OECD CITATIONS DATABASE

August 2022 edition

BACKGROUND INFORMATION

The OECD Citations database provides information on patent and non-patent literature (NPL) citations (or references) found in patent documents. Data covers citations made in patents filed at the European Patent Office (EPO) or via the Patent Co-operation Treaty (PCT), and now includes citations made in patents filed at the United States Patent and Trademark Office (USPTO). If the same patent has been published by several national or regional offices (EPO, WIPO, USPTO, JPO etc.), any of the corresponding published document can be cited in patents. For this reason, a table of patent equivalents of the cited patent documents is provided for the three publishing authorities considered. Equivalents help to better account for citations received by particular patents and thus assess, for example, the relative value of inventions. The structure of the database builds on the infrastructure first described in *Webb et al. (2005)*.

The *EPO's Worldwide Statistical Patent Database* (PATSTAT Global, Spring 2022) is the primary data source: PATSTAT database providing an extensive set of information on citations made in patents filed worldwide.

DATASET COVERAGE

The database covers all citations present in EPO and PCT (WO) patent documents published from 1978 onwards (98% of EP and PCT patents contain citations), and all citations made in USPTO patent grants from 1976 (backward citations are only available for USPTO granted patents). For each citation, the origin of the citation is indicated as well as the EPO search codes allocated (for EPO or PCT citations). Patent equivalents have been identified for 2,095,835 EPO patents, 1,577,317 PCT filings and 3,365,845 USPTO patents. Citations to non-patent literature (NPL) in EPO, PCT patents and USPTO grants have been filtered to avoid double counts: references to databases (e.g. *Patent Abstracts of Japan* and *WPI Database)* were removed from the dataset. Explicit links to NPL references cited in related patent documents (e.g. NPL text as "*See references of ...*") were used to consolidate the list of NPL citations.

Summary counts of backward citations (number of citations made) and forward citations (number of citations received at the same office) *during the search* were elaborated for EPO patents and USPTO patents. Note that it is possible to combine EPO and PCT datasets using the correspondance table between PCT and EPO for all PCT filings that entered the regional phase at the EPO.

Citation origin			
APP	Citation introduced by the **applicant**	**OPP**	The real opposition documents (citations) selected by the **opposition** division
SEA	Citation introduced during the **search** (from search report)	**APL**	Citations introduced **when filed for appeal** by applicant/proprietor/patentee
ISR	Citations from the **International Search Report**	**FOP**	When an opposition has been filed: citations **introduced by the opponent** or the proprietor
SUP	Citations from the **Supplementary Search Report**	**TPO**	Citations introduced because of **Third Party Observations** (Art 115 EPC)
PRS	"**PRe-Search**" citations (before official publication; only for US applications)	**CH2**	Citations introduced **during the Chapter 2 phase of the PCT**
EXA	Citation introduced during the **examination**		

Source: PATSTAT Data Catalog, EPO, Spring 2022.

REFERENCES

Webb, C., H. Dernis, D. Harhoff and K. Hoisl (2005) "Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks", STI Working Paper 2005/9, OECD, Paris.

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

Martinez, C. (2010) "Insight into different types of patent families", STI Working Paper 2010/2, OECD, Paris.

EPO search codes	
X	**Particularly relevant documents** when taken alone (prejudicing novelty)
I	**Particularly relevant documents** when taken alone (prejudicing inventive step) - introduced in April 2011
Y	**Particularly relevant documents** if combined with another document of the same category.
A	**Documents defining the general state of the art** and not prejudicing novelty or inventive step (not belonging to X, I or Y)
O	**Documents which refer to non-written disclosure**
P	**Intermediate documents** - documents published between the date of filing of the application being examined and the date of priority claimed
T	**Documents relating to the theory or principle underlying the invention** (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)
E	**Potentially conflicting documents** – 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
D	**Documents cited in the application** (i.e. already mentioned in the description of the patent application)
L	**Documents cited for other reasons** (e.g. a document that may throw doubt on a priority claim)
&	**Document member of the same patent family** – citing a so-called " &-document"

Note: Category I was introduced from April 2011 (publication date of the citing patent).
The former X category was split up into 2 categories, X and I. Up to three codes can be allocated (e.g. AD, XD, XP, YP, APD, XPD).

Source: EPO, DOCDB XML Exchange Format, August 2019.

DATABASE STRUCTURE

The OECD Citations database is derived from the following PATSTAT tables: CITATION, CITN_CATEG and NPL_PUBLN, combined with APPLN and PAT_PUBLN tables. If required, all patent documents included in the OECD citations database can be connected to PATSTAT (Spring 2022), using the *Appln_id* surrogate key. The patent equivalents are defined as explained in *Martinez C. (2010)*. Note that data on patent applicants and inventors, as well as the list of *International Patent Classification (IPC)* codes allocated to citing and cited patents, are no longer included in the citation database but can be recollected using the PATSTAT database.

The data are provided in distinct tables (flat files using UTF-8 format, columns separated using the pipe "|" character). Any feedback on the data (structure, content) would be highly appreciated.

RESTRICTIONS, SOURCE & CONTACT

Please note that the raw data on OECD Citations database are provided for research and analytical work. When publishing the results of your analysis, please refer to : ***"OECD, Citations database, August 2022"***.

For further information about the OECD patent project, methodological documentation and access to pre-defined patent indicators *(patent counts by technologies, by country, by regions...)*, please visit our web page at oe.cd/ipstats.

Comments and questions about this dataset should be sent to STI.Microdatalab@oecd.org.
For further information on EPO's PATSTAT, please contact patstat@epo.org.

DATABASE STRUCTURE 1/2

EPO_CITATIONS — *12,965,763 rows*
Citations made in EPO patents

Citing_pub_nbr	EPO patent publication number
Citing_pub_date	Earliest publication date (YYYYMMDD)
Citing_app_nbr	EPO patent application number
Citing_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
Cited_pub_nbr	Cited patent - publication number
Cited_pub_date	Earliest publication date of cited patent (YYYYMMDD)
Cited_app_auth	Application authority of cited patent
Cited_app_nbr	Cited patent - application number
Cited_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
Cit_Total	Total number of citations made in the citing patent document (≥ 0)
Citn_Origin	Origin of the citation [see table p.1]
Citn_Category	EPO search codes
Citn_lag_year	Citation lag in years (Citing_pub_date - Cited_pub_date)
Citn_lag_month	Citation lag in months (Citing_pub_date - Cited_pub_date)
PCT_route	=1 if Euro-PCT patent

PCT_CITATIONS — *22,927,600 rows*
Citations made in PCT patents

Citing_pub_nbr	PCT patent publication number
Citing_pub_date	Earliest publication date (YYYYMMDD)
Citing_app_nbr	PCT patent application number
Citing_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
Cited_pub_nbr	Cited patent - publication number
Cited_pub_date	Earliest publication date of cited patent (YYYYMMDD)
Cited_app_auth	Application authority of cited patent
Cited_app_nbr	Cited patent - application number
Cited_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
Cit_Total	Total number of citations made in the citing patent document (≥ 0)
Citn_Origin	Origin of the citation [see table p.1]
Citn_Category	EPO search codes
Citn_lag_year	Citation lag in years (Citing_pub_date - Cited_pub_date)
Citn_lag_month	Citation lag in months (Citing_pub_date - Cited_pub_date)
ISA	International Search Authority

US_CITATIONS — *194,899,008 rows (split into 5 files, by cited date)*
Citations made in USPTO patents[2]

Citing_pub_nbr	US patent publication number
Citing_pub_date	Earliest publication date (YYYYMMDD)
Citing_app_nbr	USPTO patent application number
Citing_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
Cited_pub_nbr	Cited patent - publication number
Cited_pub_date	Earliest publication date of cited patent (YYYYMMDD)
Cited_app_auth	Application authority of cited patent
Cited_app_nbr	Cited patent - application number
Cited_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
Cit_Total	Total number of citations made in the citing patent document (≥ 0)
Citn_Origin	Origin of the citation [see table p.1]
Citn_Category	search codes - *seldom available*
Citn_lag_year	Citation lag in years (Citing_pub_date - Cited_pub_date)
Citn_lag_month	Citation lag in months (Citing_pub_date - Cited_pub_date)

EPO_EQUIVALENT — *3,602,294 rows*
List of EPO patent equivalents to cited patents

Cited_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022 *(cited documents with EPO equivalent)*
EP_eqv_appln_id*	EPO link - Surrogate key - applications in PATSTAT, Spring 2022
Eqv_app_nbr	EPO patent application number
Eqv_pub_nbr	EPO patent publication number
Eqv_total	Total number of EPO equivalents per cited patent

PCT_EQUIVALENT — *2,309,883 rows*
List of PCT patent equivalents to cited patents

Cited_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022 *(cited documents with PCT equivalent)*
WO_eqv_appln_id*	PCT link - Surrogate key - applications in PATSTAT, Spring 2022
Eqv_app_nbr	PCT patent application number
Eqv_pub_nbr	PCT patent publication number
Eqv_total	Total number of PCT equivalents per cited patent

US_EQUIVALENT — *7,175,957 rows*
List of USPTO patent equivalents to cited patents

Cited_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022 *(cited documents with USPTO equivalent)*
US_eqv_appln_id*	USPTO link - Surrogate key - applications in PATSTAT, Spring 2022
Eqv_app_nbr	USPTO patent application number
Eqv_pub_nbr	USPTO patent publication number
Eqv_total	Total number of USPTO equivalents per cited patent

EPO_NPL_CITATIONS — *3,562,086 rows*
Citations of Non-Patent Literature[1]

Citing_pub_nbr	EPO patent publication number
Citing_pub_date	Earliest publication date (YYYYMMDD)
Citing_app_nbr	EPO patent application number
Citing_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
NPL_publn_id*	Surrogate key - NPL publication identifier PATSTAT, Spring 2022
Cited_NPL_Nbr	XP code where available (extracted from NPL text)
Same_NPL_as	Same references as other patent document (EPO or WO)
NPL_text	Full non patent literature description
NPL_cit_total	Total number of NPL citations made in the citing patent document
Citn_Origin	Origin of the citation [see table p.1]
Citn_Category	EPO search code

PCT_NPL_CITATIONS — *5,529,314 rows*
Citations of Non-Patent Literature[1]

Citing_pub_nbr	PCT patent publication number
Citing_pub_date	Earliest publication date (YYYYMMDD)
Citing_app_nbr	PCT patent application number
Citing_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
NPL_publn_id*	Surrogate key - NPL publication identifier PATSTAT, Spring 2022
Cited_NPL_Nbr	XP code where available (extracted from NPL text)
Same_NPL_as	Same references as other patent document (EPO or WO)
NPL_text	Full non patent literature description
NPL_cit_total	Total number of NPL citations made in the citing patent document
Citn_Origin	Origin of the citation [see table p.1]
Citn_Category	EPO search code
ISA	For PCT patents : International Search Authority

US_NPL_CITATIONS — *37,791,573 rows (split into 2 files, by citing date)*
Citations of Non-Patent Literature[1]

Citing_pub_nbr	US patent publication number
Citing_pub_date	Earliest publication date (YYYYMMDD)
Citing_app_nbr	USPTO patent application number
Citing_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022
NPL_publn_id*	Surrogate key - NPL publication identifier PATSTAT, Spring 2022
Cited_NPL_Nbr	XP code where available (extracted from NPL text)
Same_NPL_as	Same references as other patent document
NPL_text	Full non patent literature description
NPL_cit_total	Total number of NPL citations made in the citing patent document
Citn_Origin	Origin of the citation [see table p.1]
Citn_Category	search codes - *seldom available*

1. Exclude citations to *Patent Abstracts of Japan*, *WPI Database* and other non relevant NPL references.
2. Due to size limitation, only patents with references to patents are provided in the table, The whole list of patent publications is provided in US_Cit_Counts table (see p.3).

DATABASE STRUCTURE 1/2

EPO_CIT_COUNTS Summary table - count[1] of backward & forward citations of EP patents		***2,597,090 rows***
EP_pub_nbr	EPO patent publication number	
EP_pub_date	Earliest publication date (YYYYMMDD)	
EP_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022	
WO_pub_nbr	PCT patent publication number (WOYYYYNNNNNN) - for Euro-PCT	
WO_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022	
Grant_date	Grant date at the EPO	
Withdrawal	= 1 if the patent was withdrawn or deemed withdrawn	
Refusal	= 1 if the patent was refused	
EP_Pat_Cits	Number of patent citations made in European search	[1]
EP_NPL_Cits	Number of references to non-patent literature (NPL) in European search	[2]
WO_Pat_Cits	Number of patent citations made in international search	[3]
WO_NPL_Cits	Number of references to non-patent literature (NPL) in international search	[4]
Total_Pat_Cits	Number of all patent citations made [1]+[3]	[5]
Total_NPL_Cits	Number of all non-patent literature citations made [2]+[4]	[6]
Total_Cits	All citations made [5]+[6]	[7]
EP_Pat_XCits	Number of patents cited as X made in European search	[8]
EP_NPL_XCits	Number of references to non-patent literature (NPL) cited as X in European search	[9]
WO_Pat_XCits	Number of patents cited as X made in international search	[10]
WO_NPL_XCits	Number of references to non-patent literature (NPL) cited as X in international search	[11]
Total_Pat_XCits	Number of all patents cited as X made [8]+[10]	[12]
Total_NPL_XCits	Number of all non-patent literature citations as X made [9]+[11]	[13]
Total_XCits	All citations made as X [12]+[13]	[14]
Recd_asEP_byEP	Number of citations received as EP publication in European searches	[15]
Recd_asWO_byEP	Number of citations received as WO publication in European searches	[16]
Recd_asEP_byWO	Number of citations received as EP publication in international searches	[17]
Recd_asWO_byWO	Number of citations received as WO publication in international searches	[18]
Direct_cits_Recd	Total citations received as either EP or WO publication [15]+[16]+[17]+[18]	[19]
Recd_asEQV_byEP	Number of citations received as other national (or regional) publication in European searches	[20]
Recd_asEQV_byWO	Number of citations received as other national (or regional) publication in international searches	[21]
Total_cits_Recd	All citations received [19]+[20]+[21]	[22]
Recd_in3_asEP	Number of citations received within 3 years, as EP publication	[23]
Recd_in3_asWO	Number of citations received within 3 years, as WO publication	[24]
Direct_cits_Recd_in3	Total citations received within 3 years as either EP or WO publication [23]+[24]	[25]
Recd_in3_asEQV	Number of citations received within 3 years, as other national (or regional) publication	[26]
Total_cits_Recd_in3	All citations received [25]+[26]	[27]

Backward citations in EP patents *citations made during the search*

Forward citations of EP patents *citations received, directly as EP, as WO or as equivalents*

1. Counts are based on citations made during the search only - Citn_origin = SEA or ISR

US_CIT_COUNTS Summary table - count[1] of backward & forward citations of USPTO patents		***8,741,002 rows***
US_pub_nbr	USPTO patent publication number	
US_pub_date	Earliest publication date (YYYYMMDD)	
US_appln_id*	Surrogate key - applications in PATSTAT, Spring 2022	
Grant_date	Grant date at the USPTO	
US_Pat_Cits	Number of patent citations made during the search	[1]
US_NPL_Cits	Number of references to non-patent literature (NPL) during the search	[2]
Total_Cits	All citations made [1]+[2]	[3]
Direct_cits_Recd	Total citations received as US patent document	[4]
Recd_asEQV	Number of citations received as other national (or regional) publication during the search	[5]
Total_cits_Recd	All citations received [4]+[5]	[6]
Direct_cits_Recd_in3	Total citations received within 3 years as US patent document	[7]
Recd_in3_asEQV	Number of citations received within 3 years, as other national (or regional) publication	[8]
Total_cits_Recd_in3	All citations received [7]+[8]	[9]

Backward

Forward