2004) Guellec and van Pottelsberghe, (2004) WIPO (2008) Frietsch and Schmoch (2010)
	Estimate filing flows across different patent offices	Forecasting workload at individual patent offices based on international filing flows; and nowcasting patent statistics (to improve timeliness) based on international filing flows.	Hingley and Nicolas (1999, 2006) Hingley and Park (2003) Dernis (2007)
Micro level (*e.g.* characteristics of individual patents within given patent families)	Estimate value of patent rights	Based on models of the decision to file for protection in a set of specific countries and incur related costs (expected patent returns v. patent costs).	Putnam (1996) Lanjouw, Pakes and Putnam (1998) Deng (2007) Van Pottelsberghe and van Zeebroeck (2008)
	Estimate patent value based on citations	Analysis of forward citations received by a patent document and its equivalents.	Harhoff, Narin, Scherer and Vopel (1999) Webb, Dernis and Harhoff (2005)
	Estimate patent value based on litigation	Analysis of litigation and opposition procedures in which a patent document and its equivalents are involved after grant in different jurisdictions.	Graham and Harhoff (2006)
	Analyse applicant patent strategies	Analysis of filing strategies and use of the patent system by applicants within specific countries and internationally to protect the same or related inventions.	Harhoff (2006) Van Zeebroeck and van Pottelsberghe (2008)

Patent statistics are often used in cross-country comparisons of inventive performance, but rough numbers of patent applications tend to suffer from a home bias produced by the fact that applicants are more likely to file first in their home country and, eventually, later extend protection to other countries (Faust and Schedl, 1982; Grupp, 1982). Patent families are useful to avoid double counting when adding patent indicators from different jurisdictions, to correct the home bias associated with patent statistics from a single patent office, and to build global patent indicators related to single inventions. A condition for that to happen is that any single patent should belong to one and only one family, that is, that patent families must be mutually exclusive.

Another important issue to address when using patent statistics is their highly skewed value distribution, with very few high value patents and a majority of low value ones (OECD, 2009). Patent family data have been used to set an economic threshold, with the aim to capture only the most valuable

ones. Since filing patent applications abroad is associated with higher costs for the applicant, in terms of patent office fees, patent attorneys bills and translation costs, the intuition goes that applicants would only follow that path if the time, effort and cost associated with it, is worth it. Applicants would only seek international patent protection for their most valuable patents, as they would only be willing to do it if the expected commercial value of their invention is high enough.

The link between patent value and the size of patent family was shown by Putnam (1996) in a cross-sectional econometric study using information on patent family size, as an extension of the patent renewal model developed by Pakes and Schankerman (1984). Excluding inventions for which protection was only filed in their home country, he estimated that the international component of annual capitalised patent returns of the 1974 patent cohort represented about 21% of annual private business R&D in the countries analysed 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 with filings in major economies. More recently, Deng (2007) has examined the joint patent designation-renewal behaviour of EPO applicants finding that the European patents granted through EPO are substantially more valuable than those granted through the national route. She also finds that the value distribution of patents is highly skewed (even more so for the EPO patent families) and increases with the economic size of the country. Van Pottelsberghe and van Zeebroeck (2008) propose a new indicator to measure the value of patents filed at EPO (the scope-year index) that also uses both information on 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.

Several geographical filters of patent families have been proposed to build statistics that exclude the lowest valued patents. The most widely used filter is that requiring family members to be filed in three of the major patent offices (USPTO, JPO and EPO), resulting in the so-called triadic patent families (Grupp *et al.*, 1996; Grupp, 1998; Dernis, Guellec and van Pottelsberghe, 2001; Dernis and Khan, 2004; Guellec and van Pottelsberghe, 2004).[5] More inclusive filters have also been explored to identify high valued patents from small or developing countries that would not be filed in the triad. Henderson and Cockburn (1993) regarded patents as important if they had been applied in only two of the three major economic regions, United States, Japan and the European Community. Grupp (1998) also proposed something similar based on filings in only two triadic regions. More recently, Frietsch and Schmoch (2010) have recommended the use of "transnational patents" defined as families using either EPO or PCT supranational filing procedures, with the aim to capture globalisation and expansion to emerging markets (*e.g.* China, India, Korea).

Other researchers have used individual data on citations received by patents to estimate their value or technological importance and analyse technological spillovers across regions and applicants. 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, however their study was confined to citations registered in one single office (USPTO patents citing USPTO patents and German patents citing German patents). Data on patent families have been used to identify patent documents protecting the same invention in different jurisdictions (patent equivalents), so that citations received could comprise not only those received by the original document, but also those

---

[5]   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. It publishes statistics of triadic patent families regularly at www.oecd.org/sti/ipr-statistics.

received by its equivalents.[6] The concept of equivalents has also been used by Graham and Harhoff (2006), who recommend the adoption of a post-grant review system in the United States based on estimations made from a comparison of US litigation and EPO opposition records of "twin USPTO and EPO granted patents". They 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.

The use of patent family linkages to illustrate patent applicant strategies and estimate patent value based on filing strategies is another recent line of research. Harhoff (2006) points out how 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.

Finally, it is also worth mentioning that patent family data is used by patent offices in statistical models to forecast patent application numbers for planning future resource requirements. Hingley and Nicolas (1999, 2006) explore methods to provide simultaneous joint forecasts of patent applications at the three major patent offices: USPTO, JPO and EPO. The OECD also uses family links to nowcast triadic patent family statistics, aiming to go around the timeliness issue of patent statistics (Dernis, 2007; OECD, 2009).

3.   Most widely used patent family definitions

Each family definition may lead to a different patent count, but family definitions abound and few comparative studies have been done to date. Most evidence available to date is based on compilations of examples).[7] Indeed, as noted by Adams (2006, p.15), "the definition of a family is not defined by law, but by each database producer for their own convenience". The aim of this section is to contribute to fill this gap by providing a comparative overview of some of the most popular definitions of patent families, as presented by different patent information providers (Table 2): *i)* equivalents; *ii)* extended families; *iii)* single-priority based families; *iv)* examiners technology-based families; and *v)* commercial novelty-based families.[8] The first three definitions rely solely on linkages available in patent databases, whereas the last two use additional expert control (experts that actually read the patents to confirm they belong to the same family). The aim of this section is to describe their methodologies, point at places where data on them is provided and try to understand why and how they are different.

---

[6]   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).

[7]   One of the aims of the EPO/OECD workshop on patent families was to compare the outcomes of applying different family definitions to a sample of patent applications randomly chosen. The experiment showed that differences, if they arise, tend to be related to cases where multiple priorities are claimed (Dernis and Hingley, 2008; Paris, 2008; Fortune, 2008; Rollinson, 2008; Hingley, 2008; Torre, 2008; Dernis, 2008; Martinez and Maraut, 2008; Raffo and Lhuillery, 2008; Harhoff, 2008).

[8]   Other attempts to develop typologies of patent families have been made. Fortune (2008) included the following six types: *i)* INPADOC families; *ii)* esp@cenet families; *iii)* Equivalent families; *iv)* Triadic families; *v)* Domestic families; and *vi)* National families. Participants at the EPO/OECD workshop on patent families distinguished between three types of families (extended; single-priority based; and equivalents) and some examples of filtered subsets (*e.g.* triadic).

Table 2. Most widely used patent family definitions

Type	Interpretation	Uses	Definition	Expert quality control of patent linkages	Data availability
Equivalents	Patents that most likely protect SAME inventions.	Analysis of citations received, procedural history and legal differences of patent documents protecting the same inventions in different jurisdictions.	Applications having exactly the same priority or combination of priorities.	NO	**EPO - Esp@cenet equivalents** (www.espacenet.com) **Inno-tec equivalents** (www.inno-tec.bwl.uni-muenchen.de/personen/ professoren/harhoff)
Extended families	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 the application of filters (specific offices, number of offices) to set economic thresholds on patent indicators.	Applications directly or indirectly linked through priorities.	NO	**EPO - INPADOC extended patent** (www.espacenet.com and PATSTAT September 2008 table TLS219_INPADOC_FAM) **OECD Triadic Patent Families** (www.oecd.org/sti/ipr-statistics)
Single priority families	Each first filing is treated individually, as the ORIGIN of a different family.	Statistical analysis of patent filing flows between priority countries and offices of subsequent filings to forecast patent office workloads.	Applications originating from a single priority. In the case of multiple priorities, a given subsequent filing is assigned to multiple single-priority families.	NO	**EPO - PRI system** (see trilateral statistics reports at www.trilateral.net/tsr) **WIPO families** (see world patents report at www.wipo.int/ipstats/en/ statistics/patents)
Examiners technology-based families	Patent documents protecting SAME TECHNICAL CONTENT.	Primarily constructed by and for patent examiners to optimise their work.	Applications with exactly the same "active" priorities, understood as those adding new technical content.	YES	**EPO - DOCDB simple patent family** (DOCDB and PATSTAT September 2008 table TLS218_DOCDB_FAM)
Commercial novelty-based families	Patent documents protecting NEW TECHNICAL CONTENT.	Commercial databases, mainly addressed to help businesses make informed decisions, gain competitive intelligence and monitor industry trends.	Applications with technical content matching existing records. Based on the novelty principle.	YES	**Derwent World Patent Index (DWPI)** (www.thomsonreuters.com/ products_services/scientific/DWPI)

3.1. *Equivalents*

All applications having exactly the same priority or combination of priorities are referred to as "equivalents". They are considered a way to identify patents with the same technical content (protecting the same invention) by uniquely relying on priorities, without any additional expert judgement.

A widely used public data source for equivalents is esp@cenet, the EPO online patent information service, where we can find equivalents for a given patent document in the bibliographic data tab under the "also published as" category on the esp@cenet bibliographic search results.[9] The objective of reporting equivalents at esp@cenet is to display very similar patent documents in different languages. Two subsequent filings are referred to as "equivalents" if all their priorities are the same.[10] Along these lines, in Table 3 below, filings D2 and D3 would be equivalent to each other because they share exactly the same priorities P1 and P2.

Table 3. Equivalents

Equivalents	Patent documents		
	Subsequent filings	Priorities	
	Document D1	Priority P1		
Equivalents (P1, P2)	Document D2	Priority P1	Priority P2	
	Document D3	Priority P1	Priority P2	
	Document D4		Priority P2	Priority P3
	Document D5			Priority P3

The OECD/EPO citations database includes data on EPO esp@cenet equivalents, so that the citation impact of EPO patents and their EPC national equivalents can be measured jointly. Equivalents in the OECD/EPO database are defined as "all the publications in national patent offices pertaining to the same patent [*i.e.* sharing exactly the same priority number(s)]" (Webb *et al.*, 2005).

Another publicly available source for data on equivalents is the experimental dataset of equivalents developed by Dietmar Harhoff, available on the Inno-tec website (Graham and Harhoff, 2006).[11] Calculations are based on the principle that equivalent applications are those sharing exactly the same priorities but explicitly considering priorities in addition to subsequent filings as potential members of equivalent groups. To do so all the applications claimed as priorities in subsequent filings are combined to build "combined priority keys" and each application, *including the priorities which are included by adding self-priority claims*, is allocated to exactly one group of patent documents, provided it has the same priority key as the rest of applications within the group. Each application is assigned to one and only one equivalent group (*i.e.* equivalent groups are mutually exclusive), because the algorithm resolves patterns of

---

[9]   "Esp@cenet is Europe's network of patent databases. Its simple interface is available in most European languages and has been carefully tailored for use by people with little patent searching experience. It contains over 60 million patent documents from all over the world, draws on the same pool of data as raw patent data resources at EPO and contains the same documentation" (from www.espacenet.com).

[10]   The patent linkages considered for the esp@cenet equivalents are: Paris Convention priorities, domestic priorities and technical relations. See Annex I for more information patents linkages.

[11]   www.inno-tec.bwl.uni-muenchen.de/personen/professoren/harhoff.

chained priorities, *i.e.* if application A claims priority of application B, and application B claims priority of application C, it considers C as the ultimate priority of A.[12]

Based on the same principles, in Section 7 below we attempt to formalise a set of "business rules" to identify *mutually exclusive* equivalent groups among patent documents, where we also consider both *subsequent filings and priorities* as potential equivalents. They were developed with the aim to help researchers in cases where the simple definition of "sharing exactly the same priorities" may not be sufficient to develop algorithms that uniquely identify equivalent groups, especially when priorities are considered as potential equivalents of their subsequent filings.

3.2. *Extended families*

The aim of extended patent families is to capture any possible link (direct or indirect) between two given patent documents in order to consolidate them into a single family.

The EPO INPADOC extended patent families are the most widely known example of this type of family. They were first made available by EPO at the esp@cenet website at the end of the 1990s and are now available in an independent table within PATSTAT (since the September 2008 release of the database).[13]

INPADOC extended patent families display every document which is connected to a specific document.[14] Patent documents are first linked to a family even when they have only one priority in common. Further iterative searches are conducted for patents with common priorities with any family member of the initially built family. Thus, the family members do not necessarily have a single priority in common with the one searched for initially.[15] In Table 4 below, documents D1 to D5 belong to the same extended Family (P1,P2,P3).

Table 4. Extended patent family

Extended Family	Patent documents			
	Subsequent filings	Priorities		
Extended family (P1,P2,P3)	Document D1	Priority P1		
	Document D2	Priority P1	Priority P2	
	Document D3	Priority P1	Priority P2	
	Document D4		Priority P2	Priority P3
	Document D5			Priority P3

---

[12] From the presentation of Dietmar Harhoff at the EPO-OECD workshop on patent families held in Vienna in November 2008 (Harhoff, 2008).

[13] EPO database INPADOC includes bibliographic data from over 70 countries and legal status data from more than 40 patent authorities and makes it available at a fee in a standard XML format in the form of weekly updates, as well as cumulated backfiles. The collection comprises bibliographic and legal status data, as well as EP publications, including full text and images. It was integrated into the EPO in the 1990s to combine its particular strengths with the EPO's existing in-house bibliographic database, DOCDB, which is EPO's master database. From the presentation by James Rollinson at the EPO-OECD workshop on patent families held in Vienna in November 2008 (Rollinson, 2008).

[14] The patent linkages considered for INPADOC extended patent families are: Paris Convention priorities, domestic continuations and technical relations.

[15] http://www.piug.org/patfam.php
and http://www.epo.org/patents/patentinformation/about/families/definitions.html.

OECD Triadic Patent Families are built as filtered subsets of INPADOC extended patent families, as those including applications made at the EPO, the JPO and granted by the USPTO (Dernis, Guellec and van Pottelsberghe, 2001; Dernis and Khan, 2004).[16] The restriction to granted patents for USPTO responds to the fact that until 2001, USPTO only published granted patents. Since then, applications are also published 18 months after filing, as in most other offices in the world, but with some restrictions: applications that are not going to be extended abroad can remain unpublished at the request of the applicant. The limitation to US patent grants increases the delay in getting complete data on triadic patent families. The OECD has tried to correct this by "nowcasting" aggregate counts of triadic patent families, in order to provide estimates of most recent years based on counts of previous years (Dernis, 2007).

3.3.   *Single-priority based families*

According to the definition of single-priority based families (also called single first filing forming families), each distinct priority defines a family. A single-priority based family is a group of patent filings that claim the priority of a single filing, including the original priority forming filing itself and any subsequent filings made throughout the world (Hingley and Park, 2003; Hingley, 2009).[17] One important difference with other types of families is that they are not mutually exclusive: a subsequent filing will belong to more than one family if it claims multiple priorities. As a result, if two priority filings are claimed together in an individual subsequent application, two single-priority based families would be counted, each of them including the same subsequent filing plus one of the priorities.

In Table 5 below documents D1, D2 and D3 belong to one single-priority based family P1 and documents D2, D3 and D4 belong to Family P2, whereas documents D4 and D5 belong to Family P3. The three families overlap. Document 4 belongs to Family (P2) and Family (P3), and Documents D2 and D3 belong to both Family (P1) and Family (P2).

Table 5. Single-priority families

Single-priority families			Patent documents			
			Subsequent filings	Priorities		
		Single priority family (P1)	Document D1	Priority P1		
	Single priority family (P2)		Document D2	Priority P1	Priority P2	
			Document D3	Priority P1	Priority P2	
Single priority family (P3)			Document D4		Priority P2	Priority P3
			Document D5			Priority P3

Single-priority based families were first produced by EPO at the end of the 1990s for internal purposes and are stored in the so-called PRI system (Elliot, 1997). They are regularly used in patent filings forecasting exercises at EPO, and are also the basis for the trilateral statistics reports published jointly by EPO, JPO and USPTO.[18] They are used by the Trilateral because they can be easily used to describe the flows of demand for patent rights within and between the most economically active geographical blocs.

---

[16]   www.oecd.org/sti/ipr-statistics.

[17]   Single-priority based families can also be obtained at esp@cenet by introducing a priority number in the appropriate search field, instead of a publication or application number: esp@cenet families with "at least one priority in common".

[18]   Data on Trilateral patent families according to the EPO PRI system definition (*i.e.* active in EPC contracting states, Japan and USA, and also possibly in other countries) are discussed by the Trilateral partner offices together and published annually in the Trilateral Statistical Report, available at www.trilateral.net/tsr.

Families can be classified by their geographical bloc of origin (*i.e.* priority country that defines the family) and the set of blocs (including the bloc of origin) in which the family is active (Hingley and Park, 2003).

WIPO follows a similar methodology to build its own patent families, also based on subsequent filings to single-priorities, where each distinct priority itself defines a family.[19] According to WIPO, a patent family is defined in this way as a set of patent applications inter-related by either priority claims or PCT national phase entries, normally containing the same subject matter (Zhou, 2008).

3.4.   *Examiners' technology-based families*

This type of family is represented by the DOCDB simple patent families, which are primarily constructed by and for EPO examiners to optimise their work.[20] They include patent documents that share identical "priority pictures", understood as priorities adding new technical content. Various methods are used to exclude redundant priorities via the concept of active and inactive priorities. Priority claims that add new technical detail are "active" and included in the priority picture being the basis for a family. Priority claims that do not add new technical detail are "not active" and excluded from the priority picture. As a result, applications that claim the same "active" priorities have identical priority pictures and are considered to cover the same technical content, so that they would be members of the same DOCDB simple patent family. Active priorities would be "first filings" and filings that have properties comparable to those of "first filings". The latter include USPTO continuations in part (expected to introduce new technical detail), USPTO provisional applications (provisionally standing in for the first filing) and abandoned applications (that could have been a first filing). USPTO continuations and divisionals would not be "active priorities" but members of the family of their parent application, as they do not add new technical detail with respect to the parent.[21]

Patent linkages considered are Paris Convention priorities, domestic continuations and technical relations, but a considerable amount of effort is devoted to control that the relations included in the family refer to the same technical content. Indeed, the construction of this kind of family requires human intervention to identify active and inactive priorities: expert judgment based on the type of priority relation and the specific technical content of candidate family members. This is done through quality control and examiners requests and feedback. The stage of quality control consists of "detecting publications that have inadvertently ended up in a new family" and "manual intervention to move them into the simple patent family that covers the appropriate technical content" (Versloot-Spoelstra, 2008). Such human intervention makes DOCDB simple families different from other EPO families previously described, which were uniquely based on relations available in databases with no *ex-post* reallocation of family members or re-design of family boundaries based on any expert assessment of technical content. This difference is important because it means that DOCDB patent families cannot be replicated by individual researchers, who would need to rely on the end-result made available by EPO. Data on this type of families is currently available through two main channels: DOCDB and PATSTAT.[22]

---

[19]   www.wipo.int/ipstats/en/statistics/patents.

[20]   DOCDB is the master database of the European Patent Office. It is regularly fed with information from national patent offices on published documents. It is used by patent examiners to search prior art, and is the source of raw patent data for other EPO databases, included PATSTAT. See the manual for DOCDB database at: http://www.epo.org/patents/patent-information/raw-data/manuals.html#docdb.

[21]   From the presentation by Fenny Versloot-Spoelstra from EPO at the EPO/OECD workshop on patent families held in Vienna in November 2008 (Versloot-Spoelstra, 2008). I would like to thank Fenny Versloot-Spoelstra for further clarifications about the methodology used to build this type of families.

[22]   Table TLS218_DOCDB_FAM in PATSTAT September 2008, which only lists published applications that are family members, without indicating which ones belong to the priority picture.

3.5. *Commercial novelty-based families*

A well-known commercial database of patent families is Derwent World Patents Index (DWPI), part of Thomson Reuters.[23] DWPI is a comprehensive commercial database of enhanced patent documents where experts analyse, abstract and manually index every patent record. In the early days, DWPI only covered chemical patents from most major countries and by the mid-1970s extended to all technologies from 24 national patent offices.[24] DWPI today contains over 17.4 million records covering more than 37.2 million patent documents, with coverage from over 41 major patent issuing authorities worldwide. [25]

DWPI families are constructed based on the novelty principle where new members have matching technical content with previous ones, so that not only the priority claims are important to structure family relations, but also the timing in which applications enter the DWPI system. The methodology can be summarised in three steps. First, the priority details of new documents are analysed against those already in DWPI. Second, patents with priority details not seen before are termed 'basic', and a new family is created on the basis of them with a new DWPI database record. Third, new patents with priority data matching an existing DWPI record are termed 'equivalents', and the patent becomes a new family member within that DWPI record. A new application will only be considered 'equivalent' to an existing application in the DWPI system if the set of priorities of the new application are included in the set of priorities of the basic document used as the basis for the existing DWPI record.[26]

It is worth noting the treatment given to divisionals and USPTO continuations and continuations in part. On the one hand, divisionals and USPTO continuation applications maintain the same status as their parent applications. For instance, if a UK patent application GB1 is a basic document, and another UK application GB2 is divisional to GB1, then GB2 will be the basic document of its own family. However, if GB1 is equivalent to another document already in the DWPI database, then GB2 will join that existing family as a new equivalent. On the other hand, USPTO continuation-in-part applications are always considered as basic documents and given a new DWPI record to reflect their additional technical content[27], and cross-referenced back to the original. In most cases, a complete patent family will be gathered into a single DWPI record, but in cases such as the one just described, it may be spread across two or more records. It would thus sometimes be necessary to gather together all the members of a "scattered" patent family, as family relationships will be defined by the order in which patents appear in Derwent WPI.[28]

The methodologies used by EPO and Derwent are not very different in two respects. First, in our view, single DWPI records would be like "equivalent groups" and the combination of interrelated DWPI records forming a broader family is close to the concept of "extended family". Second, Derwent uses traditional patent linkages such as Paris Convention priorities and domestic continuations, but it also flags-up potential 'non-convention equivalents', defined as patents with the same technical content as an existing

---

[23] There exist other private providers of patent family data, such as Questel-Orbit (FAMPAT). The patent family definition used by Questel-Orbit is close to that of extended patent families, and could be described as "all applications sharing at least one priority in common belong to a single family", where "priority" is understood in a broad way, to comprise Paris Convention priorities, intellectual priorities, domestic priorities and PCT links (www.questel.com).

[24] http://scientific.thomsonreuters.com/derwent/history/.

[25] http://www.thomsonreuters.com/products_services/scientific/DWPI.

[26] Prior to mid-1992, a patent was considered to be an 'equivalent' within the DWPI system if it claimed the same latest priority as another patent already recorded in the WPI system. Since the 16th week of 1992, the priorities of a new application must exactly match the priorities of an existing DWPI record in order for the patent to be incorporated into it. If this criterion is not met, the patent in question will be placed in a separate DWPI record.

[27] In this respect the methodology to identify DWPI equivalents is similar to that used to build DOCDB simple families.

[28] http://scientific.thomsonreuters.com/support/patents/dwpiref/reftools/searchtips/searchtip_apr.

DWPI family, but not claiming the same priority, which would be like the "technical relations" identified by EPO (see Annex I for information on different types of patent linkages). However, one important aspect makes EPO and Derwent methodologies differ: the consideration of time (novelty) in DWPI families. One peculiarity of Derwent patent families, not shared by EPO families, is that the family seeds are patent documents identified by Derwent analysts as "basic" because they add novel technical content to the patent system. The use of the concept of novelty for family building implies that timing counts.

In Table 6 below, Document D1 is the first one appearing in the system (filed first), followed by Document 2. Document D3 being the last one to enter the system. The basic document for DWPI family record D1 is Document D1, with priorities P1, P2 and P3. Document D2, with priorities P1 and P2, is equivalent to D1 and thus also belongs to DWPI family record D1, because its priorities are included in the set of priorities of basic D1. However, Document D3, with priorities P1, P2 and P4, has an additional priority P4, so that it has to form a new family (DWPI family record D2) and become a basic document itself. Nevertheless, since the basic documents for DWPI family record D1 and DWPI family record D2 have two priorities in common (P1 and P2), they would be cross-referenced in Derwent´s database and appear as interrelated DWPI family records, giving the possibility to build Extended Family D1.[29]

Table 6. Derwent (DWPI) – novelty-based families

DWPI "extended family"	DWPI "equivalents"	Patent documents				
		Subsequent filings	Priorities			
Extended Family D1 (interrelated DWPI family records D1 and D2)	DWPI family record D1	Document D1	Priority P1	Priority P2	Priority P3	
		Document D2	Priority P1	Priority P2		
	DWPI family record D2	Document D3	Priority P1	Priority P2		Priority P4

4. Comparing family counts based on different definitions

Several aspects of the definitions just described can lead to differences in patent family counts, and affect statistics made on the basis of one definition or another. Five factors may cause differences in total family counts and family statistics coming from different sources: *i)* the use of expert criteria, in addition to priority links, to refine patent linkages among family members; *ii)* considering indirect priority links, in addition to direct links, to form families; *iii)* allowing a given patent document to belong to more than one family or not; *iii)* including unpublished patent documents, in addition to published ones, as family members; and *iv)* imposing geographic, technological or time filters on the definition of families.

The first factor refers to the use of expert control to check the validity of the priority links reported in the database (*i.e.* by comparing with the original documents) and, eventually, to identify additional relations based on the similarity of technological content and applicants in apparently unrelated patent documents. Simmons (2009) cites typographical errors and changes in database standardisation criteria as possible causes of misrepresentation of patent families and differences in family counts across different sources. Having experts (or patent examiners) that read all patent documents included in a family is of course the preferred option to build families but only manageable for patent offices or large corporations, and not affordable for individual researchers. We will leave the analysis of possible differences in outcomes based on the two family definitions described earlier that use expert control (DOCDB simple families and DWPI families) out of the scope of the rest of this paper, since our intention is to focus on replicable patent families methodologies that researchers using priority links reported in patent databases can use.

---

[29]   http://www.epo.org/patents/patent-information/about/families/thomson.html and
         http://www.wipo.int/meetings/en/2003/statistics_workshop/presentation/statistics_workshop_nanu.pdf.

The second factor relates to the debate about whether indirect links between two given patents should be taken into account to form broader families or not, that is, to where a patent family finishes and another family starts. Patent equivalents would be an example of narrow families that only rely on direct links, and extended patent families would be an example of broad families that consolidate both direct and indirect links. Since equivalents would be subsets of extended families, the impact of including indirect links would be to reduce the number of families with respect to the number of equivalents, as a single family with complex relations within it may include multiple sets of equivalents (see Section 7).

Third, allowing a given patent to belong to more than one family will give different results to imposing the condition that a given patent can only belong to a single family (*i.e.* mutually exclusive families). Single-priority families are an example of families that may not be mutually exclusive. Taking priority year 1999, a comparison between the number of INPADOC extended families and single-priority families using the EPO and WIPO methodologies shows the impact of building a different family from each single first-filing or not (Figure 1). Both EPO-PRI total number of first filings by country and WIPO total number of patent filings follow a similar approach, namely, whenever two first filings lead to a single subsequent filing they consider that two families are formed.[30] In contrast, the INPADOC extended family concept would consider that only one family is formed. As shown in Figure 1, single-first filing forming 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.[31] It is also worth noting that a large share of the INPADOC extended families are singletons (36% for the United States and Germany, 38% for France and as much as 85% for Japan and 56% for the UK).[32]

---

[30]   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 forming families, or the patent databases on which their algorithms are implemented. Exploring these possible additional differences is beyond the scope of this paper.

[31]   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 Germany. The rest of the countries have lower shares.

[32]   The shares for Japan and the United Kingdom are high possibly due to lack of relations among applications or to lack of information about relations in the database. PATSTAT lacks indeed information about divisional applications at JPO (EPO does not receive such data from JPO) but, to our knowledge, it includes all patents from UKPO. The data on EPO-PRI and WIPO families presented here is based on published aggregate counts which do not provide information on the number of singletons.

Figure 1. Extended families v single-priority based families in the top priority offices

Earliest priority year 1999. Including singletons
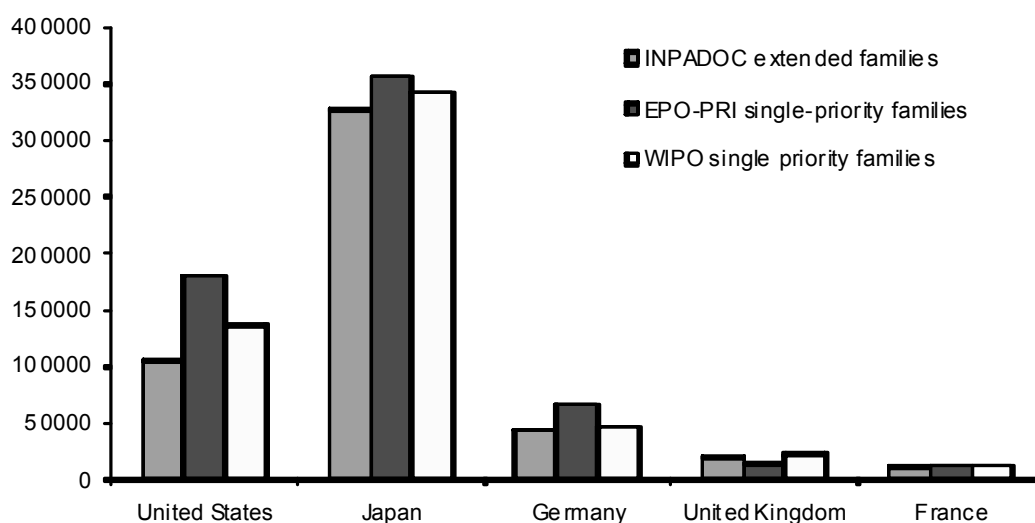
Source: Counts of EPO-INPADOC extended families from PATSTAT, data on EPO-PRI families correspond to updated figures from Table B1 in Hingley and Park (2003), provided by P. Hingley, and data on WIPO families come from www.wipo.int/ipstats/en/statistics/patents, patent families by country of origin (1990-2005), as of 3 June 2009.

The impact of the fourth factor, including unpublished documents as family members or not, can be partially assessed by comparing counts of DOCDB simple families and INPADOC extended patent families (excluding singletons).[33] A comparison between all the applications and families included in the INPADOC and DOCDB family tables in PATSTAT September 2008 shows that they have 630 856 (non-singleton) families in common: 48% of all INPADOC families and 56% of all DOCDB families with more than one member are identical (Figure 2). In addition to the expert criteria used in DOCDB family building, absent from the INPADOC methodology, the main difference between both definitions is their reliance on different types of documents as potential family members. DOCDB family members are all published patent documents, whereas INPADOC family members include not only published applications but also unpublished applications that have been cited or claimed as priorities in other published applications.[34] As seen earlier, one of the peculiarities of DOCDB simple patent families is the distinction of treatment between the applications that form the "priority picture" (generating the family relations) and the applications that become "family members", so that even though the "priority picture" that forms the DOCDB family (where the inclusion of linkages is validated by experts) can also include unpublished applications claimed as priorities, unpublished documents are never considered as family members, *i.e.* PATSTAT does not list them as part of the DOCDB simple family.

---

[33] PATSTAT September 2008 includes information on 11 456 887 applications with a filing date between 1991 and 1999 (number of distinct appln_id in table TLS_201_APPLN), forming 6 643 800 DOCDB simple families and 6 010 859 INPADOC extended patent families with earliest priorities filed during those years. Excluding families with one member only (singletons) the DOCDB family table reports 1 134 916 families and the INPADOC family table 1 311 613.

[34] In practical (PATSTAT) terms: all DOCDB family members are drawn from PATSTAT table PAT_PUBLN, whereas INPADOC family members are drawn from table APPLN. All applications included in table TLS_201_APPLN are also in table TLS_219_INPADOC_FAM, except those with appln_id above 55000001, which correspond to artificial surrogate keys created for the completeness of PATSTAT (PATSTAT September 2008 Datacatalog).

**Figure 2. Number of DOCDB and INPADOC families in PATSTAT
(excluding singletons)**

Earliest priority date: 1991-1999
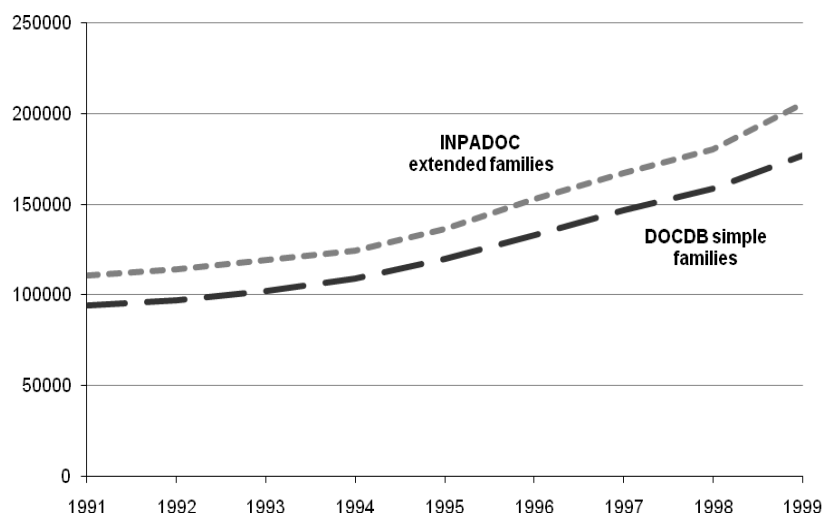
Note: The "DOCDB simple families" series consists of counts of distinct docdb_fam_id from table TLS218_DOCDB_FAM having more than one application as family member, and the "INPADOC extended families" series consists of counts of distinct inpadoc_fam_id from table TLS219_INPADOC_FAM, also having more than one application as family member.

Source: PATSTAT September 2008.

The fifth factor relates to imposing filters on existing families for the purpose of the analyses and to build indicators. Now that PATSTAT provides ready-made tables with extended families (INPADOC family table), imposing different filters on them opens very interesting avenues of research to build new indicators based on family data. Figure 3 compares the total number of INPADOC families with earliest priorities in year 1999 reported in PATSTAT with those resulting from imposing the following restrictions:

i.   "Non-domestic", having members of at least two different patent offices, thus excluding purely domestic families from the total.[35]

ii. "Transnational", including at least one PCT or one EPO application (Frietsch and Schmoch, 2010).

iii. "Triadic", having at least one USPTO grant, one EPO application and one JPO application as family members (Dernis and Khan, 2004).

Of the 165 128 INPADOC families with earliest priority in 1999, 80% are non-domestic or international (20% are domestic, *i.e.* all applications are filed at the same patent office); 61% transnational and 21% triadic. Non-domestic (or international) families are thus the least restrictive filter, whereas triadic would be the most demanding one.

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 used by researchers and statisticians aiming to analyse the most valuable patents of a country. Given the highly skewed distribution

---

[35]   This concept is similar to the foreign-oriented patent families published by WIPO, although they impose the "international" filter on WIPO patent families (which are single-priority based families) and here it is imposed on extended families.

of patent value, they tend to represent a small share of all extended patent families, as shown in Figure 3, with some variations across countries (they also show, by definition, a bias towards USPTO, JPO and EPO patents with respect to patents from other countries). At the other extreme, the least restrictive international patent families are those where the only condition imposed is to involve more than one patent office: the non-domestic families. The United States and Japan show a relatively large gap between the total number of families and the number of non-domestic ones, indicating the importance that domestic families have in these countries. Finally, imposing the condition that at least one family member has been filed through supranational filing procedures (PCT or EPO), we get the transnational families, which would be a filter of quality (only those patents with high expected commercial value are applied through supranational procedures), but less restrictive than the triadic one, which would make them more suitable to the analysis of developing countries, and unbiased to specific offices (Frietsch and Schmoch, 2010).

Figure 3. Imposing filters on extended families, family counts by priority office
Excluding singletons. Earliest priority year: 1999
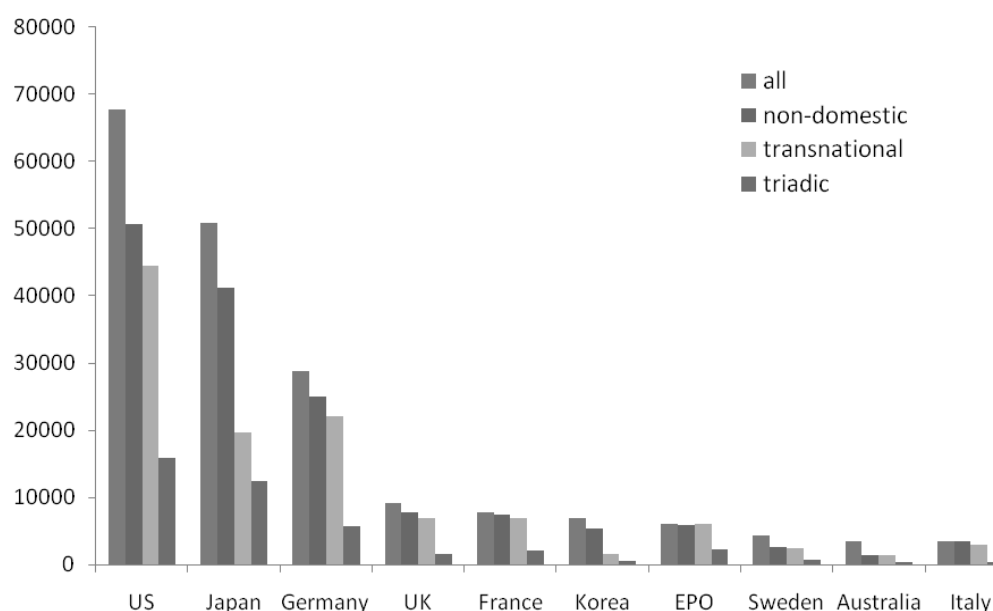
Source: Author's calculations based on PATSTAT September 2008.

5.   **Building extended families**

Apart from comparing patent family data from available sources, the aim of this paper is to provide guidance on how to build families from scratch, using raw information from the Worldwide Patent Statistics Database produced by EPO (PATSTAT).[36]

PATSTAT is a relational database containing most relevant fields for the statistical analysis of patent data and is mainly directed to researchers. Information is organised in several tables that can be connected through the unique identifier of each patent application, called "Application ID", which is unique because it is the result of combining information from three fields: application authority, application number and

---

[36]   PATSTAT was launched in the first semester of 2006 at the request of the Patent Statistics Task Force led by the OECD, which also includes EPO, USPTO, JPO, WIPO, NSF and the European Commission (represented by Eurostat and DG Research). New versions with updated information are released twice a year.

application kind.[37] Most of the raw data used to build PATSTAT comes from DOCDB, which means that the information in PATSTAT is as good as the information EPO receives on published patent documents from national offices. This means that its timeliness is determined by publication delays of patent documents imposed by the rules of the patent system, but also that not all patent offices are equally represented in PATSTAT, although completeness is improving with every new release of the database, and not all different kinds of patent linkages may be reported to the same extent by all national patent offices to the EPO.

The philosophy of the PATSTAT database has always been to allow researchers as much freedom as possible in the use of the raw data, providing sufficient information and documentation to be able to do meaningful calculations, but without imposing any specific methodology or definition of indicators. However, the difficulties perceived by researchers to build families by themselves were higher than for other indicators, and individual family data was so easily available at the EPO website esp@cenet, that the users demand for family data was accepted and two types of families were chosen for their inclusion as "ready-made tables" in PATSTAT during 2008: in the April 2008 version of PATSTAT, a new table was added with information on "DOCDB simple families", and in the September 2008 version, an additional table was included with information on "INPADOC extended families".

In this section we will describe the kind of algorithm that can be used to build INPADOC extended families and try different sources of patent linkages to come up with different variations of extended families. As for the production of the INPADOC family table, we will focus on unfiltered families, but acknowledge that it may sometimes be preferable to focus on specific geographic regions and impose filters for the selection of patent family members (*e.g.* triadic patent families would be extended patent families with filings at USPTO, JPO and EPO), based on data availability in PATSTAT and research interests. Also, we will not restrict our analysis to families composed exclusively by patent, but consider both patents and utility models. The term "applications" in PATSTAT comprises both utility models and patents. Sometimes patent families have utility models as their priority filings and patents as subsequent filings, and there are also cases (very few) where families are only made of utility models.[38]

Finally, we will limit our attention to patent families having filing dates of earliest priorities not later than 1999, in an attempt to reduce as much as possible the timeliness issue (common to all patent databases based on published documents) [39] and before 1991, as EPO has better records of patent linkages and their different types since then.[40] Our source will be PATSTAT September 2008, and our universe the 11 456 887 applications with a filing date between 1991 and 1999 which it reports.[41] It is important to

---

[37]   Although it should be kept in mind that it changes at every release of the database, as it is an internally generated ID.

[38]   There are two types of Intellectual Property Right types in PATSTAT table APPLN. Utility models are identified as "UM" and patents as "PI", standing for patents of invention. PATSTAT September 2008 has information on 61 497 371 different applications, 88% are patents and 12% are utility models, applied between 1814 and Summer 2008 at one of the 168 different application authorities for which PATSTAT has records. Among all the patent families (excluding singletons) with earliest priorities between 1991 and 2001, 4% have a utility model as the earliest priority application, but most of those families have patents as subsequent filings for the utility model. Only a few are fully made of utility models (less than 1% of the total).

[39]   The number of patent applications per extended family stays in the range 5.4-5.5 between 1991 and 1999, but decreases to 5.0 in 2000 and 4.8 in 2001. This happens for any kind of extended families, calculated using any combination of patent linkages in PATSTAT, which indicates that it may be solely due to the fact that not all the family members have yet been published for families with earliest priorities in most recent years.

[40]   As noted in the PATSTAT Data Catalog: "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.)." (European Patent Office, PATSTAT Data Catalog September 2008, 2008, p.37).

[41]   Number of distinct appln_id in table TLS_201_APPLN.

stress that our source is the September 2008 version of PATSTAT, so that the information presented here reflects the situation on patent linkages published in the patent databases that feed PATSTAT as of mid-2008. Ten years should be sufficient to have a picture as complete as possible of the families analysed, even though we might still find a few cases of families with earliest priorities in the 1990s that still add new members ten or fifteen years later (*e.g.* divisionals, continuations, etc). A family is a dynamic concept: patent families evolve over time and members might always be added as new patent documents are published and information becomes available in databases (van Pottelsberghe and van Zeebroeck, 2008).

5.1.   *Sources of family relations in PATSTAT*

The first step to build patent families is to identify the relations that create linkages among patents and understand their nature. Four different types of patent linkages among patent filings can be used to build patent families: *i)* Paris Convention priorities; *ii)* technical similarities (also called non-convention priorities, intellectual priorities or technical relations); *iii)* domestic priorities (*e.g.* continuations, continuations in part, provisionals, divisionals); and *iv)* PCT regional/national phase entries.

Paris Convention priorities, domestic priorities and PCT national phase entries are reported in PATSTAT as set out in patent documents. In turn, technical similarities are identified, and added to the database as artificial priority links, based on the expert judgement of EPO patent examiners (Table 7). Some family definitions use only a selection of these relations. Others use them all. See Annex I and Annex II for more detailed definitions of all these types of patent linkages and how they are reported in PATSTAT.

Table 7. Patent linkages

Type	Definition	Claimed by applicant in patent document
Paris Convention priorities	Allow a one year delay between first original filing and subsequent foreign filings by same applicant claiming the priority right (1883 Paris Convention).	YES
Technical similarities	Relations among patent documents with similar scope, inventor and applicant names, that nevertheless lack common priority. An artificial priority link is assigned manually by the database producers.	NO
Domestic priorities	Filed at the same office. They are mainly continuations, continuations in part and provisionals (the three of them only available at USPTO), and divisionals, which are available at most patent offices (1883 Paris Convention).	YES
National phase entries of PCT filings	Entry into regional/national phases of PCT filings.	YES

EPO validations in EPC member countries could also be considered as an additional type of patent linkage, but we do not include them in Table 7 above given their post-grant character. All the other patent linkages listed in the table are relations between patent filings, whereas EP validations happen only if the corresponding EP patent filing has been granted at EPO and the applicant takes the necessary steps to validate such a grant in the EPC countries of his choice, among more than 30 countries (*e.g.* translation, payment of national validation fees, etc) (see Annex I for more information about EPO validation procedures).

5.2. *Methodology to build extended families*

As noted earlier, the launch of PATSTAT provided an opportunity for researchers to build patent families by themselves, relying on the information on patent linkages available in the database. The only requirement was to be willing to invest in IT skills, count with help from an IT expert and learn about all the nuts and bolts of the patent system and patent databases, which may take a substantial amount of time and effort after all. In this section we summarise some important pieces of information learned during such a process, with the aim to provide some sort of PATSTAT family-building toolkit. Extended patent families are the focus of this section. Being the broadest possible type of family (as they include both directly and indirectly linked patents), they can be very useful as the basis for comparisons across different definitions. The basics of the algorithm used to build extended families using information on patent linkages are described below (see Box 1 below).

Box 1. Methodology to build extended patent families using patent linkages in PATSTAT

Step 1. Two tables are loaded based on information from PATSTAT: *i)* a table with all the applications (appln_id) wished to be considered as potential family members; *ii)* a table with relations among applications. Intuitively, we refer to these relations as being of the type "parent" (priority) and "child" (subsequent filing). Relations in PATSTAT are always of the kind "I am the Child of my Parent" (*i.e.* I am the subsequent filing of my priority). Applications for which no linkage with any other application is reported are considered as first filings.

Step 2: A list of singletons is made, comprising all the applications that are not included in the relations table, as either parent or child.

Step 3: To calculate normal families, a list of candidate family members is made from applications included in the relations table. To optimise time and calculations, only applications that are first filings or "family seeds" are selected first, *i.e.* applications that are uniquely cited as "parents" in the relations table, not as "children".

Step 4: One application is chosen from the list of potential family members, let us call it M(n, i), where n is the identification number of the new family and I is the iteration round.

Step 5. A family F(n) is created, with M(n, 1) being its first family member.

Step 6: From this point onwards, the following procedure is iterated until there is no new family member attached to Family F(n):

- Look for all applications with a relation type "parent" or "child" with respect to M(n, i), which does not yet belong to F(n).

- Integrate members M(n, i+1) into Family F(n)

- Flag member M(n, i) as "calculated" when all its relations have been calculated

- Flag new members M(n, i+1) as "not calculated" when its relations still need to be calculated

- Once all the members of a family are found (*e.g.* no new member can be added), they are eliminated from the list of candidate family members.

5.3. *Extended families using different sources of relations*

As regards the choice of patent linkages to be used in the construction of extended families, PATSTAT gives several options based on combinations of different types of patent linkages (see Annex II).[42] We have chosen four possible combinations and called each one of them a "source of family relations", as shown in Table 8 below, where **source 3 corresponds exactly to INPADOC extended patent families**, as reported in PATSTAT.

---

[42] The names given to each kind of patent linkage have been chosen to be as close as possible to PATSTAT table titles. There is no standard terminology in this field (*e.g.* domestic continuations are sometimes also called domestic priorities), however. See Annex I for more information on definitions of patent linkages.

Table 8. Selection of PATSTAT sources for family relations

Source of family relations	Types of patent linkages taken into account	PATSTAT tables on patent linkages used	PATSTAT ID of the PARENT	PATSTAT ID of the CHILD
Source 1	Paris Convention priorities	APPLN_PRIOR	Prior_appln_id	appln_id
Source 2	Paris Convention priorities Domestic continuations	APPLN_PRIOR APPLN_CONTN	Prior_appln_id Parent_appln_id	appln_id appln_id
Source 3	Paris Convention priorities Domestic continuations Technical similarities	APPLN_PRIOR APPLN_CONTN TECH_REL	Prior_appln_id Parent_appln_id Tech_rel_appln_id	appln_id appln_id appln_id
Source 4	Paris Convention priorities Domestic Continuations Technical similarities PCT regional/national phase entries	APPLN_PRIOR APPLN_CONTN TECH_REL APPLN	Prior_appln_id Parent_appln_id Tech_rel_appln_id Internat_appln_id	appln_id appln_id appln_id appln_id

Table 9 below shows that the impact of adding new patent linkages is always positive for the total number of families, but only until source 4. Thus, adding PCT national phase entries consolidates existing families, whereas the other three types of links join former singletons together to form new independent families. Another important highlight from a comparison of the sources: Paris Convention priorities alone make up more than 95%, which make them the most relevant patent linkage by far in the construction of extended patent families.

Table 9. Counts of families and applications in them, by source of family relations
Excluding singletons, earliest priorities 1991-1999

	Source 1 families		Source 2 families		Source 3 families - INPADOC extended-		Source 4 families	
	(Paris Convention)		(Paris Convention + Domestic continuations)		(Paris Convention + Domestic continuations + Technical similarities)		(Paris Convention + Domestic continuations + Technical similarities + PCT national phase entries)	
	#families	#members	#families	#members	#families	#members	#families	#members
1991	106 850	567 024	110 371	610 367	110 856	614 631	110 745	614 888
1992	107 873	571 499	113 659	624 205	114 394	629 235	114 276	629 538
1993	112 351	602 058	119 014	655 380	119 656	659 896	119 533	660 106
1994	116 602	639 485	123 818	693 431	124 606	698 404	124 362	699 194
1995	129 535	722 318	135 946	761 371	136 702	766 020	136 405	766 939
1996	146 281	805 300	152 088	849 307	152 994	854 422	152 681	855 240
1997	161 060	874 480	166 277	914 281	167 220	919 518	166 857	920 284
1998	173 243	939 352	179 345	985 989	180 095	990 508	179 618	990 914
1999	196 972	1 057 368	204 330	1 102 833	205 222	1 106 929	204 659	1 107 482
1991-1999	1 250 767	6 778 884	1 304 848	7 197 164	1 311 745	7 239 563	1 309 136	7 244 585

Note: Source 1: APPLN_PRIOR; Source 2: APPLN_PRIOR + APPLN_CONTN; Source 3: APPLN_PRIOR + APPLN_CONTN + TECH_REL; Source 4: APPLN_PRIOR + APPLN_CONTN + TECH_REL + PCT links (relations between appln_id and internat_appln_id in APPLN). Source 3 families are INPADOC families, reported in ready-made family PATSTAT table TLS_219_INPADOC_FAM.

Source: Author's calculations based on PATSTAT September 2008.

The distribution of extended patent families by country of earliest priority for the four different sources of patent linkages is quite similar (Figure 4). The main differences appear for the United States and Germany when passing from Source 1 (only Paris Convention) to Source 2 (Paris Convention and domestic continuations), given the frequent use of continuations at USPTO and "internal priorities" at the German Patent Office (see Table A1 in Annex II).

**Figure 4. Top Ten patent offices of earliest priorities in extended families
using different sources of patent linkages**
Excluding singletons, earliest priority date: 1991-1999
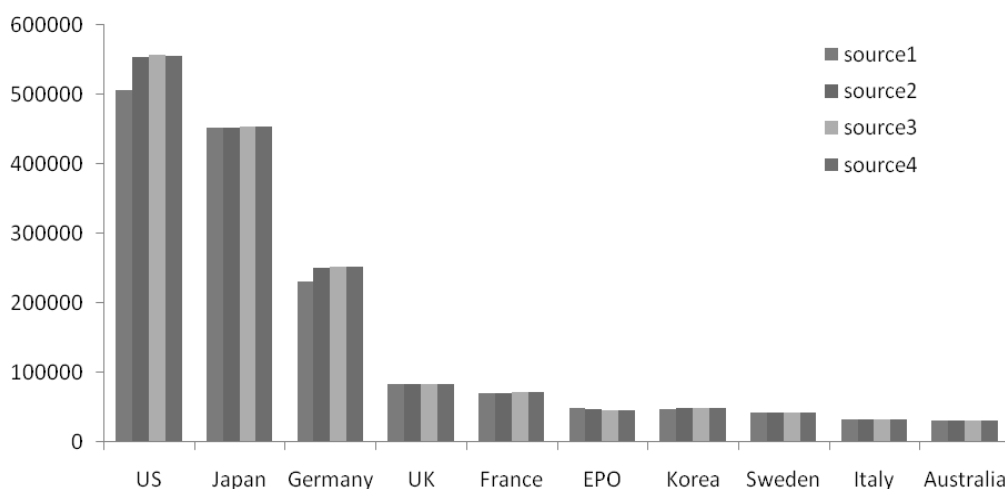
Note: Source 1: APPLN_PRIOR; Source 2: APPLN_PRIOR + APPLN_CONTN; Source 3: APPLN_PRIOR + APPLN_CONTN + TECH_REL; Source 4: APPLN_PRIOR + APPLN_CONTN + TECH_REL + PCT links (relations between appln_id and internat_appln_id in APPLN). Source 3 families are INPADOC families, reported in ready-made family PATSTAT table TLS_219_INPADOC_FAM.

Source: Author's calculations based on PATSTAT September 2008.

6.   Simple and complex family structures

Most studies of patent families treat patent families as "sets" of patents, without looking inside them to see how patent linkages bring family members together. For the first time in the analysis of patent families, to our knowledge, we have performed an exhaustive analysis of family patterns and the structure of relations within families, analysing the frequency of different family structures. This section presents the main results of such analysis, from which two main conclusions emerge: most patent families are characterised by similar structures and the most frequent ones are quite simple.[43]

As a first approach, Table 10 below summarises two main characteristics of the families having earliest priorities 1991-1999 that can be identified straight away, without using any sophisticated algorithm. For any of the four different sources considered, first, more than 85% of all families have a single earliest priority; and, second, around 30% of all families have only two family members. These two findings point to a large frequency of simple patterns among families.

---

[43]   The identification of family structures relies on information on patent linkages as reported in PATSTAT September 2008. Results may be slightly different when applied to new versions of PATSTAT as new family members may be added to existing families (as they become published), new patent linkages added and new families formed.

26

Table 10. Share of families, by source of family relations and family characteristics
Excluding singletons. earliest priorities 1991-1999

	Sources of family relations			
Family characteristics	**Source 1**	**Source 2**	**Source 3**	**Source 4**
Single earliest priority	89%	87%	86%	86%
Only two family members	29%	30%	30%	30%

Note: Source 1: APPLN_PRIOR; Source 2: APPLN_PRIOR + APPLN_CONTN; Source 3: APPLN_PRIOR + APPLN_CONTN + TECH_REL; Source 4: APPLN_PRIOR + APPLN_CONTN + TECH_REL + PCT regional/national phase entries. Source 3 families are INPADOC families, reported in ready-made family PATSTAT table TLS_219_INPADOC_FAM.

Source: Own calculations based on PATSTAT September 2008.

Based on the latter we could say that around 30% of all patent families, of any sort, are really simple, composed of just one priority and one subsequent filing. But, what happens with the remaining 70% of patent families? What do they look like?

To address this, we developed an algorithm that positions each family member in the family tree and counts both the number of generations within a family and the number of family members within each generation. Based on the results of implementing this algorithm, we were able to identify and count all the different structures adopted by (normal and cyclical) patent families during a certain priority period, and classify them accordingly. The algorithm characterised a "family structure" as the unique representation of a given family, providing information on the family pattern and number of family members (*e.g.* how many different family generations, how many applications within each generation, how many family members in total, etc).

As a result, we obtained a list of all the different family structures for all the families obtained using each of the four different sources of family relations we had selected in PATSTAT. Table 11 below presents the main results of these calculations. In addition, Table A5 and Table A5bis in Annex IV list the top 25 family structure identification numbers for each of the 4 different family sources, so that it is possible to see which structures appear most frequently in the 4 types of families, and how their position changes in the ranking for each source. They also inform on whether the structures listed follow the simple pattern of "one parent – several direct children" and the number of applications they include. Finally, Table A6 in Annex IV includes a graphical representation of family structures appearing in the four different top 10.

The first two rows in Table 11 below present the share in the total number of families characterised by the top 10 and top 25 structures, that is, the 10 and 25 different family structures that concentrate the highest percentage of families. The last row in Table 11 indicates the total number of different family structures for each type of family. Unsurprisingly, the number of different family structures increases as new sources of patent linkages are added in the family-building methodology, so that the share of families characterised by the top 10 and the top 25 structures diminishes as new linkages are added.

Focusing on **family source 3, which corresponds to INPADOC extended patent families**, we find that the top 10 family structures concentrate 73% of all the families with earliest priority 1991-1999, and the top 25 as much as 81%.

Table 11. Family structures by source of family relations
Excluding singletons. Earliest priorities 1991-1999

	Source 1	Source 2	Source 3	Source 4
Share of families with top 10 structures	80%	75%	73%	63%
Share of families with top 25 structures	89%	83%	81%	74%
Share of families with simple structure characterised by a single priority and one or several direct subsequent filings	84%	76%	75%	56%
Number of different family structures	19 584	44 161	47 437	64 228

Note: Source 1: APPLN_PRIOR; Source 2: APPLN_PRIOR + APPLN_CONTN; Source 3: APPLN_PRIOR + APPLN_CONTN + TECH_REL; Source 4: APPLN_PRIOR + APPLN_CONTN + TECH_REL + PCT regional/national phase entries. Source 3 families are INPADOC families, reported in ready-made family PATSTAT table TLS_219_INPADOC_FAM.

Source: Author's calculations based on PATSTAT September 2008.

Grouping similar family structures together, in particular those characterised by one single earliest priority and one or several direct subsequent filings emerging from that priority, we further characterise families, and discover that the apparent heterogeneity of different family structures, within each source of family relations, can be reduced. We add up the number of families by source that are characterised by structured membership of this group, and find that around 75% of all INPADOC extended families (source 3) are characterised by such a simple pattern (see third row in Table 11). This finding has important consequences for the comparison of outcomes across different family-building methodologies, as it amounts to saying that 75% of all INPADOC extended families, are also single-priority families and equivalent groups. The simple structure of "one-parent and its direct children" guarantees this is the case, all the conditions of single-priority families and equivalents described earlier in Section 3 are satisfied. The remaining 25% is nevertheless non negligible and the fact that different family definitions may provide different outcomes for them calls for transparency about family definitions and methodologies in research using patent family data, as well as on the type of patent linkages used. Adams (2006, p.15) already intuitively advanced that the main source of family variations would come from 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".

As regards the discussion about normal and cyclical families included in Annex II, it is worth noting that the simplest structure of cyclical families, composed of just two members, appears among the top 25 for family sources 3 and 4, representing 0.6% of the total number of families for each of these two sources.[44]

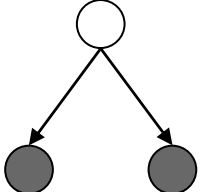## 7. Identifying equivalents based on internal family structures
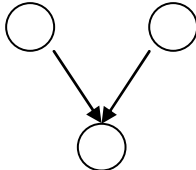
One of the main objectives of having a database of patent families is to enable the identification of patent documents protecting "the same invention" in different jurisdictions. However, we have seen that there is a large variety of family structures when considering both direct and indirect links between patents. The only reliable way to identify patent documents protecting "exactly" the same invention would thus be to ask experts to read all the patent documents within a given family and assess if they protect the same invention or not, but this is something clearly out of the scope of most economic research projects.

---

[44] Given that it appears in the ranking position 135 for family source 1, it has been given the Structure Number ID 135, and takes positions 135, 113, 14 and 19 for sources 1, 2, 3 and 4 respectively (see Table A5bis in the Annex).

Patent equivalents understood as patent documents sharing "exactly the same priorities", as defined earlier in Section 3, can be considered the safest way to identify all the documents protecting "exactly the same invention" when it is only possible to rely on patent linkages reported in databases (excluding additional expert judgement and control). They can therefore be considered as subsets or building blocks of extended patent families, where the key would be to decide where to draw the boundaries between equivalents and non-equivalents within extended patent families.
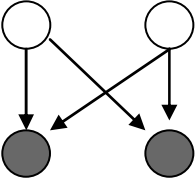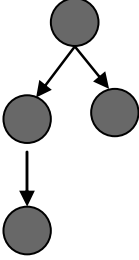
The aim of this section is to propose a refined method for the identification of mutually exclusive groups of patent equivalents within complex families using four simple rules, as an extension of the simple rule of identical priorities. We do that based on what we have learned in the previous section about the structure of patent families, as well as on the nature of the different patent linkages available. They are set out in Figure 5 below, where equivalents resulting from the application of each rule are highlighted to be distinguished from the non-equivalents according to that specific rule. Figure 6 presents two examples showing the result of implementing the four different rules together.

Figure 5. Proposed rules to identify equivalents

Rule 1 - brotherhood	Two (or more) given subsequent filings would be equivalents between them if they share exactly the same priorities.	
Rule 2 - one parent	In the case of a single priority, subsequent filings would be equivalent to their priority.	
Rule 3 - multiple parents	Equivalence is not satisfied between multiple priorities and their immediate subsequent filings, as in that case each priority would, in principle, protect a different invention, and their subsequent filings would protect a derived invention, with elements from all the priorities.	
Rule 4 - inheritance	If a subsequent filing claims the priority of an application that is a subsequent filing itself, only the earliest priority should be taken into account for the calculation of equivalents.	

Figure 6. Examples of families and equivalent groups within them
Based on the application of rules 1-4 altogether

Example 1 (the equivalent group is smaller than the extended family)		In this family structure, with two earliest priorities, the two subsequent filings are equivalents between them (rule 1), because they have exactly the same priorities, but they do not form an equivalent group with their priorities, because they are not equivalent to each other (rule 3).
Example 2 (the equivalent group is equal to the extended family)		In this family structure, with one earliest priority, the two immediate subsequent filings are equivalents between them (rule 1) and to their priority (rule 2). The subsequent filing that claims priority in one of them will be equivalent too because it will "inherit" its priority (rule 4).

As can be seen from the examples in Figure 6, the application of the rules proposed implies that all the extended families whose structures are set out graphically in Table A5 in the Annex are also equivalent groups (all family members are equivalent to each other), except for structures 10, 14, 21 and 31, which are also those not having the simple structure of "one parent and its direct children".

To finalise, it is worth noting that the application of these rules should take into account the nature of the patent linkages used. Technical relations may distort the structure of families, as the concept of priority may not have the same meaning for them as for the rest of patent linkages, which in contrast to them are all declared by the applicant rather than identified by experts or patent examiners. The fact that the "direction of the arrow" in the family relation may not always be clear, leading to circular relations between two applications, is not so surprising since they are by definition applications with similar technical content and no declared priority.

8.   Conclusion

We started this work with three questions in mind: What are patent families? What is the impact of adopting one definition or another? Are some definitions of patent families better suited than others for certain uses in statistical and economic analysis?

In general terms, a patent family is usually understood as the set of all patents filed with the objective to protect the same invention. However, some family definitions are broader than others, and in some cases different family members may not seek protection for exactly the same inventions. We have tried to shed light onto the different methodologies underlying the most commonly used family definitions and provide guidance to replicate them using raw data from PATSTAT. In addition, we have shown that possible differences in family counts (and thus, family statistics) using different existing family definitions would only become apparent when the structure of a patent family is complex, which was only the case in about 25% of all families with priorities in the 1990s.

In fact, 75% of all the INPADOC extended patent families with earliest priorities between 1991 and 1999 have a very simple structure: one single parent (priority) and its direct children (subsequent filings). This result is important as it shows that most definitions will give the same results for 75% of the families. In particular, equivalents, single-priority families and extended patent families will effectively provide the same results for families characterised by one single parent and direct subsequent filings. It may however

be the case that more complex families (within the remaining 25%) are the ones that matter most for economic analysis after all, or those more highly correlated with patent value. Recent research by van Zeebroeck and van Pottelsberghe (2008) has pointed at the high correlation between patent value indicators and the complexity of filing and drafting strategies of applicants at the European level, so there are already indications that it might be so. Further research will address this issue.

The third question, what family definitions are better suited for what uses, is difficult to answer. Among the families that uniquely rely on priority links (*i.e.* excluding families built using expert criteria in addition to priority links), equivalents would be the preferred definition when trying to identify the legal duplicates of an application in other offices. Nevertheless, further research on what it actually means to protect "the same invention" would be useful as in 75% of the cases analysed (families with simple structures) one extended family corresponds to exactly one set of equivalents, so both definitions can be used interchangeably. For the families with more complex structures (the remaining 25% for the period analysed), extended families could in turn be the most useful definition to analyse patent applicant strategies and protection of inventions having a common origin and, in principle, the same applicants.

The extended families definition will also be the one best suited to create new indicators by imposing filters *ex-post* given their breadth, and also because ready-made tables of INPADOC extended families are already provided with PATSTAT. We have compared the total number of INPADOC extended families with those resulting from imposing filters such as non-domestic, transnational and triadic, but many others can be thought of. For instance, filters might be imposed on regions of origin and regions of destination (along the lines of the statistics provided in the Trilateral co-operation website and by WIPO). As regards the analysis of patent strategies, extended families could also be the best resource to analyse time lapsed between first priority and latest filing date within extended families, the span of technology fields included in a family, or even possible changes of assignees across countries and over time within families. Single-priority based families would instead be probably best fitted to forecast filing flows within and across patent offices, as a single subsequent filing would belong to different single-priority families in case it claims multiple priorities.

In sum, both research "on" patent families and research "with" patent families seem to have a long future on an increasing range of economic and statistical studies. A better understanding of their architecture and underlying methodologies would certainly be useful to improve upcoming analyses and facilitate the replicability of results. This study aims to contribute to that.

REFERENCES

Adams, S. R. (2006), *Information Sources in Patents*, 2nd edition, K.G. Saur, Munich.

Deng, Y. (2007), "Private Value of European Patents", *European Economic Review* 51, pp.1785-1812.

Dernis, H. (2007), "Nowcasting Patent Indicators", OECD Science, Technology and Industry Working Paper 2007/3, Directorate for Science, Technology and Industry, OECD, Paris, www.oecd.org/sti/working-papers.

Dernis, H., D. Guellec 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 Science, Technology and Industry Working Paper 2004/2, Directorate for Science, Technology and Industry, OECD, Paris, www.oecd.org/sti/working-papers.

Dernis, H. (2008), "OECD Triadic Patent Families", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

Dernis, H. and P. Hingley (2008), "Variations on the Theme of the Paris Convention", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

Elliott, B. (1997), "European Patent Office, Supply of Patent Family Data from DOCDB for Statistical Analysis by the Controlling Office, Ref. FAMSTATB", internal EPO report.

European Patent Office (2008), PATSTAT Data Catalog September 2008.

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.

Fortune, E. (2008), "Patent Families Analysis", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

Frietsch, R. and U. Schmoch (2010), "Transnational Patents and International Markets", *Scientometrics* 82 (1), pp.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., G. Münt 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, United Kingdom.

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., F. Narin, F.M. Scherer 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, 23-24 October 2006, available at: http://academy.epo.org/schedule/2006/ac04/_pdf/monday/Harhoff.pdf

Harhoff, D. (2008), "Thoughts on Patent Families, Equivalents and Priorities", presentation at the OECD/EPO Patent Families Workshop, Vienna, 20-21 November 2008.

Henderson, R. and I. Cockburn (1993), "Scale, Scope and Spillovers: the Determinants of Research Productivity in Ethical Drug Discovery", Working Paper WP 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. (2008), "Patent Families Defined as Priority Forming Filings and their Descendents", presentation at the OECD/EPO Patent Families Workshop, Vienna, 20-21 November 2008.

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., A. Pakes 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.

Martinez, C. and S. Maraut (2008), "Identifying Different Kinds of Patent Families in PATSTAT", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

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.

Nishimura. Y. (2008), "Prediction of R&D Project Size of Firms from Patent Family Data: Evidence from Japan Inventors Survey", presented at the EPO/OECD Patent Families Workshop, 20-21 November 2008, Vienna.

OECD (1994), *Patent Manual*, OECD, Paris.

OECD (2009), *Patent Statistics Manual*, OECD, Paris.

Paris, P. (2008), "Patent Families. Is There an Obvious Concept Covering All Aspects?", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

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.

Raffo, J. and S. Lhuillery (2008), "We Can Choose our Friends, but Can We Pick our Patent Families?" presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

Rollinson, J. (2008), "Extended Priority INPADOC Patent Family", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

Simmons, E.S. (2009), "'Black Sheep' in the Patent Family", World Patent Information, 31, pp. 11-18.

Torre, M. de la (2008), "Study on Patent families from PATSTAT Data", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

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.

Versloot-Spoelstra, F. (2008), "DOCDB Simple Patent Family", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

Webb, C., H. Dernis, D. Harhoff and K. Hoisl (2005), "Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks", OECD Science, Technology and Industry Working Paper 2005/9, Directorate for Science, Technology and Industry, OECD, Paris, www.oecd.org/sti/working-papers.

WIPO (2008), *World Patent Report: A Statistical Review*, WIPO, Geneva.

Zhou, H. (2008), "WIPO Patent Family Database", presentation at the EPO/OECD Patent Families Workshop, Vienna, 20-21 November 2008.

ANNEX I. GENERAL OVERVIEW OF LINKAGES BETWEEN PATENT FILINGS

Paris Convention priorities

The priority right for the international extension of patent protection was created by the 1883 Paris Convention for the protection of industrial property, administered by WIPO. Article 4 of the Convention states a right of priority of twelve months for the extension of patent protection to other countries in the following terms: [45]

> "4A. (1) Any person who has duly filed an application for a patent, or for the registration of a utility model, or of an industrial design, or of a trademark, in one of the countries of the Union, or his successor in title, shall enjoy, **for the purpose of filing in the other countries**, a right of priority during the periods hereinafter fixed. (2) Any filing that is **equivalent to a regular national filing** under the domestic legislation of any country of the Union or under bilateral or multilateral treaties concluded between countries of the Union shall be recognized as giving rise to the right of priority. (3) **By a regular national filing is meant any filing that is adequate to establish the date** on which the application was filed in the country concerned, whatever may be the subsequent fate of the application."[46]

> "4C. (1) The periods of priority referred to above shall be **twelve months for patents** and utility models, and six months for industrial designs and trademarks. (2) These periods shall start from the date of filing of the first application; the day of filing shall not be included in the period. (3) If the last day of the period is an official holiday, or a day when the Office is not open for the filing of applications in the country where protection is claimed, the period shall be extended until the first following working day." [47]

Therefore the right of priority, as set out in the Paris Convention, necessarily relates to cases where the first filing has been done in a different country to the subsequent filing, and has to be explicitly claimed by the applicant.

The question of whether the invention in a first filing has to be the exactly "the same" or not as the one protected by subsequent filings claiming its priority, is marginally addressed in Article 4 C (4), where it is stated that the priority date would be that of domestic subsequent filings provided they cover "the same subject" and the first filing has been abandoned, withdrawn or refused before being published. In addition, the European Patent Convention, in Article 87 (1), uses the term invention instead of subject matter when it refers to Paris Convention subsequent filings[48]:

> "87. (1) A person who has duly filed in or for any State party to the Paris Convention for the Protection of Industrial Property, an application for a patent or for the registration of a utility model or for a utility certificate or for an inventor's certificate, or his successors in title, shall enjoy, for the purpose of filing a European patent application in respect of **the same invention**, a right of priority during a period of twelve months from the date of filing of the first application." [49]

---

[45]   http://www.wipo.int/treaties/en/ip/paris/trtdocs_wo020.html.

[46]   Emphasis added.

[47]   Emphasis added.

[48]   The Opinion of the Enlarged Board of Appeal in G2/98 on this point is as follows: "The requirement for claiming priority of "the same invention", referred to in Article 87(1) EPC, means that priority of a previous application in respect of a claim in a European patent application in accordance with Article 88 EPC is to be acknowledged only if the skilled person can derive the subject-matter of the claim directly and unambiguously, using common general knowledge, from the previous application as a whole". Case number G0002/98, decision of 31 may 2001, available at http://legal.european-patent-office.org/dg3/biblio/g980002e.htm.

[49]   Emphasis added.

This requirement is important as it implies that, in principle, all family members directly linked only through Paris Convention priorities have to protect exactly the same invention. However, when family definitions rely on other types of patent relations and take complex and broad structures, the question of whether it is possible to find different inventions within the same patent family is more difficult to answer. As noted by Simmons (2009), "most family members claim priority under the Paris Convention for Protection of Industrial Property or bilateral treaties and are equivalent, but not necessarily identical in their claim scope or disclosures".

Technical similarities

There is another type of international relation between patents that, in contrast to the previous ones, is not claimed by applicants. Technical similarities are identified by patent examiners and analysts based on the similarity of the inventions described, the inventors and applicants mentioned in two given applications. These may refer to patent applications in countries that have not ratified the Paris Convention, or to filings that have simply exceeded the 12 months delay to benefit from the right of priority. In that case, artificial priorities are added to create the formerly inexistent patent linkages.

Technical similarities reported in EPO databases (*e.g.* esp@cenet, PATSTAT) receive the name of "technical relations". In turn, analysts at Thomson Reuters call them "non-convention" or "intellectual" priorities and take them into account for the calculations of Derwent patent families. Equivalency is determined through manual checking of inventors, subject matter, etc. They are patents with the same technical content but not claiming the same priority.[50]

Domestic priorities

Relations between first filings and subsequent filings may also be the result of domestic procedures, giving rise to "domestic priorities". The most common ones are continuations, continuations in part, provisional patent filings (all of them only available at the United States Patent and trademark Office), and divisionals (available in many patent offices),[51] as included in Article 4 of the Paris Convention as follows:

> "4G. (1) If the examination reveals that an application for a patent contains more than one invention, the applicant may divide the application into a certain number of **divisional applications** and preserve as the date of each the date of the initial application and the benefit of the right of priority, if any. (2) The applicant may also, on his own initiative, divide a patent application and preserve as the date of each divisional application the date of the initial application and the benefit of the right of priority, if any. Each country of the Union shall have the right to determine the conditions under which such division shall be authorized."[52]

USPTO continuations and continuation in part (CIP) applications claim the domestic priority of an earlier non-provisional application and are filed with the objective to add new claims to the invention disclosed in the original application, in the case of continuations, or to disclose new subject matter in the case of CIP. In turn, USPTO provisional applications (available since June 1995) allow an inventor to disclose its invention one year before doing a regular filing and claim the priority date of the provisional

---

[50]   http://www.piug.org/patfam.php.

[51]   The effective filing dates of continuations and divisionals are those of their parent applications, except for the new matter added in CIPs.

[52]   Emphasis added.

filing.[53] The date of priority for international extensions according to the Paris Convention would be the date of filing of the provisional application, unless it was not published.

Article 4 of the Paris Convention treats the relation between domestic priorities and Paris Convention priorities in Article 4C (4) as follows.

> "4C. (4) A subsequent application concerning **the same subject** as a previous first application within the meaning of paragraph (2), above, **filed in the same country** of the Union shall be considered as the first application, of which the filing date shall be the starting point of the period of priority, if, at the time of filing the subsequent application, the said previous application has been **withdrawn, abandoned, or refused, without having been laid open** to public inspection and without leaving any rights outstanding, and if it has not yet served as a basis for claiming a right of priority. The previous application may not thereafter serve as a basis for claiming a right of priority."[54]

PCT regional/national phase entries

International patent protection has been largely facilitated by the introduction of the Patent Cooperation Treaty (PCT), which was signed in 1970 and entered into force in 1978. The PCT, administered by the World Intellectual Property Organisation (WIPO), provides the possibility to seek patent protection in a large number of countries by filing a single international application (PCT application) with a single patent office (receiving office) and then entering the national stage in the desired countries at a later date. In that sense, a PCT application can be considered an option for future applications to patent offices around the world. The PCT application starts with the filing of an international application either at the national (or regional) patent office or with the international bureau of WIPO. This has to be done in the 12-month period following the priority filing, but it can be done immediately as a priority filing itself. The applicant must be a national or resident of one of the PCT signatory states. Since January 2004, a PCT application automatically includes all PCT signatory states as designated states. At 30 months from the priority date, the international phase ends and the international application enters the national or regional phase where the applicant wants to actually apply for a patent (OECD, 2009).[55]

Validations of EPO grants in EPC member states

The European Patent Office, which started to operate in 1978, offers a harmonised application and examination path for applicants seeking patent protection in signatory states to the EPC. If the application passes the examination process at EPO successfully, the applicant has the right to validate his/her right in the EPC member countries of his/her choice, provided they were designated in the application.

As of April 2009 there is automatic designation of all EPC countries in EPO applications, but two previous changes made a "de facto" automatic designation system at EPO earlier than that. First, since July 1999, a flat designation fee was imposed for designation of seven or more countries. Second, since January 2004, the Euro-PCT applications have to abide to the automatic designation of all PCT countries (which includes EPC countries) in Euro-PCT applications (OECD, 2009).

---

[53]   US provisional applications can be limited to a written description of the invention (no need to have formal patent claims), are not examined and can be marked as "patent pending". They are US national applications filed in the USPTO under 35 U.S.C. 111(b).

[54]   Emphasis added.

[55]   See an overview of the PCT system at http://www.wipo.int/pct/en/activity/pct_2007.html#P351_20133.

ANNEX II: PATENT LINKAGES AS REPORTED IN PATSTAT

Table APPLN_PRIOR

In contrast to what would be expected, not all the priority relations included in the PATSTAT table of Paris Convention priorities (APPLN_PRIOR) correspond to international relations. In fact, 57% of the applications filed between 1995 and 2006 are claimed as a domestic priority (42% of them within the United States and 23% within Japan). Also, 62% of all the US domestic priority applications reported in table APPLN_PRIOR are of application kind "P", which indicates "provisional application". Relations between US provisional applications and subsequent non-provisional applications are thus not to be found in the domestic continuations table in PATSTAT (APPLN_CONTN), but in the Paris Convention priorities table (APPLN_PRIOR). We would nevertheless continue to refer to the table APPLN_PRIOR as the one providing information about Paris Convention priorities, to be coherent with the PATSTAT Data Catalog.

Table APPLN_CONTN

As shown in Table A1 below, the United States is by far the country with the highest number of applications claiming domestic priorities, with 76% of the applications included in the continuations table in PATSTAT September 2008, followed by the patent office of Germany (8%), the European Patent Office (6%), Switzerland (4%), Australia (2%) and Canada and the United Kingdom (with 1% each). PATSTAT breaks down domestic continuations into ten different types, with the large majority of them being US continuations, US divisionals and US continuations in part. It is worth noting, however, that PATSTAT does not have currently information on divisionals from JPO.[56]

---

[56]   JPO does not incorporate data on divisionals in the data provided to EPO. I thank Davide Lingua for this information, as well as Tomoya Yanagisawa and Yun Suzuki for their help with understanding the JPO database.

Table A1. Types of domestic priorities in PATSTAT continuations table
Included in PATSTAT September 2008 Table TLS216_APPLN_CONTN

Type	Continuation Type	Description	Patent Office where each type of continuation is available*	Total number of continuations broken down by type	Share of subsequent filings in continuation relations broken down by patent office
Continuation	CON	Abandoned application claimed for a continuation	US	506 259	99% US
		Prior application claimed for a continuation	US, PH		
Divisional	DIV	Abandoned application claimed for a division	US	436 775	70% US, 14% EP, 4% AU, 3% CA, 3% GB
		Prior application claimed for a division	AU, BA, CA, CH, CS, CZ, DE, DK, EP, ES, FI, FR, GB, HK, HU, IE, IL, IN, KR, LU, LV, NL, NOP, NZ, PH, US, YU		
Continuation in part	CIP	Abandoned application claimed for a continuation in part	US	427 937	99% US
		Prior application claimed for a continuation in part	US, PH, CA		
Internal priority	INN	Domestic priority claimed for patent	DE	89 948	98% DE, 2% JP
		Domestic priority claimed for utility model	DE		
		Domestic priority	JP, RU, SU, YU		
Addition	ADD	Prior application claimed for an addition	AU, CH, IE, IL, IN, KR, NZ, PL, TW	32 689	98% CH, 1% IL, 1% AU
Change of IPR type	--	Cited application changed from patent to utility	AT, JP, KR, MX	6 573	22% HR, 14% CA, 11% EP, 10% LV, 9% LT
		Cited application changed from utility to patent	AT, DE, JP, KR, MX		
Reissue	REI	Request for re-examination number	US	5 716	100% US
		Claimed application is original reissue serial number	US		
Cognate	CGT	Cognate application	IE, IN, NO, NZ	212	87% NZ, 13% IE
Supplementary disclosure	SUP	Claimed application is a supplementary disclosure	CA	79	100% CA
Substitute	SBS	Prior application claimed for a substitute	US	24	100% US

*The list of acronyms identifying national patent offices is available at www.wipo.org

Source: Own elaboration based on PATSTAT September 2008 Data Catalog (page 90) and extractions from Table TLS216_APPLN_CONTN (European Patent Office, 2008).

Table TECH_REL

Technical similarities are not reported by applicants in patent documents, but identified *ex post* by examiners in databases by comparing technical content and bibliographic information. The Data Catalog of the September 2008 version of PATSTAT says the following about technical relations (European Patent Office, 2008):

- "The technical relations are entered when detected by examiners or the EPO bibliographic data experts and when no other priority-like relation exists between the applications. There can however be no guarantee of completeness. This relation is also not published by Patent Offices" (p. 18).

- "Technically related documents are those patent documents whose technical content has been identified within the EPO as being considered equivalent. This relation is identified in the EPO master documentation database DOCDB by setting the indicator LMI=T for 'Technical'. The "T" indicator has allowed extraction of most of the technical relations in table TLS205_TECH_REL. However, due to the manual intervention needed to create technical relations, it is known that a certain number of technical relations do not have the indicator set to "T", thus appearing in PATSTAT as a PARIS convention priority" (p.88).

Applications with technical relations can be linkages with other domestic or international applications and are mostly found in France (38%), United Kingdom (21%), Germany (14%) and United States (11%) (see Table A2 below)[57]. As regards the lapse between the application filing date of the application and its "technical relation", 45% of the applications have been filed within the same year, 45% one year later and 3% two years later.

Table A2. Top patent offices for applications with identified "technical similarities"
Included in PATSTAT September 2008 Table APPLN_TECH_REL

Country breakdown of all applications claiming the technical relation	% domestic % foreign	Country breakdown of foreign technical relations
France (38% of all)	64% within France 36% non-French	34% Germany 18% United States 16% United Kingdom 8% Switzerland
United Kingdom (21% of all)	14% within UK 86% non-UK	34% United States 26% Germany 14% France 6% Switzerland
Germany (14% of all)	13% within Germany 87% non-German	27% United States 16% France 16% United Kingdom 10% Switzerland
United States (11% of all)	3% United States 97% non-United States	29% Germany 22% United Kingdom 10% France 8% Canada

Source: Author's elaboration based on PATSTAT September 2008.

---

[57] All the other patent offices have less than 5%: Switzerland 5%, Belgium 2%, Netherlands 2%, Austria 2%, Canada 1%, Japan 1%, European Patent Office 1%, Australia 1%, etc.

Table APPLN

Information on PCT regional/national phase entries included in PATSTAT is presented in table APPLN as the relation between appln_id and internat_appln_id, where the former would be the child and the latter the parent application in a PCT national phase entry type of linkage. It should however be stressed that, for the moment, it is limited to national phase entries of PCT applications that have entered the regional phase at EPO.[58]

The fact that PATSTAT only includes information on PCT filings that have entered EPO regional phase is an important caveat that needs to be taken into account when using PCT links to build families. Information on PCT regional/national phase entries from PATSTAT, albeit incomplete, is quite relevant and informative of patent linkages not available elsewhere, and in any case different from those included in PATSTAT tables reporting Paris Convention priorities, domestic continuations or technical relations.[59]

EPO post grant information: validations in EPC designated member states

PATSTAT does not include post-grant information about filings at EPO or any other patent office. Nevertheless, it sometimes reports national validations of EPO grants as national equivalents of EPO applications (*i.e.* sharing exactly the same priorities as the EPO application). In these cases, it is possible to include them as additional patent family members, but it may not always be easy to distinguish them from original national filings. In some cases, the national patent office of the EPC member state gives EP validations a new application number and it is not possible to identify them without additional information from national patent databases on EP validations. In other cases, they are identified in PATSTAT by appln_kind 'T' (from translation) and have the same application number as the EP application they correspond to, which makes easier to identify them.

---

[58] The Australian Patent Office publishes all PCT filings, whether they enter the national phase in Australia or not (Paris, 2008; Simmons, 2009). This is one peculiarity of PCT publications in Australia, but signals the need to be careful when interpreting PCT information in PATSTAT. Simmons (2009) calls them "phantom family members".

[59] Only a few of the 77 962 internat_appln_id with filing year 2006 are also identified as "parents" in other types of patent relations: 410 are also reported as being Paris Convention priorities, 13 as parents of domestic continuations and 15 as prior technical relations. At present, INPADOC extended patent families do not use PCT national phase as reported in PATSTAT as patent linkages to form families, whereas OECD triadic patent families do.

ANNEX III: CYCLICAL FAMILIES IN PATSTAT

Our family calculations based on PATSTAT data revealed that the relations between applications are sometimes circular, although not very frequently. Based on this, we distinguished between normal or cyclical families. A "normal" family is a family having one or several originating applications or family seeds, that is, applications which do not depend on any other application but on which other applications depend (*i.e.* "Parents who are not Children"). In turn, a "cyclical family" is a family in which it is not possible to identify a family seed because each application declares having a relation with at least another application within the family (*i.e.* "All family members have parents").[60]

One possible explanation for the existence of cyclical families is the occurrence of mistakes in the recording of family relations in the original documents or in the databases. In fact, individual PATSTAT tables on patent linkages have some circularity already (before consolidation into families through interrelation with other types of patent linkages). In the September 2008 version of PATSTAT we find that 51 006 applications in priority relations, 568 applications in domestic continuations and 114 955 applications in technical relations are involved in circular relations within their own tables.[61] Therefore, the large majority of "circular relations" found in PATSTAT come from the artificial priorities reported as technical similarities. Actually, the large share of circularity in the table of technical relations (around 6% of all applications in the table) raises the question of whether the "priority" concept has the same meaning for technical similarities as for the rest of patent linkages. If, as seems logic, by definition a technical similarity would tend to satisfy transitivity, this indicates that this type of linkage may need to be treated differently, it is also produced differently.

To facilitate the understanding of the concept of cyclical families, a graphical representation of two examples of normal and cyclical families is set out below in Figure A1 below.

Figure A1. Normal and cyclical families
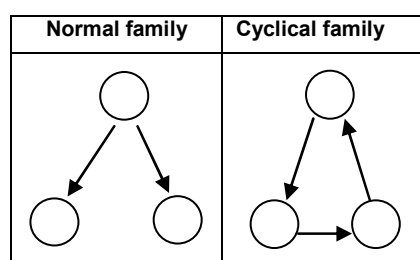
One implication of the distinction between normal and cyclical families is the distribution of families over time. If we order families by the filing date of their earliest priorities, in normal families such a date will correspond to the filing date of the application from which the rest of family members are derived

---

[60]   There may also be cyclical relations within normal families, but we do not flag them separately insofar as it is possible to identify a family seed, which defined the family as "normal".

[61]   These figures are the result of counting the number of different applications (appln_id) in tables APPLN_PRIOR, APPLN_CONTN, TECH_REL and APPLN for which the following happens: application A declares application B as its priority, and application B declares application A as its priority in another record.

(*i.e.* the family seed). In the case of cyclical families, this concept does not apply because there is no family seed, so that cyclical families would be distributed over time according to the earliest filing date among all the filing dates of their members.

A comparison of the total number of cyclical families across different sources of family relations with earliest priority filings comprised between 1991 and 1999 shows that most circular relations come from technical similarities, as expected (Table A3 below). We find only 519 and 999 cyclical families using respectively sources 1 and 2 for family relations, but the count rises to 21 326 when technical similarities are added (source 3). Only a few additional cyclical families are found when PCT links are also used (21 609 cyclical families with source 4).

Table A3. Counts of cyclical families and applications in them, by source of family relations
Earliest priorities 1991-1999

	Source 1 families		Source 2 families		Source 3 families (INPADOC extended)		Source 4 families	
	(Paris Convention)		(Paris Convention + Domestic continuations)		(Paris Convention + Domestic continuations + Technical similarities)		(Paris Convention + Domestic continuations + Technical similarities + PCT national phase entries)	
	#families	#members	#families	#members	#families	#members	#families	#members
1991	68	518	85	584	1 545	8 012	1 599	8 400
1992	92	488	129	696	2 122	9 947	2 155	10 242
1993	63	326	118	664	2 145	9 581	2 175	9 828
1994	57	254	135	575	2 271	10 247	2 297	10 476
1995	46	218	116	575	2 444	10 943	2 465	11 109
1996	40	183	98	539	2 854	12 553	2 901	12 908
1997	39	272	91	599	2 901	12 601	2 943	12 921
1998	38	196	103	523	2 528	10 519	2 540	10 630
1999	76	312	124	576	2 516	9 660	2 534	9 796
1991-1999	519	2 767	999	5 331	21 326	94 063	21 609	96 310

Note: Source 1: APPLN_PRIOR; Source 2: APPLN_PRIOR + APPLN_CONTN; Source 3: APPLN_PRIOR + APPLN_CONTN + TECH_REL; Source 4: APPLN_PRIOR + APPLN_CONTN + TECH_REL + PCT links (relations between appln_id and internat_appln_id in APPLN). Source 3 families are INPADOC families, reported in ready-made family PATSTAT table TLS_219_INPADOC_FAM.

Source: Author's calculations based on PATSTAT September 2008.

ANNEX IV: ADDITIONAL TABLES

Table A4. Total number of INPADOC extended families and filtered subsets, priority country
Earliest priority year 1999

	All	Non-domestic	% All	Transnational	% All	Triadic	% All
United States	67 554	50 535	75%	44 376	66%	15 824	23%
Japan	50 695	41 110	81%	19 645	39%	12 334	24%
Germany	28 727	24 897	87%	21 972	76%	5 595	19%
United Kingdom	9 079	7 734	85%	6 864	76%	1 596	18%
France	7 659	7 335	96%	6 857	90%	1 995	26%
Korea	6 882	5 308	77%	1 611	23%	537	8%
EPO	6 014	5 858	97%	6 014	100%	2 206	37%
Sweden	4 291	2 518	59%	2 382	56%	718	17%
Australia	3 452	1 376	40%	1 460	42%	266	8%
Italy	3 435	3 403	99%	2 968	86%	401	12%
TOTAL	205 222	165 128	80%	124 985	61%	43 394	21%

Source: Author's calculations based on PATSTAT September 2008.

Table A5. Top 10 family structures by source of family relations
Excluding singletons. Earliest priorities 1991-1999
Structures are numbered according to the full ranking of normal and cyclical families built using Paris Convention priorities only (source 1)
and having earliest priorities 1991-1999

Source 1 (Paris Convention)				Source 2 (Paris Convention + Domestic Continuations)				Source 3: INPADOC FAMILY (Paris Convention + Domestic Continuations + Technical Similarities)				Source 4 (Paris Convention + Domestic Continuations + Technical Similarities + PCT nacional phase)			
Structure Number ID	*Structure with single priority and direct subsequent filings*	*Number of applications in family structure*	*% all families*	*Structure Number ID*	*Structure with single priority and direct subsequent filings*	*Number of applications in family structure*	*% all families*	*Structure Number ID*	*Structure with single priority and direct subsequent filings*	*Number of applications in family structure*	*% all families*	*Structure Number ID*	*Structure with single priority and direct subsequent filings*	*Number of applications in family structure*	*% all families*
1	yes	2	29.0	**1**	yes	2	30.4	**1**	yes	2	29.7	**1**	yes	2	29.6
2	yes	3	13.9	**2**	yes	3	12.2	**2**	yes	3	11.9	**2**	yes	3	8.4
3	yes	4	8.9	**3**	yes	4	8.0	**3**	yes	4	7.8	**3**	yes	4	5.5
4	yes	5	8.2	**4**	yes	5	7.0	**4**	yes	5	6.9	**14**	no	3	5.2
5	yes	6	6.3	**5**	yes	6	5.3	**5**	yes	6	5.2	**4**	yes	5	4.9
6	yes	7	4.6	**6**	yes	7	3.7	**6**	yes	7	3.7	**5**	yes	6	3.1
7	yes	8	3.3	**7**	yes	8	2.6	**7**	yes	8	2.6	**10**	no	3	1.9
8	yes	9	2.4	**10**	no	3	2.0	**10**	no	3	1.9	**6**	yes	7	1.8
9	yes	10	1.8	**8**	yes	9	1.9	**8**	yes	9	1.8	**21**	no	4	1.6
10	no	3	1.6	**14**	no	3	1.6	**14**	no	3	1.6	**31**	no	4	1.5
Top 10	-	-	80	**Top 10**	-	-	75	**Top 10**	-	-	73	**Top 10**	-	-	63

Note: Structures with Number ID 11, 14, 21 and 31 are highlighted in the table because they are not included in the top 10 of families with source 1, but appear in the top 10 of families with other sources.

Source: Author's calculations based on PATSTAT September 2008.

Table A5bis. Top 25 family structures by source of family relations
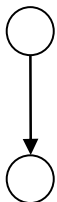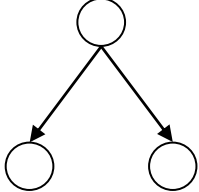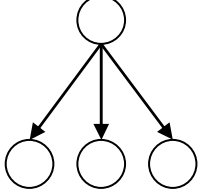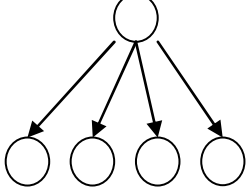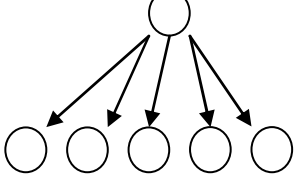Excluding singletons. Earliest priorities 1991-1999
Structures are numbered according to the full ranking of normal and cyclical families built using on Paris Convention priorities only (source 1)
and having earliest priorities 1991-1999

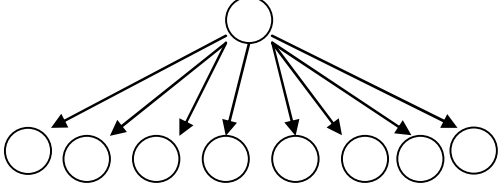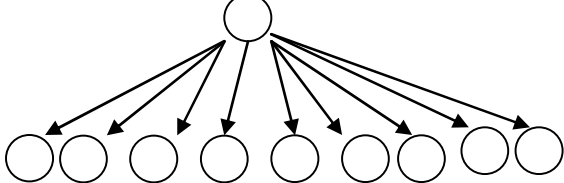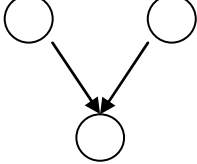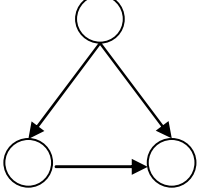| Source 1 (Paris Convention)				Source 2 (Paris Convention + Domestic Continuations)				Source 3: INPADOC FAMILY (Paris Convention + Domestic Continuations + Technical Similarities)				Source 4 (Paris Convention + Domestic Continuations + Technical Similarities + PCT nacional phase)			
Structure Number ID	Structure with single priority and direct subsequent filings	Number of applications in family structure	% all families	Structure Number ID	Structure with single priority and direct subsequent filings	Number of applications in family structure	% all families	Structure Number ID	Structure with single priority and direct subsequent filings	Number of applications in family structure	% all families	Structure Number ID	Structure with single priority and direct subsequent filings	Number of applications in family structure	% all families
11	yes	11	1.3	9	yes	10	1.3	9	yes	10	1.3	29	no	5	1.2
12	yes	12	1.0	11	yes	11	1.0	11	yes	11	1.0	26	no	6	1.2
13	no	4	0.8	12	yes	12	0.7	12	yes	12	0.7	7	yes	8	1.0
14	no	3	0.8	13	no	4	0.6	135	no	2	0.6	88	no	7	0.9
15	yes	13	0.8	15	yes	13	0.5	13	no	4	0.6	74	no	5	0.8
16	yes	14	0.6	31	no	4	0.5	15	yes	13	0.5	91	no	6	0.8
17	no	6	0.5	24	no	4	0.5	31	no	4	0.5	32	no	7	0.8
18	no	5	0.5	21	no	4	0.4	21	no	4	0.5	103	no	8	0.7
19	yes	15	0.4	16	yes	14	0.4	24	no	4	0.5	135	no	2	0.6
20	no	7	0.4	39	no	5	0.4	16	yes	14	0.4	8	yes	9	0.6
21	no	4	0.4	41	no	6	0.4	39	no	5	0.4	116	no	9	0.5
22	yes	16	0.3	43	no	7	0.4	41	no	6	0.4	39	no	5	0.4
23	no	8	0.3	27	no	4	0.3	43	no	7	0.3	191	no	8	0.4
24	no	4	0.3	17	no	6	0.3	27	no	4	0.3	9	yes	10	0.4
25	yes	17	0.3	18	no	5	0.3	17	no	6	0.3	65	no	4	0.4
Top 25	-	-	89	**Top 25**	-	-	83	**Top 25**	-	-	81	**Top 25**	-	-	74
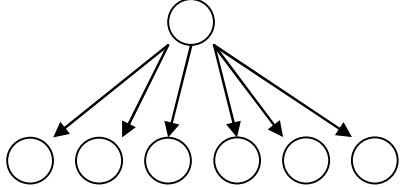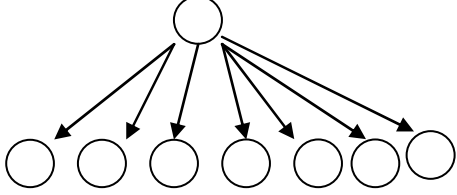
Note: Family Structure ID 135 (highlighted in the table), in position top 14 for family source 3 and top 19 for family source 4, is the only cyclical family within the top 25 of the sources considered. It has only two members but cannot be identified with the simple pattern of one single priority and direct subsequent filings, because no application can be considered as the "earliest priority" due to circular relations.

Source: Author's calculations based on PATSTAT September 2008.

Table A6. Graphical representation of most frequent structures of families with priorities in 1991-1999, excluding singletons

Structure Number ID	Number of earliest priorities	Number of applications	Structure with single priority and direct subsequent filings only	Family structure
1	1	2	**Yes**	
2	1	3	**Yes**	
3	1	4	**Yes**	
4	1	5	**Yes**	
5	1	6	**Yes**	

**Table A6. Graphical representation of most frequent structures of families with
priorities in 1991-1999, excluding singletons (contd.)**

Structure Number ID	Number of earliest priorities	Number of applications	Structure with single priority and direct subsequent filings only	Family structure
6	1	7	Yes	
7	1	8	Yes	
8	1	9	Yes	
9	1	10	Yes	
10	2	3	No	
14	1	3	No	

48

Table A6. Graphical representation of most frequent structures of families with priorities in 1991-1999, excluding singletons (contd.)

Structure Number ID	Number of earliest priorities	Number of applications	Structure with single priority and direct subsequent filings only	Family structure
21	1	4	**No**	
31	1	4	**No**	

Note: This table presents a graphical representation of the family structures that are positioned in the top 10 most frequent families for all family sources considered. See top 25 ranking in Table A4, with top 10 highlighted. Structures are numbered according to the full ranking of normal and cyclical families built using only Paris Convention priorities (source 1) and having earliest priorities 1991-1999, as listed in Table A4.

Source: Author's elaboration based on calculations using data from PATSTAT September 2008.