



OECD Science, Technology and Industry Working Papers  
2005/09

Analysing European  
and International Patent  
Citations: A Set of EPO  
Patent Database Building  
Blocks

**Colin Webb,  
Hélène Dernis,  
Dietmar Harhoff,  
Karin Hoisl**

<https://dx.doi.org/10.1787/883002633010>

**Unclassified**

**DSTI/DOC(2005)9**



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

**18-Oct-2005**

**English text only**

**DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY**

**DSTI/DOC(2005)9**  
**Unclassified**

**ANALYSING EUROPEAN AND INTERNATIONAL PATENT CITATIONS:  
A SET OF EPO PATENT DATABASE BUILDING BLOCKS**

**STI WORKING PAPER 2005/9**  
**Statistical Analysis of Science, Technology and Industry**

**Colin Webb and Hélène Dernis, OECD; Dietmar Harhoff and Karin Hoisl, LMU**

**JT00192168**

Document complet disponible sur OLIS dans son format d'origine  
Complete document available on OLIS in its original format

**English text only**

### **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.

---

*<http://www.oecd.org/sti/working-papers>*

---

**Copyright OECD, 2005**

**Applications for permission to reproduce or translate all or part of this material should be made to:  
OECD Publications, 2 rue André-Pascal, 75775 Paris, Cedex 16, France.**

**ANALYSING EUROPEAN AND INTERNATIONAL PATENT CITATIONS:  
A SET OF EPO PATENT DATABASE BUILDING BLOCKS\***

Colin Webb\*\* and H el ene Dernis, OECD; Dietmar Harhoff and Karin Hoisl, LMU\*\*\*

**ABSTRACT**

This paper presents a preliminary set of European and international citation data tables designed to enable researchers to become familiar with the subject and undertake a range of analyses. It addresses analytical and statistical issues such as why citations from international searches and European searches should be combined; the calculation of citation lags; use of International Patent Classification (IPC) codes for grouping patents according to technology; treatment of references to Non-Patent Literature (NPL); and the notion that many publications, from different patenting authorities, covering the same invention can be cited. Differences between US and European citations are also discussed.

The database was built using data kindly provided by the European Patent Office (EPO) and its development has been carried out under the auspices of OECD's Patent Statistics Task Force. The work is an OECD contribution to a wider project to construct a worldwide patent database mainly co-ordinated by EPO and the World Intellectual Property Organisation (WIPO). It has already provided useful input into EPO's current PATSTAT database project to make their raw patent data more accessible for statistical and econometric analyses.

It is assumed that potential users are reasonably familiar with recent citation analysis literature. Possible avenues of research such as knowledge spillovers (geographical or across fields of science), knowledge flows and value of patents are not addressed in this paper but dealt with in a related document, by the same authors, "European Patent Citations – How to count and interpret them?". For the development of citations-based indicators and examples of a wide range of related analyses, reference to the significant body of work based on US citations, by researchers such as Hall, Jaffe and Trajtenberg, is encouraged.

The data tables currently cover all patent applications published by EPO and WIPO, under the Patent Co-operation Treaty (PCT), from their introduction in 1978 up until November 2001. They are available on CDROM, on request from OECD, for research purposes only.

---

\* This paper would not have been possible without expert advice from members of the OECD's Patent Statistics Task Force, in particular comments by Franco Cordera, Dominique Guellec, Robert Heijna and James Rollinson of EPO; and Mosahid Khan, Dirk Pilat and Andrew Wyckoff of OECD. Comments from test users of this database were also greatly appreciated – particular thanks to Paola Criscuolo and Paroma Sanyal for their valuable input.

\*\* Principal contact: colin.webb@oecd.org

\*\*\* University of Munich (Ludwig-Maximilians-Universit at).

## L'ANALYSE DES CITATIONS DE BREVETS EUROPÉENS ET INTERNATIONAUX UNE SÉLECTION DE MODULES DE LA BASE DE DONNÉES DE BREVETS DE L'OEB\*

Colin Webb\*\* et Hélène Dernis, OCDE ; Dietmar Harhoff et Karin Hoisl, LMU\*\*\*

### RÉSUMÉ

Ce document présente une première sélection de données sur les citations européennes et internationales. Ces modules ont été construits afin de permettre aux chercheurs de se familiariser avec les citations et d'engager un certain nombre d'analyses. Plusieurs questions analytiques et statistiques sont abordées, notamment : pourquoi combiner les citations issues des recherches internationales avec celles des recherches européennes ; calculer des délais de citation ; utiliser les codes de la Classification internationale des brevets (CIB) pour regrouper les brevets par technologies ; comment considérer les références à la littérature non-brevet (NPL) ; et comment comptabiliser les citations lorsque plusieurs brevets, publiés par différents offices de brevet, pour la même invention sont cités. Les divergences entre citations américaines et européennes sont également étudiées.

La base de données a été développée sous l'égide du Groupe de travail de l'OCDE sur les statistiques de brevet, avec les données aimablement fournies par l'Office européen des brevets (OEB). Ce travail de l'OCDE participe à un plus vaste projet de construction d'une base de données de brevets mondiale, projet coordonné principalement par l'OEB et l'Organisation mondiale de la propriété intellectuelle (OMPI). Ce projet a offert une bonne contribution au projet de base de données PATSTAT actuellement mené par l'OEB, visant à améliorer l'accessibilité de leurs données brutes pour l'analyse statistique.

Ce document s'adresse à des utilisateurs potentiels des données, qui seraient déjà familiarisés avec la littérature récente sur l'analyse des citations. Les axes de recherche potentiels tels que les retombées (*spillovers*) de la connaissance (géographiques ou entre domaines scientifiques), les flux de connaissance et la valeur des brevets ne sont pas abordés ici mais traités dans un autre papier par les mêmes auteurs : "Les citations de brevets européens - comment les compter et les interpréter?" ("European Patent Citations - How to count and interpret them?"). Pour le développement d'indicateurs basés sur les citations et d'autres exemples d'analyse sur le sujet, il convient de se référer aux travaux menés sur les citations américaines par des chercheurs tels que Hall, Jaffe et Trajtenberg.

Les tables de données couvrent actuellement toutes les demandes de brevets publiées par l'OEB et l'OMPI, via le Traité de coopération en matière de brevets (PCT), depuis leur introduction en 1978 jusqu'en novembre 2001. Elles sont disponibles sur CD-ROM, sur demande auprès de l'OCDE et uniquement pour des fins de recherche.

---

\* Ce document s'appuie en grande partie sur l'expertise des participants au Groupe de travail de l'OCDE sur les statistiques de brevet, notamment sur les commentaires de Franco Cordera, Dominique Guellec, Rob Heijna et James Rollinson de l'OEB ; Mosahid Khan, Dirk Pilat et Andrew Wyckoff de l'OCDE. Les commentaires d'utilisateurs-test de la base de données ont été également appréciés – nous remercions tout particulièrement Paola Criscuolo et Paroma Sanyal.

\*\* Contact principal: colin.webb@oecd.org

\*\*\* Université de Munich (Ludwig-Maximilians-Universität).

## TABLE OF CONTENTS

1. INTRODUCTION .....	6
2. PATENT PROCESSES .....	7
2.1 Brief notes on international (PCT) and European application and search processes .....	7
2.2 Comparing with USPTO citations data .....	7
3. GENERAL DESCRIPTION OF THE DATA SET .....	9
3.1 Organisation .....	9
3.2 The main tables .....	9
3.3 Tables of Cited Patents and Non-Patent Literature (NPL) .....	14
3.4 Link tables .....	15
3.5 Supplementary tables .....	16
3.6 Calculating citation lags .....	17
3.7 Grouping citations by IPC .....	18
4. NEXT STEPS .....	19
4.1 Further enhancements to the data set .....	19
4.2 Next version of the data set .....	19
REFERENCES .....	21
ANNEX 1. EPO AND PCT APPLICATION PROCESSES .....	23
ANNEX 2. EPO SEARCH PROCEDURES .....	26
ANNEX 3. COMBINING REFERENCES FROM EPO AND INTERNATIONAL SEARCHES .....	28
ANNEX 4. EXAMPLE OF EPO REFERENCES TO NON-PATENT LITERATURE .....	30

## 1. INTRODUCTION

The objective of OECD's citations database project is to build a database so that international citations data can be easily exploited and used for comparative analyses of flows of knowledge and innovation. By linking with OECD's established main Patent database, and its wealth of patent information, pertinent indicators can be developed and a wide range of analyses should be possible - such as those undertaken by Hall, Jaffe, Trajtenberg *et al.* during their extensive work over the years with US patent citations data.

First efforts, using information provided by the European Patent Office (EPO), started in 2003 and led to a progress report being presented to the OECD Patents Statistics Task Force meeting<sup>1</sup> at the World Intellectual Property Organisation (WIPO) in Geneva<sup>2</sup> later that year. This was followed by discussions between EPO, OECD and INNO-tec at the University of Munich leading to a better understanding of the issues involved when dealing with EPO and international (PCT) citations information. The INNO-tec team undertook preliminary analyses of citations using the database which further enhanced understanding. Development of the database has also benefited from expert advice from EPO staff and become a useful input into their own PATSTAT database project – a project whose principal aim is to make patent information at EPO more accessible for statistical analysis.

During the October 2004 meeting of the Patents Statistics Task Force it was agreed that OECD tables of EPO patent citations information should be made available to other researchers to refine methods, design indicators and help determine optimum data structures for various analyses. At the same meeting EPO agreed to continue its work on a glossary to aid researchers in understanding patenting processes.

This document presents a preliminary set of OECD/EPO patent citations database 'building blocks' for wider distribution and discusses methodological issues, limitations of the data and further work required. The idea is to provide researchers with a structured set of raw data that can be easily downloaded into local analytical and statistical systems (such as SQL, SAS, STATA, SPSS, TSP, etc.). It should be stressed, however, that this data set is not yet complete – information still needs to be added in certain areas, particularly concerning cited patents. Also, additional measures could be added such as priority numbers and dates and counts of references of Non-Patent Literature (NPL). An improved version of the data set, including data for more recent years, is scheduled for release in early 2006.

Finally, it is assumed that potential users are reasonably familiar with recent citation analysis literature. Possible avenues of research such as knowledge spillovers (geographical or across fields of science), knowledge flows and value of patents, are not addressed in this paper but dealt with in the related document, "European Patent Citations – How to count and interpret them?" (Harhoff *et al.* 2005).

- 
1. Attended by representatives from OECD, WIPO, EPO, United States Patent and Trademark Office (USPTO), Japanese Patent Office (JPO), European Commission (EC), EuroStat and US National Science Foundation (NSF).
  2. [www.wipo.int](http://www.wipo.int)

## 2. PATENT PROCESSES

### 2.1 Brief notes on international (PCT) and European application and search processes

International patent applications filed under the Patent Co-operation Treaty (PCT) at WIPO are options for future applications to patent offices around the world (signatories to PCT) including EPO. International searches are carried out by International Search Authorities (ISAs)<sup>3</sup> though the results are not legally binding and WIPO cannot grant patents. Within 18 months of the priority date, usually following an international search, applicants have three options: *i*) to move their PCT applications to designated national patent offices or regional patent offices (such as the EPO), entering the “*national or regional phase*”; *ii*) to elect to have an *international preliminary examination*, deferring transfer to *national or regional phase* for up to 30 months after the priority date; or *iii*) withdraw the application (or simply abandon it). Applicants can also apply directly to the EPO, though increasing numbers are filing under PCT first. See Annex 1 for more details of the EPO and PCT application processes.

European searches follow the *Guidelines for Examination in the European Patent Office*<sup>4</sup> and are carried out by the Search Divisions of the EPO. The main objective is to discover prior art relevant for determining whether the invention meets the novelty and inventive step required. European searches differ substantially from USPTO searches, and so do the citations generated in the two processes. This means that researchers wishing to employ patent citation analysis need to be aware of these differences. The European search report should only include (as references) the most important documents; or the earliest of equally important documents. According to EPO philosophy, a good search report contains all relevant information within a minimum number of citations (Michel and Bettels 2001). Details of EPO search procedures are presented in Annex 2.

A very important feature of European search reports is the allocation of search codes to each reference signifying its relevancy to the patent application in question in terms of the three criteria of patentability: novelty, inventive activity and industrial applicability (see Table 1). These characteristics allow researchers to use the classification for weighting or filtering purposes, there is evidence that the composition of patent citations may matter considerably.<sup>5</sup>

### 2.2 Comparing with USPTO citations data

Applicants to USPTO are legally required to include a full list of prior art known, or believed, to be relevant (‘duty of candor’) – these can be subsequently filtered and/or supplemented by the examiner. At EPO no such requirement exists – citing of prior art within the application document on the part of the applicant or the applicant’s patent attorney is facultative. The vast majority of the references are added by the EPO examiner during the search - although some researchers suggest that there is an incentive for EPO applicants to list all prior art to avoid future patent opposition (Akers, 2000). References which were already listed by the applicant *and* have been adopted by the examiner are characterised by the letter ‘D’.

- 
3. Current ISAs are EPO and the national patent offices in Australia, Austria, China, Japan, Korea, Russia, Spain, Sweden and the United States (USPTO).
  4. Online at [www.european-patent-office.org/legal/gui\\_lines/](http://www.european-patent-office.org/legal/gui_lines/)
  5. See Harhoff, Hoisl and Webb (2005).



Some commentators have noted that applicants to USPTO may provide more references than necessary and that USPTO examiners cannot always check them all. This, combined with EPO examiners' minimalist approach, goes some way to explain the fact that the average number of citations in USPTO patents is significantly greater than those in found in EPO patents (Michel and Bettels 2001).<sup>6</sup>

**Table 1. EPO search codes and their meaning**

Code <sup>1</sup>	Meaning
<b>X</b>	<b>Particularly relevant documents</b> when taken alone (a claimed invention cannot be considered novel or cannot be considered to involve an inventive step)
<b>Y</b>	<b>Particularly relevant documents</b> if combined with one or more other documents of the same category, - such a combination being obvious to a person skilled in the art
<b>A</b>	<b>Documents defining the general state of the art</b> (but not belonging to X or Y)
<b>O</b>	<b>Documents which refer to non-written disclosure</b>
<b>P</b>	<b>Intermediate documents</b> - documents published between the date of filing of the application being examined and the date of priority claimed
<b>T</b>	<b>Documents relating to the theory or principle underlying the invention</b> (documents which were published after the filing date and are not in conflict with the application, but were cited for a better understanding of the invention)
<b>E</b>	<b>Potentially conflicting documents</b> – Any patent document bearing a filing or priority date earlier than the filing date of the application searched but published later than that date, and the content of which would constitute prior art
<b>D</b>	<b>Documents cited in the application</b> <i>i.e.</i> already mentioned in the description of the patent application
<b>L</b>	<b>Documents cited for other reasons</b> ( <i>e.g.</i> a document that may throw doubt on a priority claim)

1. Up to three codes can be allocated *e.g.* AD, XD, XP, YP, APD, XPD

Source: EPO (2003).

Until recently<sup>7</sup> USPTO generally only published granted patents while EPO publishes all applications whether eventually granted, refused or withdrawn. Strictly speaking, a patent is not a patent until it is granted so the term 'patent application' is preferred here. Granted EPO patent applications can be identified in this data set.

Finally, the widely used NBER citations data set (Hall *et al.* 2001) only includes references to US patents - references to patents published in other offices, including EPO, have been removed. So while it is an excellent data set for studying the US market, it may have its limitations for certain analyses in an international context - see Harhoff *et al.* (2005) for a discussion. In fact, by discarding citations of patents published by other offices (in particular EPO or WIPO, under PCT) citations to some US patents have also been discarded (see discussion of equivalents in 3.4). This EPO data set includes all references to earlier published patent applications.

6. In the context of EPO patent applications, some experts prefer the term 'reference' to 'citation' arguing that in many cases, the documents listed to define prior art have not been submitted by the applicant(s) and are therefore not citations in the commonly understood sense of the word. Both terms are used interchangeably in this document.

7. Since 2001, USPTO has published applications (called pre-grant publications) and although they include the applicant's citations, they do not include those of the examiner. The latter are only published if and when the patent is granted (Mogee 2005).

### 3. GENERAL DESCRIPTION OF THE DATA SET

#### 3.1 Organisation

The data are presented as a set of zipped text files (delimited by semi-colons or vertical bars) dividing a range of EPO patent information into a set of tables in order to allow researchers to construct their own databases according to the type of analyses they wish to carry out. They have been generated from SQL tables at OECD and are thus relational in nature. *The recommended citation is OECD/EPO patent citations database, 2004.*

Citation information for EPO and international (PCT) publications are kept in separate tables to allow separate analyses of citations resulting from European and international searches, respectively, to be performed. The set of distinct cited patents are kept in a separate table with publication and application dates, a 4-character IPC sub-class<sup>8</sup> and the total number of citations received. The set of cited Non-Patent Literature (NPL) codes are kept separately with just the number of citations received – no other information is currently provided. There are two ‘link’ tables that *i)* link pairs of EP and PCT publication numbers (prefixed by WO) for patents published under both regimes to facilitate analysis of this set of patent publications, and *ii)* link EP publications with their equivalent publications in national patent offices. The structure allows a broad range of analyses to be performed via links with the supplementary tables from OECD’s main patent database and its wealth of information concerning inventors, applicants, International Patent Classification (IPC) and outcomes. The information from the supplementary tables can be linked via the EPO publication numbers.

#### 3.2 The main tables

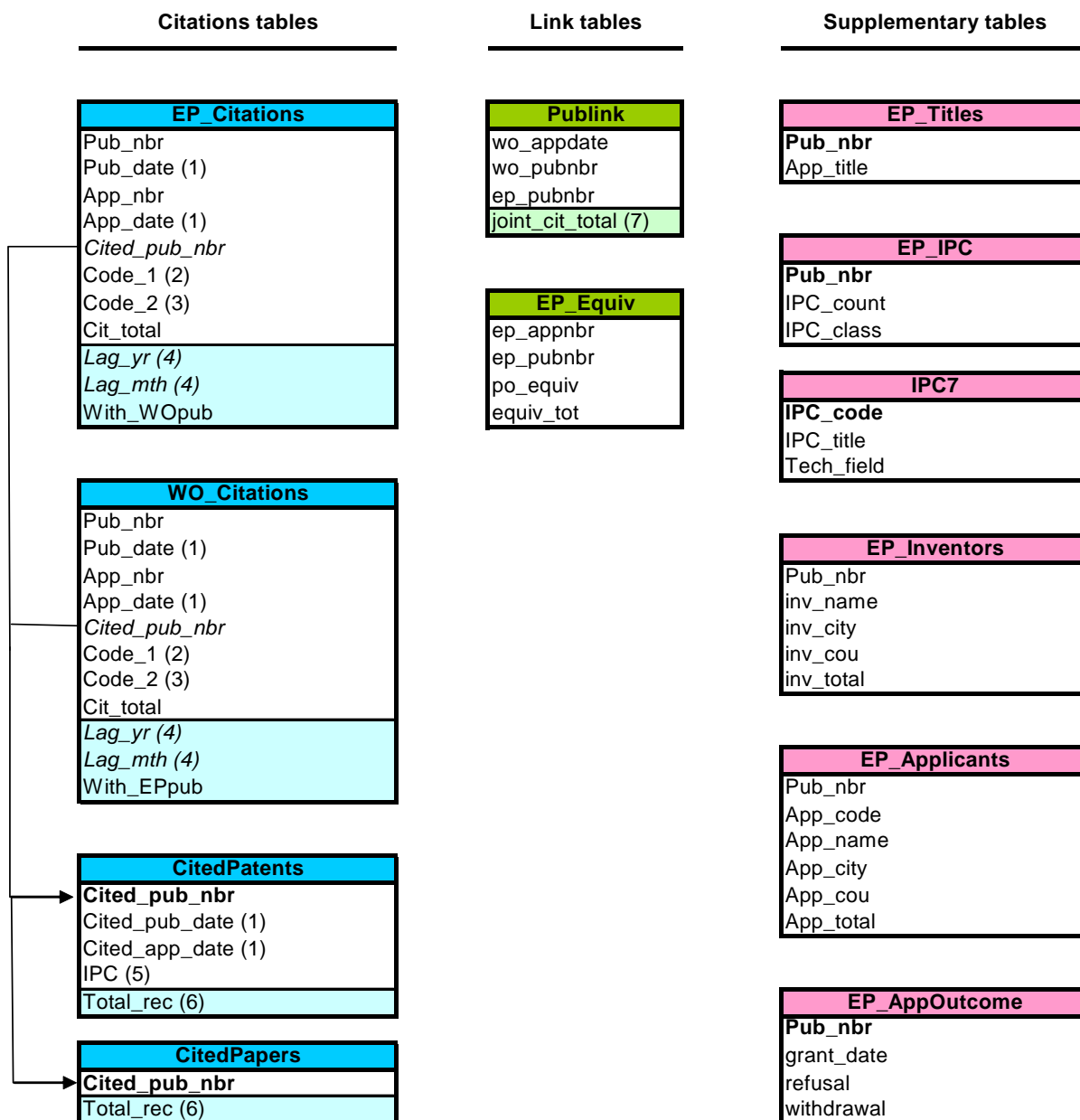
The citations tables, based on data supplied by EPO at the end of 2001, represent all **patent applications published by EPO and under PCT from their introduction in 1978 up until November 2001**. The data are divided into two groups:

- **EP\_Citations** [3 835 727 rows representing about 1 158 000 different publications]: Patent applications published by EPO, with publication numbers prefixed by “EP”, showing citations resulting from EPO “European searches”.
- **WO\_Citations** [2 732 348 rows representing about 607 000 different publications]: Patent applications published under PCT, with publication numbers prefixed by “WO” - citations resulting from “International searches” performed by ISAs according to standardised procedures.

---

8. Currently the first IPC sub-class shown in the list allocated to a patent application - often considered by researchers as the primary IPC. However, at EPO there is no obligation to designate a primary IPC first in the list although some examiners may do so (*e.g.* German examiners following German practice). Some examiners may list the IPC sub-classes alphabetically.

Figure 1.



(1) Of the form YYYYMMDD. Filled by information from country patent data sets as far back as possible. For older patents in CitedPatents, other sources required - for example implied year from publication number (expressed as YYYY0000).

(2) Document-type code for cited patent application. Mainly 'A' = first publication.

(3) Search code(s) which, for the patent application in question, grade the cited documents according to relevancy, taking into account the three criteria of patentability: novelty, inventive activity and industrial applicability (see table 1).

(4) Calculated here as Pub\_date - Cited\_pub\_date. However, using citing patent's application date may be preferable in some circumstances. Lag\_years are given for cases where months not available.

(5) Currently the first IPC code given in the search report (*cf.* EP\_IPC).

(6) Combined count of citations received from both European and International (PCT) searches.

(7) Combined count of citations made in both European and International (PCT) search reports. Adjusted for cases where the same citation is made in both search reports for particular citing patent applications (about 7 000 cases).

In both cases, all published patent applications are present, whether or not they had any citations associated with them. For each published patent application these tables contain:

- The application number and date (repeated where there is more than one citation).
- The publication number and date<sup>9</sup> (repeated where there is more than one citation).
- Publication numbers of cited patents and/or non-patent literature, if any.
- A code indicating the document type of each cited publication (Code\_1).
- Search code(s) for each patent application/citation combination (Code\_2) - see Table 1.
- The total number of patent and (coded) NPL references in the search report.
- Calculated 'citation lags' (difference between publication date and cited patent's publication date, where available) - *not provided by EPO*.

In all, over 6 180 000 citations are present in this data set. Full information on individual EPO and PCT patent applications can be found at <http://register.epoline.org/espacenet/ep/en/srch-reg.htm>

While it is useful to keep these tables separate, for many analyses of citations it is essential that they are combined since many EPO applications (in EP table) that were previously submitted under PCT, and thus subjected to an international search (in WO table), do not have any references attributed to them. This is mainly because when EPO is the ISA, there is no European search report. For an illustration of this see Annex 3.

Figure 2 summarises how the main **EP\_Citations** and **WO\_Citations** tables are related. As the status of the patent applications published by WIPO, but not yet EPO at the time of the original EPO citations data extraction (end-2001) is not known, it is recommended that studies should concentrate on the EPO data merged with the WIPO data (B+C) or just the merged set (C). The period 1987-1999 for certain analyses of citations made (particularly counts) would seem reasonable for the moment – this eliminates possible biases in the early years when the international patenting process was in transition and later years where *truncation*<sup>10</sup> effects abound.

---

9. The publication date when the data were extracted. In most cases (99%) this is the first publication date (18 months after the priority date, document A1).

10. Truncation is a commonly used term in patent analyses to describe the impact of the, sometimes lengthy, patenting processes which can result in biases in certain indicators for recent years. For example, since the grant of a patent can typically occur between 2 to 10 years from first filing, when looking at the evolution of the percentage of patents granted by priority year, figures for recent years should be treated with caution since full information is not yet available for many patents. This can be a particular problem for analyses of citations received since a patent may not come near its full potential to be cited for many years.

**Figure 2.**

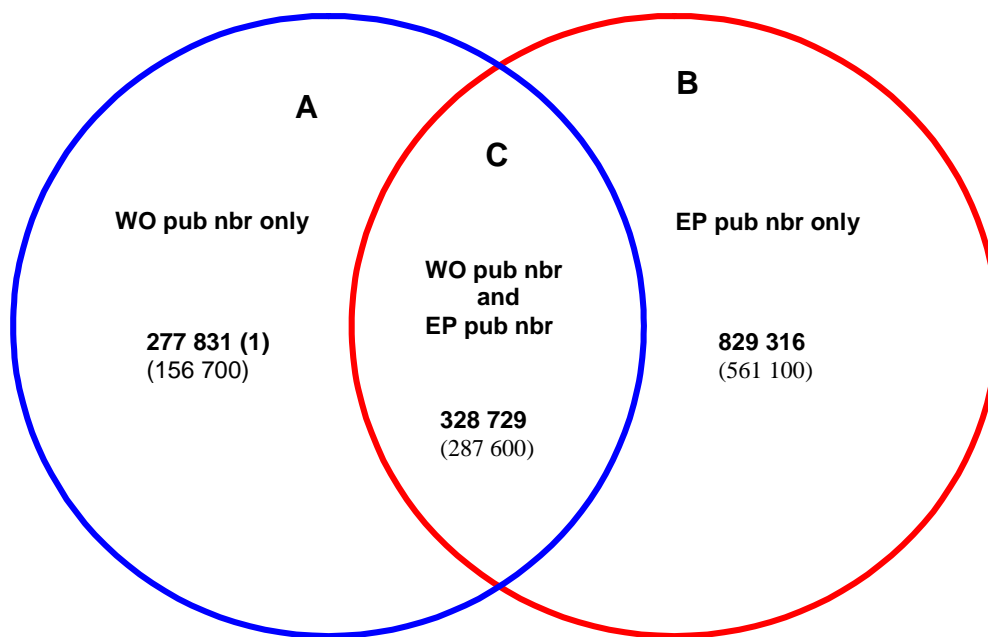
Total number of different patent publication numbers provided:  
**1 764 605** (1 293 000 in period 1987-99)

**WO publication numbers (PCT)**

Citations from international search (by ISA)  
**606 560** (444 300)

**EP publication numbers**

Citations from European search  
**1 158 045** (848 700)



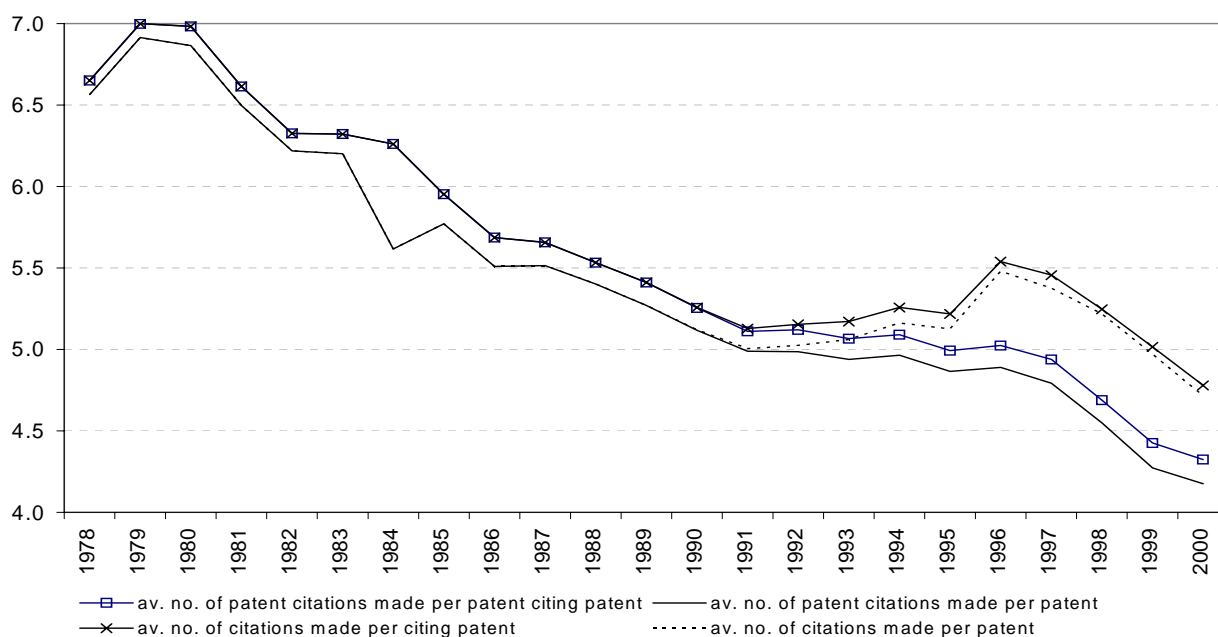
Total after adjusting for distinct patents published by both EPO and under PCT (**A+B+C**):  
**1 435 876** (1 005 400)

(1) Includes over 67 500 PCT (WO) publications that had not yet been published by EPO at the time of EPODOC data extraction (November 2001) but have been published since - an aspect of truncation.

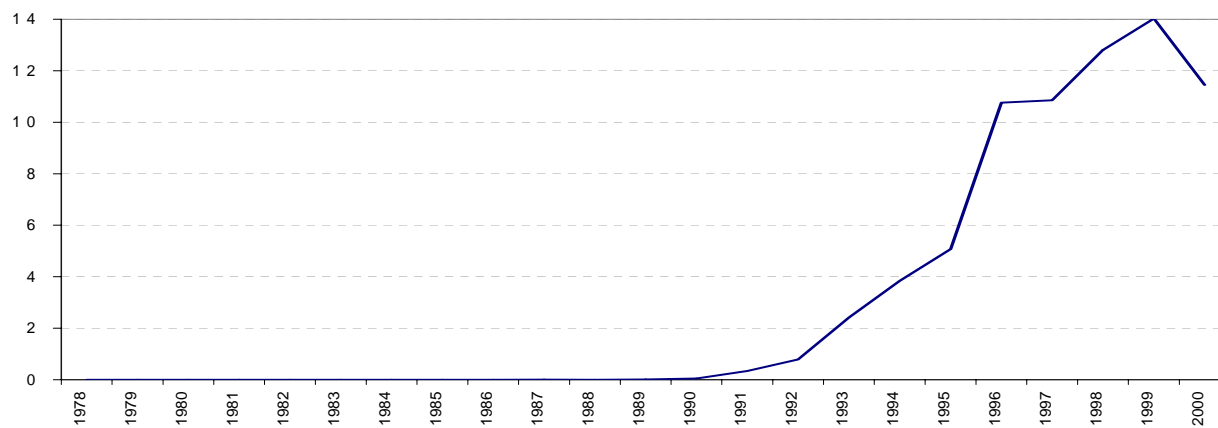
Using the merged set of patents some basic analyses can be carried out such as the average number of citations per patent application. Figure 3 presents four such measures: *i*) all citations per patent application; *ii*) all citations per citing patent application (*i.e.* excluding those with no citations); *iii*) patent citations per patent application (*i.e.* excluding NPL citations); and *iv*) patent citations per patent citing patents (*i.e.* excluding those that only cite NPLs).

The fall over time in the average number of citations made, shown in Figure 3, is due to a reduction in the average numbers of citations in the international search reports - the average number of citations made in European search reports has remained relatively stable during the last two decades. Finally, Figure 4 suggests that it is preferable to consider patent citations separately from NPL citations since the latter have only been systematically coded (with the prefix 'XP') since the early 1990s.

**Figure 3. Citations made per patent application**  
 Patents published by EPO and WIPO – by PCT application year



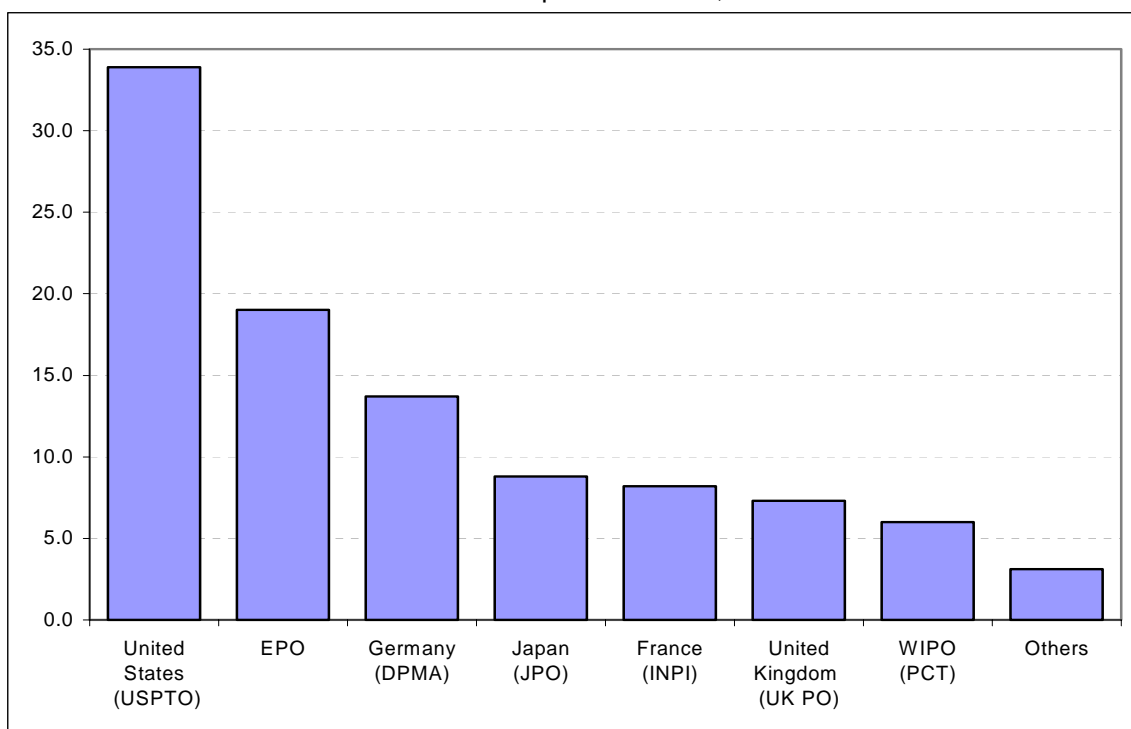
**Figure 4. Coded non-patent literature (NPL) as a percentage of citations made**  
 Patents published by EPO and WIPO – by PCT application year



### 3.3 Tables of Cited Patents and Non-Patent Literature (NPL)

The database currently contains over 3 400 000 *different* cited patents, dating back to the mid-nineteenth century, and over 230 000 different cited NPLs. 97 per cent of patents cited in EPO and WIPO publications originate from seven main patenting authorities – see Figure 5 below.

**Figure 5. Distribution of patent references in EP publications**  
from PCT and European searches, 1978-2001



**CitedPatents** [3 414 470 rows]: This table, generated from EP\_Citations and WO\_Citations, contains each different cited patent publication together with its publication and application dates as well as an IPC sub-class and the total number of citations received so far. This table is useful - while full information for cited EP and WO publications can be taken from the other tables provided, information for cited publications from national patent offices needs to be obtained from other sources.

In fact, even with the minimum requirements here, many data cells still need to be filled (see Table 2). In general, information for patents published after the 1950s by the main patenting authorities exists in electronic form and is available from national offices (*e.g.* USPTO) and is collected by EPO for their DOCDB database.<sup>11</sup> Using OECD's version of the country data sets derived from DOCDB, publication dates were filled for British, French and US patents. Tables provided by the USPTO enabled publication years going back to 1870 to be entered for US patents. Publication dates for German and Japanese patents still need to be filled - this will require matching the codes given in the country databases with those entered in the European search report.

11. This information can be accessed online via the EspaceNet service:  
[www.espacenet.com/access/index.en.htm](http://www.espacenet.com/access/index.en.htm)

Compiling IPC information is complicated by the fact that older patent publications have codes from earlier versions of the IPC allocated to them. The latest classification is the 7th revision. Older patents can be classified according to IPC revisions 4, 5 or 6.

**Table 2. OECD CitedPatents table status 2004**

Publishing Authority	Number of patents	% of total	Missing pub_date	%	No IPC	%
USA	1205423	35.3	1015	0.1	439882	36.5
DPMA (DE)	489100	14.3	<b>489100</b>	<b>100.0</b>	<b>489100</b>	<b>100.0</b>
JPO	449052	13.2	<b>449052</b>	<b>100.0</b>	<b>449052</b>	<b>100.0</b>
EPO	408405	12.0	24	0.0	25	0.0
INPI (FR)	283157	8.3	44069	15.6	<b>283157</b>	<b>100.0</b>
UK PO	249804	7.3	34227	13.7	<b>249804</b>	<b>100.0</b>
WIPO	172109	5.0	7180	4.2	612	0.4
Others	157420	4.6	157420	100.0	157420	100.0
Total	3414470	100.0	1182087	34.6	2069052	60.6

**CitedPapers** [230 415 rows]: This table contains each different cited NPL together with total number of citations received. Analysis of NPL citations is currently limited since they have only been formally coded from the mid-1990s (see Figure 4 above). Only coded NPL (prefixed by XP) are included in the EP\_ and WO\_Citation tables. However, full text files of NPL references are available (see Annex 4).

### 3.4 Link tables

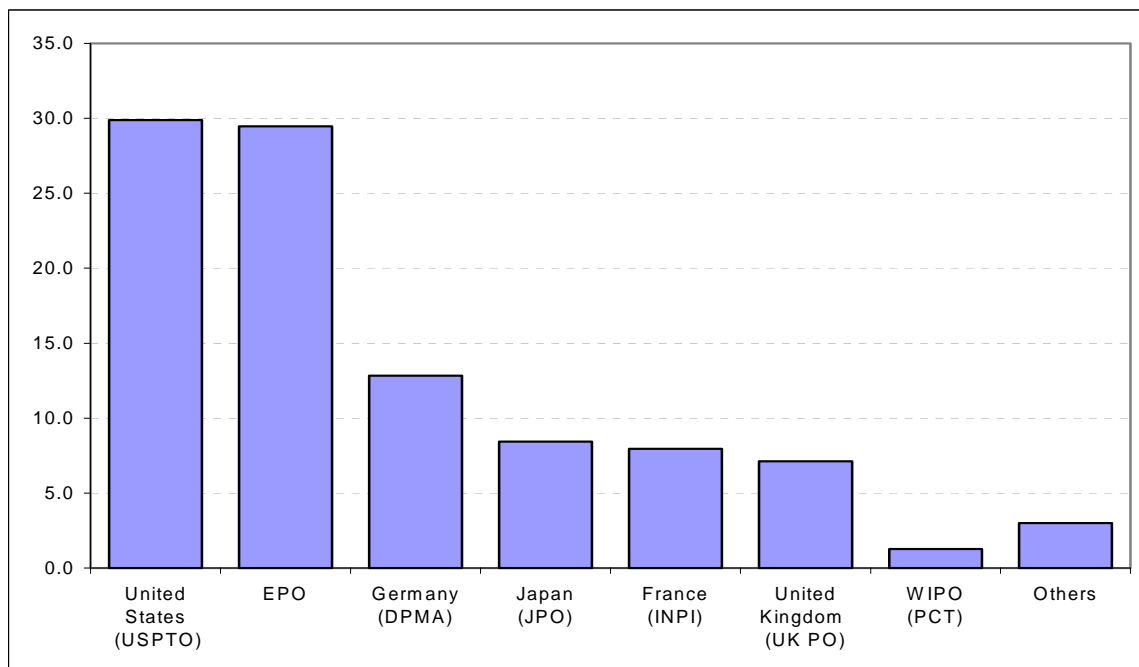
When a patent is published by more than one office (*i.e.* part of a patent family), any version of it can be cited in a search report. The choice of publication to be cited depends on the ease of access to patent information (technical, legal and linguistic) by the searching authorities. Moreover, the respective guidelines recommend a preference for documents that are in the same language as the application for which the references are sought. This issue has implications for the calculation of citations lags, discussed below, and the number of citations received by particular inventions. For certain analyses, families of cited patents need to be identified so that *i)* the number of citations received by particular inventions is not understated; *ii)* counts of different citations received grouped by IPC, for example, are not over stated; and *iii)* citation lags are calculated in a more meaningful manner.

**Publink** [328 729 rows]: This table matches EP publication numbers with PCT (WO) publication numbers, where applicable, to allow citations from European and international searches to be combined.

**EP\_EQUIV** [5 844 192 rows]: For each EPO publication, this table, based on EPO's *Espace* database, provides a list of all the publications in national patent offices pertaining to the same patent [*i.e.* sharing *exactly* the same priority number(s)]. Using this information, citations of national publications can be replaced by the 'EPO equivalent' so that the impact of particular inventions is better measured. Also, introducing EPO equivalents into a modified CitedPatents table, would allow more publication date and IPC information to be added. Compare Figure 6 below with Figure 5.



**Figure 6. Distribution of patent references in EP publications after adjusting for EPO equivalents from PCT and European searches, 1978-2001**



### 3.5 Supplementary tables

These tables allow rich analyses to be performed beyond simple counts of citations. For example, grouping by country of inventors, by IPC, looking at granted patents only etc. If the set of EP-EP citations is just considered, a wide range of dynamics can be studied. These tables were generated in November 2004 and therefore contain EPO applications published after those in the main EP\_ and WO\_ Citations tables.

- **EP\_Titles** [1 479 468 rows]: Provides the title for each EP publication.
- **EP\_IPC** [1 479 468 rows]: Contains the full list of IPC codes pertinent to the EP publication and the total number present. The first code given is not necessarily the primary IPC (see footnote 8.).
- **IPC7** [750 rows]: International Patent Classification Revision 7: 3-digit and 4-digit codes and descriptions ([www.wipo.int/classifications/fulltext/new\\_ipc/index.htm](http://www.wipo.int/classifications/fulltext/new_ipc/index.htm)). Technical fields have not yet been entered.
- **EP\_Applicants** [1 602 152 rows]: For each EP publication, contains information concerning applicants: name, city, country and total number of applicants. There is also an applicant code although these have not been harmonised – *i.e.* if an applicant name has a one character difference from an existing applicant name, a new code will be assigned even if it is the same applicant. Note that EPO can not alter applicant names – the legal status of documents means they accept names as given.

- **EP\_Inventors** [3 531 465 rows]: For each EP publication, contains information concerning the inventors – name, city, country and total number of inventors. Note that inventor names have not been harmonised.
- **EP\_Outcome** [1 479 468 rows]: Shows whether a patent was granted (via the presence of a grant date), refused or withdrawn. All fields empty means that the outcome was not known, the patent was pending, at the time of the data extraction. This table could be reduced to two fields: Outcome set as G,R,W or empty, and a Grant\_date where applicable.

### 3.6 Calculating citation lags

The term “citation lag” refers to a measure of time passing between a characteristic date of the referencing patent application and a characteristic date of the cited document. Citation lags can be computed in various ways, *e.g.* based on priority, application, or publication dates. The lag measure computed in the present dataset is defined as the time between the publication of the cited patent application (in general, a patent or NPL cannot be cited before it is published) and the publication date of the referencing search report. Thus, the measure developed here answers the question: how much time passed between the publication of a patent application or patent grant and the appearance of a reference to said patent in the search report of the referencing patent.

Some implications of this choice need to be pointed out. For most of the cited patent documents originating with European patent offices or the JPO, publication (including the revelation of search results to the public at large) occurs exactly 18 months following the priority date. Hence, for the computation of citation lags of European patents, it does not matter if we choose the date of the search report or the date of the publication of the application. Moreover, we could choose to take the differences between priority dates, and the measure would be equivalent to the one computed here.

If the cited document is a US patent which was only pursued within the United States, the earliest publication date is equivalent to the grant date. If the cited US patent has an international equivalent at the EPO, the corresponding international applications are again published 18 months after the US priority date. Under the American Inventors Protection Act (AIPA), all patents which seek some form of patent protection outside of the US are published by the USPTO, again 18 months after the US priority date. That does not change the timing of the earliest publication, but the publication is now available from the USPTO and will show up in our data even if we do not detect European equivalents of the US patent.

For patents with the international search report published by WIPO and a supplementary search report published by EPO, we have multiple publication dates. If the referenced documents have no overlap, we compute the time lag between the date of the publication of the relevant search report. If the international search report and the EPO supplementary search report reference the same document, we drop the later entry from the list and take the earliest publication date of the two search reports for computing the citation lag.

Under certain circumstances citation lags can be negative. These include *i)* when the search code E is present; *ii)* when a publication date later than the first publication date is given for the referenced patent and *iii)* coding errors in the originating EPO REFI database – though this is quite rare.

### 3.7 Grouping citations by IPC

Table 3 shows those ten 4-digit IPC classes that had the highest average number of citations. The table hints at the predominance of the biotechnology and information and communications technology (ICT) fields with regard to citation activity. However, only about half of the entries in CitedPatents table currently have an IPC sub-class allocated to them, notably those concerning EPO and WO patent applications and US patents from 1976. These tables are therefore just illustrative.

The IPC code in CitedPatents was taken as the first 4-character IPC sub-class given in the patent publication. However, as mentioned earlier (see footnote 8) this should not necessarily be considered the primary IPC sub-class. For EPO publications, by linking to the EP\_IPC table, other codes in the list can be taken into consideration. Tests involving whole counting or fractional counting of all given IPC codes should be carried out to determine whether or not using the first IPC code is a good proxy for certain analyses. Preliminary analysis suggests that the impact of the different approaches can vary across sub-classes.

By compiling IPC information for cited patents coming from the seven main patenting authorities, 97% of the cited patents will be covered. Finally, note that exploiting the Equivalents table would increase coverage via the use of the information given for EP publications

**Table 3. Top 10 counts of citations received by IPC sub-class**

IPC sub-class	different cited patents	citations received	av. no. of times cited	% of total	IPC description
A61K	48970	126739	2.59	4.34	PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES
G06F	42292	101113	2.39	3.46	ELECTRIC DIGITAL DATA PROCESSING
C07D	27576	77838	2.82	2.66	HETEROCYCLIC COMPOUNDS
G01N	29219	69691	2.39	2.38	INVESTIGATING OR ANALYSING MATERIALS BY DETERMINING THEIR CHEMICAL OR PHYSICAL PROPERTIES
H04N	23698	62712	2.65	2.15	PICTORIAL COMMUNICATION, <i>e.g.</i> TELEVISION
H01L	29035	60901	2.10	2.08	SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR
A61B	23781	59324	2.49	2.03	DIAGNOSIS; SURGERY; IDENTIFICATION
C07C	23500	51743	2.20	1.77	ACYCLIC OR CARBOCYCLIC COMPOUNDS
C12N	15724	46279	2.94	1.58	MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF
G11B	18344	45918	2.50	1.57	INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER
<b>All</b>	<b>1345418</b>	<b>2922862</b>	<b>2.17</b>		

## 4. NEXT STEPS

### 4.1 Further enhancements to the data set

Many improvements could be made to the data set to expand the range of possible analyses. These include:

- **Introduction of priority dates and numbers:** The first date the patent application was filed anywhere in the world. A patent application can arise from many priority filings while some priority filings can lead to more than one patent application. They allow broader definitions of patent families to be introduced – for example, OECD’s definition of triadic patent families.<sup>12</sup> They are not included in this first release of the data set.
- **Harmonisation of applicant names:** Partial work has been carried out by some researchers and Eurostat initiated a project to harmonise EPO applicant names in early 2005. For the moment users could attempt to harmonise names when research focuses on certain areas – *Users are encouraged to share any results of harmonisation efforts for possible inclusion in later releases of the data tables.*
- **Exploitation of full lists of NPL citations,** including those not coded (see Annex 4). A coding system that identifies journals would be very useful.
- **Self-citations:** For example, add a dummy variable (1 or 0) to EP\_Citations to denote whether a citation includes the names of one or more of the applicants in the citing patent. Requires harmonised applicant names in patent publications. Ideally, any similar analyses by inventor would also require the inclusion of NPL citations with harmonised names.
- **International Search Authorities (ISAs):** Including a code, in WO\_Citations, for the ISA which carried out the international search could reveal varying citation practices by different national or regional patent offices when acting as an ISA.
- **More procedural data** such as information on opposition.

### 4.2 Next version of the data set

The mostly likely improvements in a future version of this data set are:

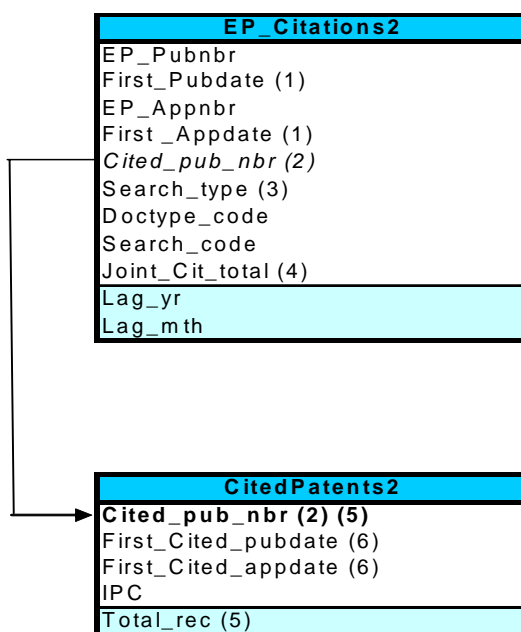
- Inclusion of EPO citations data for later years.
- A re-organisation of EP\_IPC supplementary table so that there is one entry per publication number (pub\_nbr) / IPC sub-class combination.

---

12. See [www.oalis.oecd.org/olis/2004doc.nsf/linkto/dsti-doc\(2004\)2](http://www.oalis.oecd.org/olis/2004doc.nsf/linkto/dsti-doc(2004)2).

- More complete information for cited patents published by national patent offices - notably those of France, Germany, Japan and the United Kingdom.
- Total number of NPL citations, including those not coded.
- A set of summary statistics so that researchers can emulate the results to verify their databases.
- A table introducing priority dates.
- Experimental merged tables combining **EP\_Citations** and **WO\_Citations** and modifying **CitedPatents** using relations in **Publink** and **EP\_EQUIV** (see proposal in Figure 7).

**Figure 7. Possible combined EP citations tables**



(1) Earliest publication/application date between EPO application and PCT application (if international application made). Usually PCT date(s). Further refinement would be to check against publication/application dates at National offices.

(2) After replacing non-EP publication numbers where possible: first WO numbers using PUBLINK where applicable, then remaining non-EP numbers using EP\_EQUIV where applicable.

(3) EPO or PCT.

(4) Combined count of citations made in both European and International (PCT) search reports.

(5) Merge Cited\_pub\_nbr so that there are no duplicate EP pub numbers and recalculate Total\_rec = total citations received.

(6) Earliest publication/application date determined after comparing dates from international (PCT), European (EPO) and National patent applications.

## REFERENCES

- Akers, N (2000), “The referencing of prior art documents in European patents and applications”. *World Patent Information*, 22(3): 306-315.
- Criscuolo, P. and B. Verspagen (2005), “Does it matter where patent citations come from? Inventor versus examiner citations in European patents”, Merit – Infonomics Research Memorandum series, Maastricht: [www.merit.unimaas.nl/publications/rmpdf/2005/rm2005-018.pdf](http://www.merit.unimaas.nl/publications/rmpdf/2005/rm2005-018.pdf)
- Dernis, H., D. Guellec and B. van Pottelsberghe (2001), “Using Patent Counts for Cross-country Comparisons of Technology Output”. *OECD STI Review No. 27*, pp. 129-146: [www.oecd.org/dataoecd/26/11/21682515.pdf](http://www.oecd.org/dataoecd/26/11/21682515.pdf)
- European Patent Office (2004), *The European Grant Procedure*, Munich: [www.european-patent-office.org/epo/pubs/brochure/europat/pdf/europat\\_e\\_20040708.pdf](http://www.european-patent-office.org/epo/pubs/brochure/europat/pdf/europat_e_20040708.pdf)
- European Patent Office (2005), *Guidelines for examination in the European Patent Office*, Munich: [www.european-patent-office.org/legal/gui\\_lines/](http://www.european-patent-office.org/legal/gui_lines/)
- European Patent Office (2005), *Annual Report 2004*, Munich: <http://annual-report.european-patent-office.org/2004/>
- Guellec, D. and B. van Pottelsberghe (2002), “The Value of Patents and Patenting Strategies: Countries and Technology Areas Patterns”, *Economics of Innovation and New Technologies (EINT)*, 11(2), pp.133-148: [www.ulb.ac.be/cours/solvay/vanpottelsberghe/resources/Pap\\_EINT\\_1.pdf](http://www.ulb.ac.be/cours/solvay/vanpottelsberghe/resources/Pap_EINT_1.pdf)
- Hall, B. H., A.B. Jaffe, and M. Trajtenberg (2001), “The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools”. NBER Working Paper 8498: [www.nber.org/patents](http://www.nber.org/patents).
- 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*, Vol. 81, No. 3, pp. 511-515: <http://ideas.repec.org/a/tpr/restat/v81y1999i3p511-515.html>
- Harhoff, D., K. Hoisl and C. Webb (2005), “European Patent Citations - How to count and how to interpret them?”, Presentation to 2nd ZEW Conference on Economics of Innovation and Patenting: [www.zew.de/en/veranstaltungen/details.php?LFDNR=495](http://www.zew.de/en/veranstaltungen/details.php?LFDNR=495)
- Harhoff, D. und S. Wagner (2004), *Modeling the Duration of Patent Examination at the European Patent Office*, Munich: [www.vwl.uni-mannheim.de/stahl/io\\_ausschuss/paper/04\\_harhoff.pdf](http://www.vwl.uni-mannheim.de/stahl/io_ausschuss/paper/04_harhoff.pdf)
- Hinze, S., T. Reiss and U. Schmoch (1997), *Statistical Analysis on the Distance Between Fields of Technology*: [www.tema.liu.se/tema-t/sirp/PDF/325\\_6.pdf](http://www.tema.liu.se/tema-t/sirp/PDF/325_6.pdf)
- Jaffe, A. and M. Trajtenberg (2002), “Patents, Citations and Innovations: a Window on the Knowledge Economy”, *MIT Press*, ISBN 0-262-10095-9.

- Khan, M. and H. Dernis (2005), “Impact of Patent Co-operation Treaty data on EPO patent statistics and improving the timeliness of EPO indicators”, OECD, STI Working Paper 2005/2:  
[www.oecd.org/olis/2005doc.nsf/linkto/dsti-doc\(2005\)2](http://www.oecd.org/olis/2005doc.nsf/linkto/dsti-doc(2005)2)
- Maurseth, P.B. and B. Verspagen (2002), “Knowledge spillovers in Europe: a patent citations analysis”, *Scandinavian Journal of Economics*, 104(4): 531-545.
- Michel, J. and B. Bettels (2001), “European Patent Office, Patent citation analysis: a closer look at the basic input data from patent search reports”, *Scientometrics*, Vol. 51, No. 1, 185–201.
- Mogee, M.E. (2005, *forthcoming*), Comparing Citation Patterns from Patent Publications of the USPTO, the EPO and the PCT, *Report for the National Science Foundation*.
- Narin, F. (2000), “Tech-Line Background Paper”, in From Knowledge Management to Strategic Competence, *Series on Technology Management - Vol. 3*, Imperial College Press, 155-195.
- National Science Foundation (2000), *Science and Engineering Indicators 2000*, Vol. 1, Arlington.
- OECD (2004), *Compendium of Patent Statistics* (Dernis, H. and M. Khan):  
[www.oecd.org/dataoecd/60/24/8208325.pdf](http://www.oecd.org/dataoecd/60/24/8208325.pdf)
- Smith, L.C. (1981), “Citation Analysis”, *Library Trends*, Vol. 30, No. 1, pp. 83-106.
- US Patent and Trademark Office (2002), Issue dates and patent numbers since 1836, *USPTO*.
- Verbeek A, K. Debackere and M. Luwel (2003), “Science Cited in Patents: A geographic 'flow' analysis of bibliographic citation patterns in patents”, *Scientometrics*, Vol. 58, No. 2, 241-263.
- WIPO (2001), *Handbook on industrial property information and documentation*.
- WIPO (2002), *Basic Facts about the Patent Cooperation Treaty (PCT)*. WIPO Publication No. 433 (E). Geneva.

## ANNEX 1. EPO AND PCT APPLICATION PROCESSES

The European Patent Office (EPO) offers a harmonised application and examination path for applicants seeking patent protection in signatory states to the European Patent Convention (EPC). Until 2002, an applicant to EPO designated the EPC member states for which patent protection was requested. From 2003, all member states are automatically designated and the applicant chooses within which member states they would like their patent to be granted - if the EPO has issued the right to do so.<sup>13</sup> The examination process at the EPO consists of an initial search and subsequent examination phase. Once an EPO application has been filed, a search report is generated for the applicant by an EPO patent examiner. This report describes the state of prior art regarded as relevant according to EPO guidelines for the patentability of the invention, *i.e.* it contains a list of references to prior patents and/or non-patent sources. Unlike in the US system, applicants at the EPO are not required to supply a full list of prior art (see Michel and Bettels, 2001, 191ff). Normally, patent applications are published by the EPO 18 months after the priority date.<sup>14</sup> In most cases this is after the search report and the document is designated the code A1 to indicate this. If publication occurs before the search has been carried out, the document is designated the code A2, becoming A3 when the search report is eventually produced.

After the search, the application is normally under examination at the EPO.<sup>15</sup> After an examination has been performed, the EPO either informs the applicant that they have the right to take out patents in EPC member states, as specified in the original application, or requires the applicant to agree to changes in the application. Once an agreement has been found between the applicant and the examiner, the applicant has three months to designate member states after which the patent is duly issued and translated into the relevant national languages. In this process, the applicant may again decide not to pursue the patenting effort, in some or all designated countries, since the prospect of actually obtaining an economically valuable patent may be weak. If the EPO refuses to 'grant'<sup>12</sup> a patent, the applicant may file an appeal. The most frequent outcome, about two-thirds of the cases, is an actual patent 'grant'.

Applications filed under the Patent Cooperation Treaty (PCT) now constitute a large and increasing share of all filings at the EPO (see Figures 8, 9 and 10) and are subject to specific institutional treatments. A PCT filing gives the filing party the option to launch patent applications in up to 115<sup>16</sup> PCT signatory countries within 30 months of the filing date (which becomes the priority date). Any patent application already filed can be turned into a PCT filing within the priority year. PCT filings are advantageous for several reasons. First, they allow the expansion of patent protection to a large number of countries without incurring the full costs and complexity of national applications. Second, applicants will receive an international search report within a relatively short time period, informing them about prior art that may be relevant for their own application's likelihood of being granted. Third, the PCT filing, when compared to a national or regional application, has greater option value since it allows applicants to delay decisions about

- 
13. Although the term 'EPO grant' is often used, EPO does not really grant patents, but issues the right to have patents granted in EPC member states.
  14. In line with the 'Paris criteria' (1883 Paris convention) requiring publication with search report within 18 months of the priority date.
  15. After the initial search, the applicant has six months to file a request for examination.
  16. As of 1<sup>st</sup> April, 2002. See WIPO 2002, p. 12.



the countries for which they seek patent protection for up to 30 months (and not just for the duration of the priority year, as with national and regional applications). PCT filings can also receive a preliminary international examination which is authoritative, but not binding for the national or regional offices finally granting the patent.

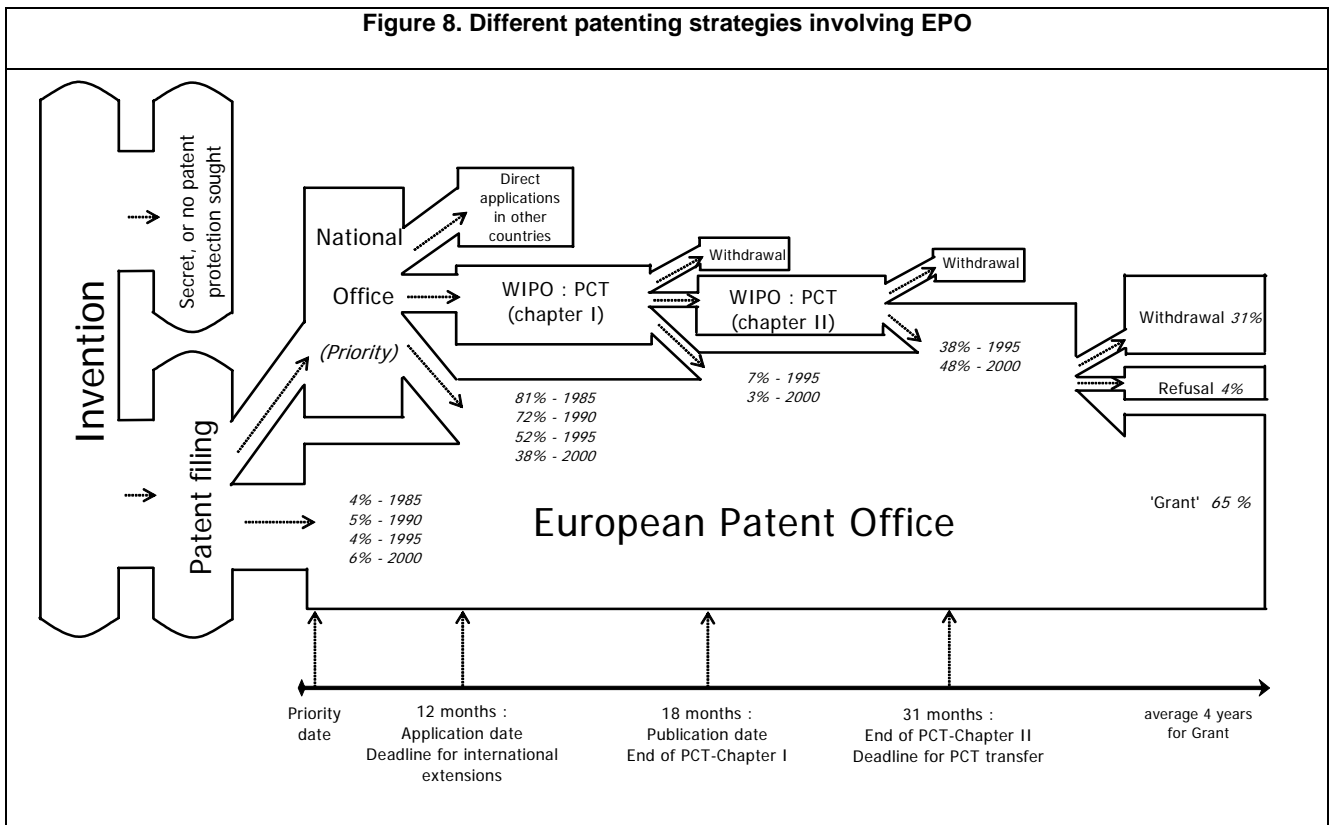


Figure 9. Distribution of lags between priority date and EPO publication date, 1978-2001

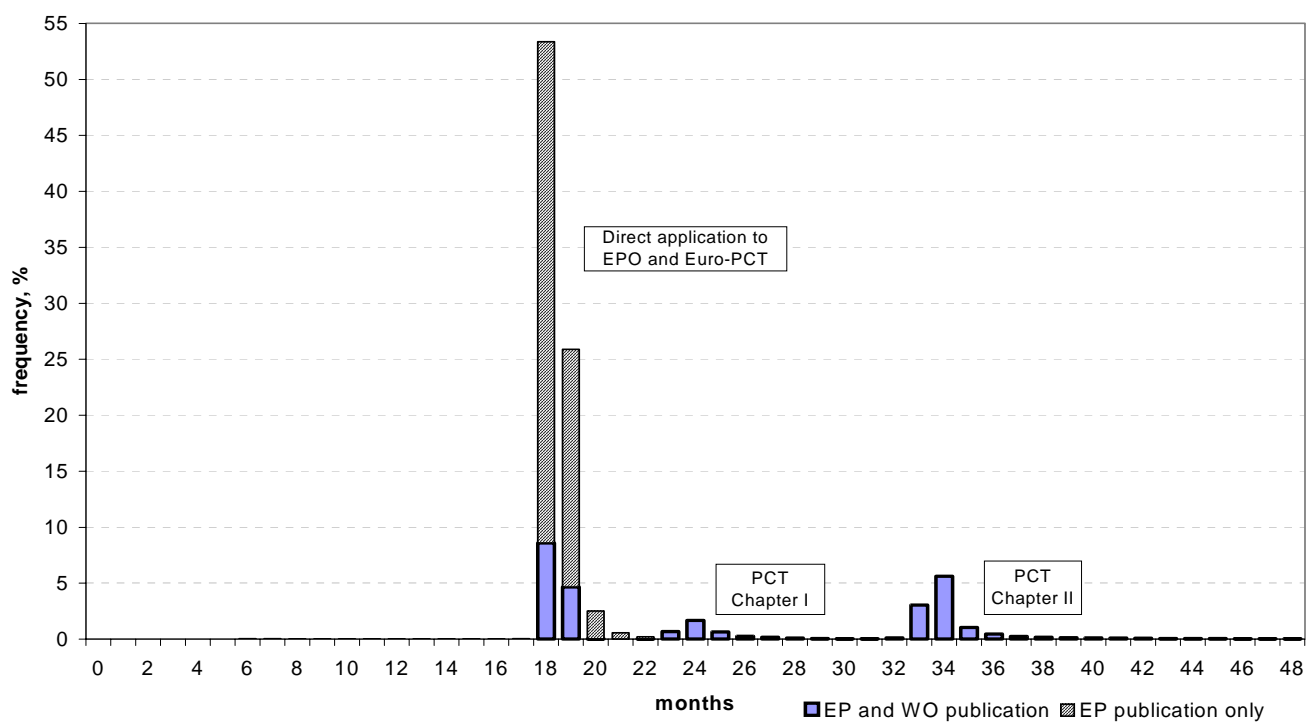
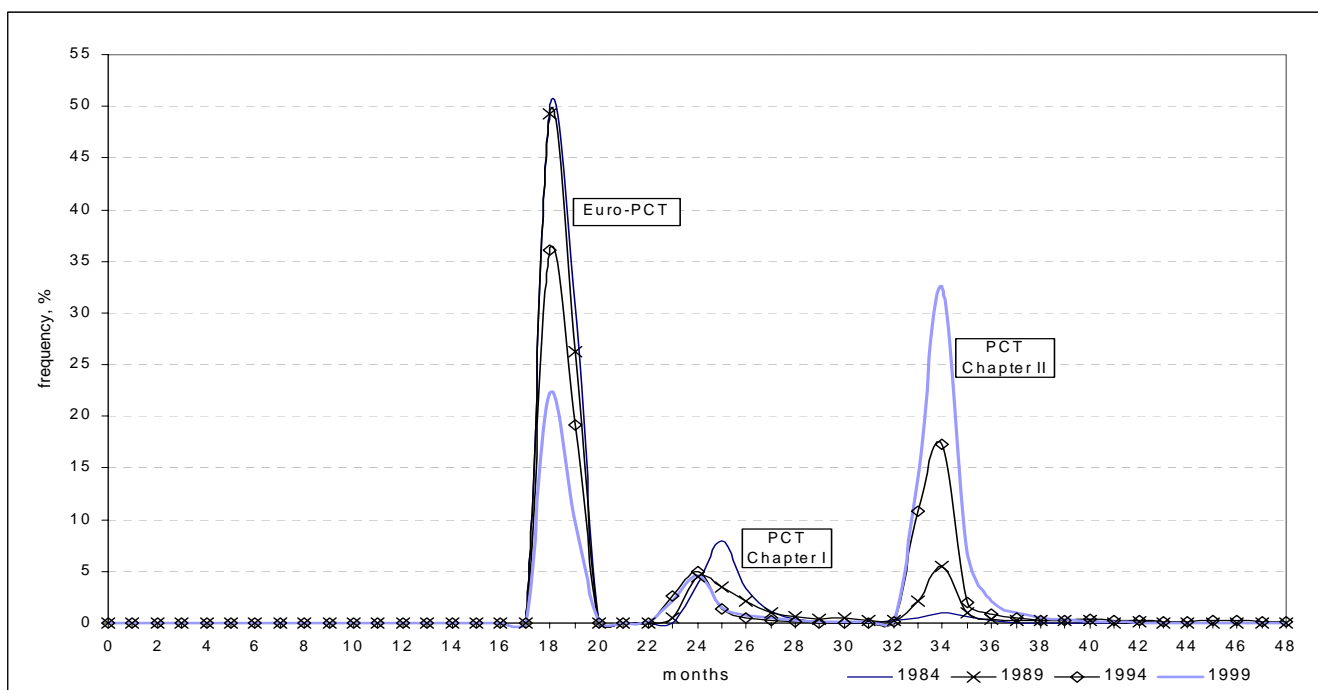


Figure 10. Distribution of lags between priority date and EPO publication date  
EPO publications with PCT (WO) publications, by application year, 1984, 1989, 1994 and 1999



## ANNEX 2. EPO SEARCH PROCEDURES<sup>17</sup>

The search process seeks to identify patent and non-patent documents constituting the relevant prior art to be taken into account in determining whether the underlying invention is new and involves an inventive step. The objective of the substantive examination is to decide whether the invention meets the requirements for patentability: novelty, inventive step, and commercial applicability. The search is especially important for citation analysis because European citation analyses use references to patent and non-patent literature derived from the EPO search report. In order to enable appropriate use and interpretation of citation data, the European search and citation procedure underlying the EP search reports is described in the following paragraphs.

Until recently, overall responsibility for searches lay with the Directorate General (DG) for Searching, located in The Hague, while examinations were carried out at the DG for Examination at the EPO's headquarters in Munich. However, since 2005, these DGs have been merged so that search and substantive examination are carried out by the same examiner (the implementation process enjoying the motto *BEST - bringing search and examination together*). European searches are carried out by the EPO but may also take place in the national patent offices of certain contracting states. The search should be directed to the most important characteristics of the invention and is, therefore, conducted on the basis of the claims. The patent claims describe the scope of protection for which patent protection is designated. According to the guidelines, the examiner should carry out the search focusing primarily on novelty. At the same time he should pay attention to any document that may be important for the inventive step requirement as well as for other reasons, such as conflicting applications or documents facilitating the understanding of the underlying invention. The prior art search should be continued until the probability of discovering further relevant documents is very low in relation to the effort needed. The search may also be terminated when documents have been discovered which doubtlessly demonstrate a violation of novelty of the claimed invention. Note that the prospective value (private or social) of an invention is not taken into account in determining the search effort.

After completion of the search, the examiner has to select the documents to be cited in the search report. The report should only include the most important documents. If the search results in several documents of equal relevance, the search report should normally contain no more than one of them. The decision on which one to use for citation, is made according to the expert knowledge of the examiner. In case of two documents which are of equal relevance, one document published before the date of priority and the other published between priority date and filing date, the search examiner should choose the earlier one. Michel and Bettels (2001) state that “ (...) according to the EPO philosophy a good search report contains all the technical relevant information within a minimum number of citations.” Basically, the relevant information is obtained from one to two documents (Michel and Bettels 2001). Citing no more than what is absolutely necessary can lead to an understatement of certain documents. Additionally, the examiner is obliged to favour early documents over later ones. This could also result in a lack of relevant references for economic analyses.

---

17. Mostly taken from Harhoff, Hoisl and Webb (2005).

If an invention is protected in more than one country and, therefore, several documents exist belonging to the same patent family, the examiner should preferably cite the patent document in the language of the application - patent documents belonging to the same family being listed in the annex of the search report. This language convenience on the part of the examiner may lead researchers to overestimate the influence of the applicant's home country. In this regard, Michel and Bettels (2001) show that 90% of the total number of patent citations made by the EPO refer to EP, DE, GB, WO, or US documents. At the USPTO and the JPO, 90% or more of the references in the search reports refer to national documents.

EPO search reports cite only documents to which EPO examiners have access so that no doubt exists about the contents of the documents cited. However, this procedure may result in an understatement of the documents relevant to prior art. In particular, non-patent literature which is not accessible in databases or arranged in the library of the EPO is likely to be missing in the search report.

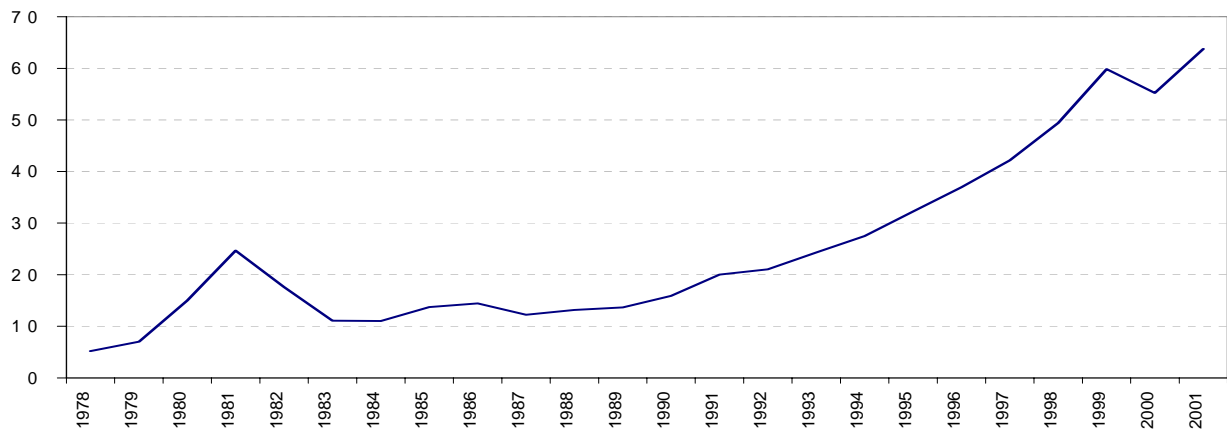
The Search Division does not determine the verification of a claimed priority date. Therefore, the date of filing of the EP patent application is used as a reference date for the search. Exceptionally, documents published after the filing date may be cited. Examples are a later document containing the principle of a theory underlying the invention, which is instrumental in better understanding the invention, or a later document showing that matters of the invention are incorrect. These documents could also be chosen for citation in the search report. Cited documents that were published after the filing date explain negative citation lags. Using the priority date instead of the filing date as a reference date in citation analysis can even increase the number of negative citation lags.

Documents cited by the applicant should be considered in the search report only if they are crucial in defining the state of the art, or when they are necessary for the understanding of the application. Citations which do not fulfil these requirements may be disregarded. Verbeek *et al.* (2003) describe the difference between examiner and applicant citations as follows: Whereas the examiner has to ensure the novelty of an invention, the inventor aims at identifying work "either related to, but significantly different from, or else a useful step towards, the new invention or a use of the invention" (Collins and Wyatt 1988). "Examiner citations, as a result, usually complement, rather than duplicate, the citations given by the inventor" (Verbeek *et al.* 2002). Nevertheless, both examiner and inventor could refer to the same publications.

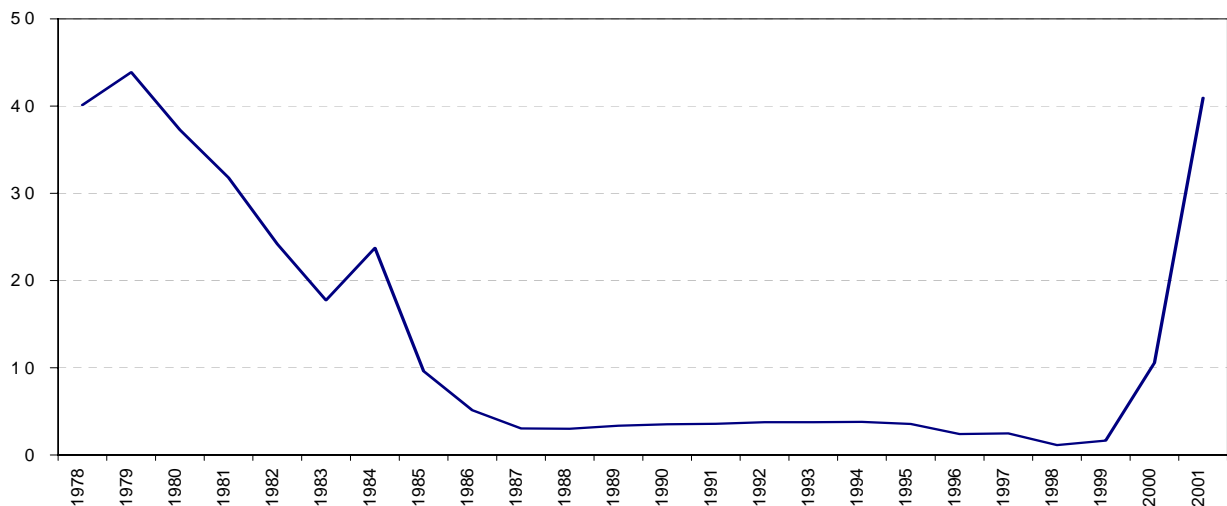
**ANNEX 3. COMBINING REFERENCES FROM EPO AND INTERNATIONAL SEARCHES**

A notable feature of the EPO data is that the number of EP\_ publications with no reference allocated (from a European search) rose from 10% in the mid-1980s to over 60% by end of the 1990s, while for WO publications, the percentage with no references (from an international search) remained low and stable between 1987 and 1999. See Figures 11 and 12 below.

**Figure 11. Percentage of EP publications with no references**



**Figure 12. Percentage of WO publications with no references**

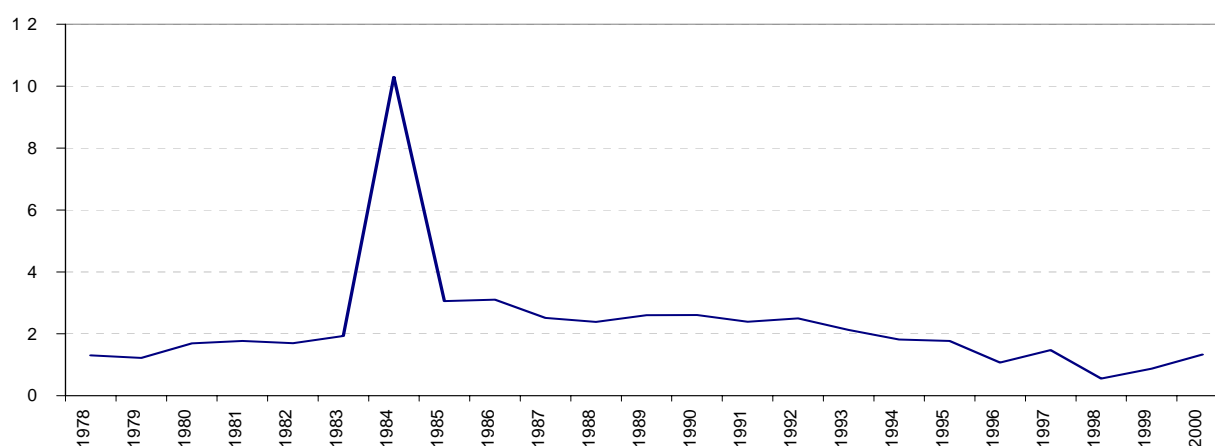


The main explanation is that many patent applications have been examined under PCT, had an international search report published by WIPO, and then have been subsequently published by EPO without undergoing a European search. This is mainly because EPO was the designated ISA carrying out the International search and examination and thus no supplementary examination by EPO was necessary. However, it should be noted that a small proportion of EP publications have no citations allocated to them because at the time of the original data extraction, they had only been published as an A2 document, *i.e.* published before a search was carried out (see Annex 1). For many of these publications subsequent updates of the database (from more recent versions of EPO's REFI database, for example) will introduce citations.

If we examine the set of patent applications published by both EPO and WIPO, under PCT, and combine the references from the respective searches, the percentage with no references is low and stable throughout 1978-2001 with the notable exception of a significant peak in 1984 – the year when major changes to the PCT were implemented (see Figure 13 below). *This peak needs to be investigated further.*

When combining the references from both searches, we are assuming that the European search provides complementary information to the international search rather than superseding it. There are about 7 000 cases where the same patent is cited in both searches – we assume that each search has focused on different aspects of the cited patent.

**Figure 13. Percentage of patents published by EPO and under PCT with no references**



## ANNEX 4. EXAMPLE OF EPO REFERENCES TO NON-PATENT LITERATURE

Pub\_nbr|App\_nbr|NPL\_count|[Search\_code]NPL\_title<sup>18</sup>

EP0348360|EP19890830264|3|[XP] NEUROCHEM. INT., vol. 15, no. 1, 1989, pages 107-113; L.F. AGNATI et al.: "Effects of chronic treatment with uridine on striatal dopamine release and dopamine related behaviours in the absence or the presence of chronic treatment with haloperidol"

EP0348360|EP19890830264|3|[X] ACTA PHYSIOL. SCAND., vol. 132, no. 2, February 1988, pages 209-216; C. FARABEGOLL et al.: "Chronic uridine treatment reduces the level of [3H]spiperone-labelled dopamine receptors and enhances their turnover rate in striatum of young rats: relationship to dopamine-dependent behaviours"

EP0348360|EP19890830264|3|[Y] MED. INT., vol. 12, nos. 8-9, August/September 1977, pages 425-426; CL. MARIE: "Etude de l'uteplex\* en psychiatrie"

EP0348357|EP19890810482|2|[X] PATENT ABSTRACTS OF JAPAN, vol. 8, no. 87, (M-291)[1524], 20th April 1984; & JP-A-59 005 020 (INOUE EMU TEE PII KK) 11-01-1984

EP0348357|EP19890810482|2|[A] PATENT ABSTRACTS OF JAPAN, vol. 12, no. 364 (M-747)[3211] 29th September 1988; & JP-A-63 118 231 (KASAI KOGYO CO. LTD) 23-05-1988

EP0348353|EP19890810464|4|[X] BIOLOGICAL ABSTRACTS, vol. 63, no. 7, 1977, abstract no. 41007, Biological Abstracts, Inc., Philadelphia, PA, US; T. KITA et al.: "Analgesic effect of neurotropin in mice, and comparison between analgesic effects of some drugs in SART-stress mice and normal mice", & FOLIA PHARMACOL. JPN. 72(5): 573-584. 1976

EP0348353|EP19890810464|4|[X] BIOLOGICAL ABSTRACTS, vol. 85, no. 8, 1988, abstract no. 82541, Biological Abstracts, Inc., Philadelphia, PA, US; C. HARA et al.: "Effects of neurotropin on immunodeficiency and ulcer development of rats exposed to activity-stress", & JPN. J. PHARMACOL. 45(4): 459-466. 1987

EP0348353|EP19890810464|4|[X] BIOLOGICAL ABSTRACTS, vol. 87, no. 4, 1989, abstract no. 41156, Biological Abstracts, Inc., Philadelphia, PA, US; T. HATA et al.: "Mechanism of the analgesic effect of neurotropin", & JPN. J. PHARMACOL. 48(2): 165-174. 1988

EP0348353|EP19890810464|4|[XP] BIOLOGICAL ABSTRACTS, vol. 87, no. 11, 1989, abstract no. 117949, Biological Abstracts, Inc., Philadelphia, PA, US; M. TANAKA et al.: "Effects of neurotropin on regional brain noradrenaline metabolism in rats", & JPN. J. PHARMACOL. 49(2): 187-196. 1989

EP0348351|EP19890810450|3|[Y] JOURNAL OF ORGANIC CHEMISTRY

EP0348351|EP19890810450|3|[A] JOURNAL OF ORGANIC CHEMISTRY

EP0348351|EP19890810450|3|[ ] MIHALY NOGRADI: "STEREOSELECTIVE SYNTHESIS"

EP0348349|EP19890810448|1|[Y] PATENT ABSTRACTS OF JAPAN vol. 11, no. 32 (P-541)(2479) 30 January 1987, & JP-A-61 201251 (FUJI PHOTO FILM) 05 September 1986.

---

18. Note that references to "Patent Abstracts of Japan" are effectively references to Japanese patents and should already be included as such in the published patent document.