

Data Catalog

PATSTAT Biblio

PATSTAT Legal Status

Author:	European Patent Office
Distribution:	EPO PATSTAT customers
PATSTAT Edition	2016 Autumn Edition
Version:	5.08
Date:	01.10.2016

DATA CATALOG for PATSTAT Biblio and PATSTAT Legal Status

The European Patent Office has, on behalf of the OECD Taskforce on Patent Statistics, prepared a database designed to assist in statistical research into patent information.

This database is called the PATSTAT. It is distributed from the Vienna sub-office of the European Patent Office by the Publication department.

Please direct enquiries about the database to patstat@epo.org. The discussion forum can be found in <http://forums.epo.org/patstat/> .

Further information can be found on the PATSTAT information pages of the EPO web site at <http://www.epo.org/searching-for-patents/business/patstat.html> .

The EPO is an active member of the Patent Statistics Task Force led by the Organisation for Economic Co-operation and Development (OECD). Other members are the World Intellectual Property Organisation (WIPO), the Japanese Patent Office (JPO), the US Patent and Trademark Office (USPTO), Korean Intellectual Property Office (KIPO), the US National Science Foundation (NSF) and European Commission (EC). The EC is represented by Eurostat and by DG Research.

Document update history

Version	Date	Author	Explanation
0.3	31-05-2005	Heijna	First version
1.0	01-06-2005	Heijna	
2.01	23-06-2005	Heijna	
2.02	07-07-2005	Heijna	Improved relation types
2.1	15-07-2005	Heijna	Technical relations added, Citations re-modelled
2.11	20-07-2005	Heijna	Replenishment explained more clearly
2.2	25-07-2005	Heijna	Classes added
2.21	25-07-2005	Heijna	Improved business rules tables 001 and 010
2.22	26-07-2005	Heijna	Duplicate publication ID processing described, Primary keys highlighted, continuation type definition improved
2.23	02-08-2005	Heijna	IPC re-modelled, extracted citations definition improved
2.24	08-08-2005	Heijna	Various clarifications
3.0	09-09-2005	Rollinson	Enhancement of lay-out and format
		Heijna	Extension with physical model
3.1	13-09-2005	Heijna	Physical model refinements
3.11	22-09-2005	Heijna	Change in physical model
3.12	05-10-2005	Heijna	Textual changes and small improvements
3.13	07-10-2005	Heijna	Physical model diagram
3.2	21-11-2005	Heijna	Citation model upgraded, Conceptual model broken out
3.21	10-03-2006	Heijna	Comments added
3.22	20-04-2006	Rollinson	Foreword, Implementation rules, Directions for use added
3.23	08-09-2006	Heijna	Detailed improvements
3.24	02-05-2007	Rollinson	Updated "Directions for use", internet links (EPO site was rebuilt, many links changed)
3.25	27-08-2007	Rollinson	Added table for Classifications ICO-Y01N; US Applicant addresses ; improved country code coverage; added IPC advanced symbols; changed IPC symbol order to 'as-in' DOCDB ;
3.26	31-10-2007	Lingua	Revision of text
3.27	01-04-2008	Rollinson	Added table for DOCDB simple Family;
3.28	14-05-2008	Lingua	Small revision of text
3.29	23-09-2008	Rollinson/ Lingua	Added new table INPADOC family; redefined priorities with LMI=A and kind code W as equivalent to PARIS priorities.
3.30	15-10-2008	Lingua	Revision of text
3.31	16-04-2009	Lingua	Added new element PUBLN_FIRST_GRANT
3.32	22-05-2009	Lingua	Revision of text

4.00	17-06-2009	Rollinson	Major revision - replaced DOCDB database by DOCDB XML Exchange product as source. Removed access authorisation from element descriptions. Amended IPR_TYPE to include Design Patents, such as US with kind code 'S1'. References in the EPO Supplementary Search Reports to the original WO Search Report publication are now included as NPL citations.
4.06	14-07-2009	Rollinson/ Lingua	Text revised, updated diagrams, links, tables, element description
4.07	21-09-2009	Lingua	Adaptation to September 2009 edition
4.08	15-10-2009	Lingua	Renamed TLS217_APPLN_I_CLS with nanotech to TLS217_APPLN_ECLA as ECLA codes and all ICO codes (with nanotech) are included. Updated after Specs meeting with contractor 29th September.
4.09	18-02-2010	Lingua	Adaptation to April 2010 edition. Changes in loading IPC classes; reduction of artificial applications "D2" originating from to citations.
4.10	11-10-2010	Lingua	Routine for element PUBLN_FIRST_GRANT replaced by source from DOCDB XML, table TLS_221_INPADOC_PRS added, new citation kinds added in TLS_212_CITATION. Adaptation to September 2010 edition.
4.22	12-04-2011	Lingua	New permanent unique application identifier introduced in APPLN_ID, IPC CORE Level symbols no longer maintained by WIPO.
4.31	11-10-2011	Lingua	Adaptation to October 2011 edition, including creation of additional tables TLS222 and TLS223 for JP and US national classifications, adding elements CITN_GENER_AUTH and CITED_APPLN_ID in TLS212 and PUBLN_CLAIMS in TLS211
4.40	13-04-2012	Lingua	Adaptation to April 2012 edition
4.41	10-10-2012	Lingua	Adaptation to October 2012 edition
4.50	02-04-2013	Kracker	Complete revision of this document; Integration of physical model into the logical model; Deletion of table TLS217_APPLN_ECLA; Addition of TLS224_APPLN_CPC; Re-formulation of the domains of attributes in an DBMS-independent way; Various corrections and clarifications; Changes in data model (leading spaces removed from attributes APPLN_NR and PUBLN-NR)
5.00	14-10-2013	Kracker	Adaptation to October 2013 edition; Description of all elements of table TLS221_INPADOC_PRS; Inclusion of the extensions of PATSTAT Online; Minor clarifications and corrections;
5.01	01-04-2014	Kracker	Adaptation to 2014 Spring Edition; Various clarifications and minor corrections;

5.01.01	02-06-2014	Kracker	Correction of Diagram Section 3.2 "Logical Model Diagram"
5.02	15-10-2014	Kracker	Adaption to 2014 Autumn Edition: See section 7 "History of major changes to tables and attributes". Clarification of stability of certain IDs: See section "4.3.2 Stable IDs". Various other clarifications and minor corrections; Corrected errors in Logical Model Diagram (§3.2); Added chapter 8 "Known Deficiencies"
5.03	01-04-2015	Kracker	Adaption to 2015 Spring Edition: See section 7 "History of major changes to tables and attributes". Pre-computed attributes which are only available in PATSAT Online: Their computation is specified in the new section "SQL scripts for computed attributes". Various other clarifications and minor corrections;
5.04	01-10-2015	Kracker	Adaption to 2015 Autumn Edition: See section 7 "History of major changes to tables and attributes". Section "SQL scripts for computed attributes" has been removed, because differences between the database schemas of PATSTAT Raw Data and PATSTAT Online data model have been removed.
5.05	01-12-2015	Kracker	Adaption to 2015 Spring - Amended Edition: See section 7 "History of major changes to tables and attributes".
5.06	08-02-2016	Kracker	URLs to the EPO homepage have changed; Minor clarifications
5.07	01-04-2016	Kracker	Adaption to 2016 Spring Edition: See section 7 "History of major changes to tables and attributes".
5.08	01-10-2016	Kracker	Adaption to 2016 Autumn Edition: See section 7 "History of major changes to tables and attributes".

Table of contents

1	Introduction	13
2	Domain model	18
3	Logical model	21
4	Design principles	24
5	Table description	37
6	Attribute description	75
7	History of major changes to tables and attributes	279
8	Known deficiencies	284

Detailed Table of contents

1	Introduction.....	13
1.1	About this document.....	13
1.2	About PATSTAT.....	13
1.3	PATSTAT product line.....	13
1.4	Data currentness and coverage.....	14
1.5	Data sources.....	14
1.5.1	DOCDB bibliographic data.....	14
1.5.2	Data sources for legal events.....	14
1.5.3	Data sources for person data.....	14
1.5.4	Data sources for harmonized names.....	15
1.5.5	Data sources for number of claims.....	15
1.6	Other databases for statistical purposes.....	16
1.7	Suggested reading on patent statistics.....	16
1.8	Correct citation of PATSTAT; copyright and trademark.....	16
1.9	Disclaimer.....	16
1.10	Help desk, discussion forum, feedback.....	17
2	Domain model.....	18
2.1	Overview.....	18
2.2	Applicants and inventors.....	18
2.3	Priorities, national / regional phases, continuations, technical relations.....	19
2.4	Citations and publications.....	19
2.5	Classifications.....	20
2.6	Families.....	20
3	Logical model.....	21
3.1	Table naming convention.....	21
3.2	Logical model diagram.....	22
4	Design principles.....	24
4.1	Handling of double quotes and line breaks.....	24
4.2	Handling of missing or unknown values.....	24
4.3	Surrogate database keys.....	24
4.3.1	Pro and cons.....	24
4.3.2	Stable IDs.....	25
4.4	Application replenishment.....	26
4.4.1	Application replenishment for priorities.....	27
4.4.2	Application replenishment for citations.....	27
4.4.3	Allocating the APPLN_ID.....	28
4.5	Publication replenishment.....	30
4.5.1	Publications.....	30
4.5.2	Publication replenishment for citations.....	30
4.6	Relation types (Paris Convention priorities, continuations, etc.).....	31
4.6.1	Rules.....	32
4.6.2	Continuation types.....	34
5	Table description.....	37
5.1	TLS201_APPLN: Application.....	37
5.2	TLS202_APPLN_TITLE: Application title.....	40

5.3	TLS203_APPLN_ABSTR: Application abstract	41
5.4	TLS204_APPLN_PRIOR: Paris convention priority	42
5.5	TLS205_TECH_REL: Technical relation	43
5.6	TLS206_PERSON / TLS906_PERSON: Person	44
5.7	TLS207_PERS_APPLN: Link between Person and Application	46
5.8	TLS209_APPLN_IPC: International Patent Classification	47
5.9	TLS210_APPLN_N_CLS: National classification	49
5.10	TLS211_PAT_PUBLN: Patent publication	50
5.11	TLS212_CITATION: Citation	51
5.12	TLS214_NPL_PUBLN: Non patent literature publication	55
5.13	TLS215_CITN_CATEG: Citation category	56
5.14	TLS216_APPLN_CONTN: Application continuation	57
5.15	TLS221_INPADOC_PRS: Legal event PATSTAT Legal Status	58
5.16	TLS222_APPLN_JP_CLASS: Japanese classification	60
5.17	TLS223_APPLN_DOCUS: US classification	61
5.18	TLS224_APPLN_CPC: Cooperative Patent Classification	62
5.19	TLS226_PERSON_ORIG: Unmodified person data	63
5.20	TLS227_PERS_PUBLN: Link between person and publication	65
5.21	TLS228_DOCDB_FAM_CITN: Citation between DOCDB families	66
5.22	TLS229_APPLN_NACE2: NACE2 industry classification	67
5.23	TLS230_APPLN_TECHN_FIELD: Classification by technical field	68
5.24	TLS801_COUNTRY: Reference table of country codes	69
5.25	TLS802_LEGAL_EVENT_CODE: Reference table of legal event codes PATSTAT Legal Status	70
5.26	TLS901_TECHN_FIELD_IPC: Mapping between technology fields and IPC	71
5.27	TLS902_IPC_NACE2: Mapping between IPC and industrial sectors	72
5.28	TLS904_NUTS: NUTS regional codes	73
5.29	TLS906_PERSON: Person	74
6	Attribute description	75
6.1	Explanation of attribute description	75
6.2	ADDRESS_1, ADDRESS_2, ADDRESS_3, ADDRESS_4, ADDRESS_5	76
6.3	ADDRESS_FREEFORM	78
6.4	APPLN_ABSTRACT	80
6.5	APPLN_ABSTRACT_LG	81
6.6	APPLN_AUTH	82
6.7	APPLN_FILING_DATE	84
6.8	APPLN_FILING_YEAR	86
6.9	APPLN_ID	87
6.10	APPLN_KIND	89
6.11	APPLN_NR	91
6.12	APPLN_NR_EPODOC	93
6.13	APPLN_NR_ORIGINAL	95
6.14	APPLN_TITLE	96
6.15	APPLN_TITLE_LG	97
6.16	APPLT_SEQ_NR	98
6.17	AUTH_CC PATSTAT Legal Status	100
6.18	CITED_APPLN_ID	101
6.19	Note that this first <patcit> element can be a reference to a patent <i>publication</i> (then see section 6.22 "CITED_PAT_PUBLN_IDCITED_NPL_PUBLN_ID	102

6.20	CITED_DOCDB_FAMILY_ID	103
6.21	CITED_NPL_PUBLN_ID	104
6.22	CITED_PAT_PUBLN_ID	105
6.23	CITN_CATEG.....	109
6.24	CITN_GENER_AUTH.....	111
6.25	CITN_ID	113
6.26	CITN_ORIGIN	114
6.27	CITY	116
6.28	CONTINENT.....	117
6.29	CONTN_TYPE	118
6.30	CPC_CLASS_SYMBOL	120
6.31	CPC_GENER_AUTH	122
6.32	CPC_POSITION.....	123
6.33	CPC_SCHEME.....	125
6.34	CPC_VALUE	126
6.35	CPC_VERSION.....	127
6.36	CTRY_CODE	128
6.37	DISCONTINUED	129
6.38	DOCDB_FAMILY_ID.....	130
6.39	DOCDB_FAMILY_SIZE.....	132
6.40	DOC_STD_NAME	133
6.41	DOC_STD_NAME_ID	134
6.42	DOCUS_CLASS_SYMBOL.....	135
6.43	EARLIEST_FILING_DATE	136
6.44	EARLIEST_FILING_ID.....	137
6.45	EARLIEST_FILING_YEAR.....	138
6.46	EARLIEST_PAT_PUBLN_ID.....	139
6.47	EARLIEST_PUBLN_DATE	140
6.48	EARLIEST_PUBLN_YEAR	141
6.49	EPO_MEMBER	142
6.50	EU_MEMBER.....	143
6.51	FIRST_NAME.....	144
6.52	GRANTED.....	145
6.53	HAN_HARMONIZED.....	146
6.54	HAN_ID	147
6.55	HAN_NAME.....	148
6.56	IMPACT PATSTAT Legal Status	149
6.57	INPADOC_FAMILY_ID.....	150
6.58	INT_PHASE.....	152
6.59	INTERNAT_APPLN_ID	154
6.60	INVT_SEQ_NR.....	156
6.61	IPC	158
6.62	IPC_CLASS_LEVEL.....	159
6.63	IPC_CLASS_SYMBOL.....	160
6.64	IPC_GENER_AUTH.....	162
6.65	IPC_MAINGROUP_SYMBOL.....	163
6.66	IPC_POSITION	164
6.67	IPC_VALUE.....	165
6.68	IPC_VERSION	166
6.69	IPR_TYPE	167

6.70	ISO_ALPHA3.....	168
6.71	JP_CLASS_SCHEME	169
6.72	JP_CLASS_SYMBOL.....	170
6.73	L501EP PATSTAT Legal Status.....	171
6.74	L502EP PATSTAT Legal Status.....	172
6.75	L503EP PATSTAT Legal Status.....	173
6.76	L504EP PATSTAT Legal Status.....	174
6.77	L505EP PATSTAT Legal Status.....	175
6.78	L506EP PATSTAT Legal Status.....	176
6.79	L507EP PATSTAT Legal Status.....	177
6.80	L508EP PATSTAT Legal Status.....	178
6.81	L509EP PATSTAT Legal Status.....	179
6.82	L510EP PATSTAT Legal Status.....	180
6.83	L511EP PATSTAT Legal Status.....	181
6.84	L512EP PATSTAT Legal Status.....	182
6.85	L513EP PATSTAT Legal Status.....	183
6.86	L515EP PATSTAT Legal Status.....	184
6.87	L516EP PATSTAT Legal Status.....	185
6.88	L517EP PATSTAT Legal Status.....	186
6.89	L518EP PATSTAT Legal Status.....	187
6.90	L519EP PATSTAT Legal Status.....	188
6.91	L520EP PATSTAT Legal Status.....	189
6.92	L522EP PATSTAT Legal Status.....	190
6.93	L523EP PATSTAT Legal Status.....	191
6.94	L524EP PATSTAT Legal Status.....	192
6.95	L525EP PATSTAT Legal Status.....	193
6.96	LAST_NAME	194
6.97	LEC_DESCR PATSTAT Legal Status.....	195
6.98	LEC_ID PATSTAT Legal Status.....	196
6.99	LEC_NAME PATSTAT Legal Status.....	197
6.100	LECG_DESCR PATSTAT Legal Status.....	198
6.101	LECG_ID PATSTAT Legal Status.....	199
6.102	LECG_NAME PATSTAT Legal Status.....	200
6.103	MIDDLE_NAME.....	201
6.104	NACE2_CODE	202
6.105	NACE2_DESCR	203
6.106	NACE2_WEIGHT	204
6.107	NAME_FREEFORM.....	205
6.108	NAT_AUTH_CC PATSTAT Legal Status.....	208
6.109	NAT_LEC_NAME PATSTAT Legal Status.....	209
6.110	NAT_CLASS_SYMBOL.....	210
6.111	NAT_PHASE	211
6.112	NB_APPLICANTS	212
6.113	NB_CITING_DOCDB_FAM.....	213

6.114	NB_INVENTORS.....	214
6.115	NOT_WITH_IPC.....	215
6.116	NPL_BIBLIO.....	216
6.117	NPL_CITN_SEQ_NR.....	218
6.118	NPL_PUBLN_ID.....	219
6.119	NPL_TYPE.....	221
6.120	NUTS.....	223
6.121	NUTS_LEVEL.....	225
6.122	NUTS3.....	226
6.123	NUTS3_NAME.....	227
6.124	OECD_MEMBER.....	228
6.125	PARENT_APPLN_ID.....	229
6.126	PAT_CITN_SEQ_NR.....	231
6.127	PAT_PUBLN_ID.....	232
6.128	PERSON_ADDRESS.....	233
6.129	PERSON_CTRY_CODE.....	235
6.130	PERSON_ID.....	238
6.131	PERSON_NAME.....	239
6.132	PERSON_ORIG_ID.....	241
6.133	PRIOR_APPLN_ID.....	242
6.134	PRIOR_APPLN_SEQ_NR.....	243
6.135	PRS_CODE PATSTAT Legal Status	244
6.136	PRS_EVENT_SEQ_NR PATSTAT Legal Status	246
6.137	PRS_GAZETTE_DATE PATSTAT Legal Status	247
6.138	PSN_ID.....	248
6.139	PSN_LEVEL.....	249
6.140	PSN_NAME.....	250
6.141	PSN_SECTOR.....	251
6.142	PUBLN_AUTH.....	252
6.143	PUBLN_CLAIMS.....	253
6.144	PUBLN_DATE.....	255
6.145	PUBLN_FIRST_GRANT.....	256
6.146	PUBLN_KIND.....	258
6.147	PUBLN_LG.....	260
6.148	PUBLN_NR.....	261
6.149	PUBLN_NR_ORIGINAL.....	262
6.150	REG_PHASE.....	263
6.151	RESIDENCE_CTRY_CODE.....	264
6.152	ROLE.....	265
6.153	SOURCE.....	266
6.154	SOURCE_VERSION.....	267
6.155	ST3_NAME.....	268
6.156	STATE.....	269
6.157	STATE_INDICATOR.....	270
6.158	STREET.....	271
6.159	TECH_REL_APPLN_ID.....	272
6.160	TECHN_FIELD.....	273
6.161	TECHN_FIELD_NR.....	274
6.162	TECHN_SECTOR.....	275
6.163	UNLESS_WITH_IPC.....	276

6.164	WEIGHT	277
6.165	ZIP_CODE.....	278
7	History of major changes to tables and attributes.....	279
8	Known deficiencies.....	284

1 INTRODUCTION

1.1 About this document

This document describes the structure of the PATSTAT Biblio and the PATSTAT Legal Status databases. It contains diagrams showing the high level structures, business rules, design principles as well as detailed description of the tables and attributes.

1.2 About PATSTAT

Patent statistics are used as indicators of the inventive activity of companies or countries, and as indicators of the patent system itself. While the patent activity rose substantially over the last decades, the demand for patent data and statistics followed the same trend. Demand is increasing to monitor this development to better understand the innovation process and in the end to support policy decisions.

The EPO is an active member of the Patent Statistics Task Force led by the Organisation for Economic Co-operation and Development (OECD). Other members are the World Intellectual Property Organisation (WIPO), the Japanese Patent Office (JPO), the US Patent and Trademark Office (USPTO), Korean Intellectual Property Office (KIPO), the US National Science Foundation (NSF) and European Commission (EC). The EC is represented by Eurostat and by DG Research.

Upon request of the Task Force, the EPO has created PATSTAT as the backbone data set for statistical analysis.

1.3 PATSTAT product line

PATSTAT consists of 3 individual products (<http://www.epo.org/searching-for-patents/business/patstat.html>), tag “Getting started” and “Conditions”, of which 2 parts are covered in this document:

- **PATSTAT Biblio:**
This is the core of PATSTAT (EPO product 14.24). It has a worldwide coverage and contains bibliographic information about applications and publications.
Most of this document deals with this data.
- **PATSTAT Legal Status:**
This is another component to PATSTAT Raw Data which may be additionally subscribed (EPO product 14.24.1). It, too, has a worldwide coverage and contains legal information about patents.
It is derived from the INPADOC EPO worldwide legal status database. This data is marked in this document by **PATSTAT Legal Status**

- **PATSTAT Register:**
It contains detailed bibliographic, procedural and legal event information for EP patents (EPO product 14.24.2).
This data is not described in this document, but in the PATSTAT Register Data Catalog.

1.4 Data currentness and coverage

This statistical database is a 'snapshot' of the source databases at a single point in time. Therefore if you double-check specific details against the register of a specific patent granting authority and differences are detected at the data level such as different names or dates, please consider that the authority may have corrected their databases in the time since this statistical database 'snapshot' was made.

Typically, the date of data extraction from the source databases is end of January for the PATSTAT Spring Edition and end of July for the PATSTAT Autumn Edition.

The back-files of the two main databases DOCDB EPO worldwide bibliographic database and INPADOC EPO worldwide legal status database used to produce the 2016 Autumn Edition contain all publications present in the databases at the end of week 2016/31 for DOCDB as well as for Legal Status.

Information about the coverage of EPO's master bibliographic database DOCDB and the INPADOC EPO worldwide legal status database can be found in <http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/weekly.html> .

1.5 Data sources

1.5.1 DOCDB bibliographic data

Much of the raw data is extracted from the DOCDB EPO worldwide bibliographic database. The manual for this database is called 'EXCHANGE FORMAT EPO - Patent Information Resource' and can be downloaded from the EPO internet site as DOCDB User Documentation (<http://www.epo.org/searching-for-patents/technical/docdb.html>), tab "Further information").

1.5.2 Data sources for legal events

For table TLS221_INPADOC_PRS the raw data are extracted from the INPADOC worldwide legal status database, (<http://www.epo.org/searching-for-patents/legal/inpadoc.html>).

1.5.3 Data sources for person data

For improved quality, names and addresses of applicants / owners and inventors are taken from several sources which are described in detail below.

1.5.3.1 EPO data

The person data for the EPO applications is taken from the EP Patent Register data.

The EP Patent Register data holds also historical information. This allows PATSTAT to link the applicants and inventors to old publications, even if these persons have changed in the meantime.

1.5.3.2 USPTO data

The US data for names and addresses for published grants patents published after 1976-01-01 is taken from the USPTO's patent databases, as published weekly on their ftp site.

Starting with the publications of September 29th 2005, we also take the names and address data for published applications from USPTO's patent database.

The US data for names and addresses for patents published before 1976-01-01 (published grants) and September 25th 2005 (published applications) is taken from the EPO's DOCDB Database.

1.5.3.3 DOCDB

DOCDB data is used for all non-EP, non-US applications and for US applications before 1976.

1.5.4 Data sources for harmonized names

There are several types of harmonized names available:

- DOCDB Standardized Name:
This name is retrieved from DOCDB.
- PATSTAT Standardized Name (PSN):
This effort for harmonizing names and allocation of assignee sectors is done by ECOOM (K.U. LEUVEN; <http://www.ecoom.be/en/EEE-PPAT>).
- OECD HAN:
The Harmonized Applicant Name computed by OECD is taken from <http://www.oecd.org/sti/innovationinsciencetechnologyandindustry/oecdpatentdatabases.htm>.

1.5.5 Data sources for number of claims

The number of claims for EPO and US publications are taken from special data feeds directly from EP and USPTO.

1.6 Other databases for statistical purposes

Almost all national and regional (EPO, EAPO, ARIPO, OAPI, ...) patent offices offer online internet access to their registers. These are needed if you wish to conduct in-depth research on a single national or regional patent granting authority.

The EPO offers a range of on-line databases for EP patent data you may use for further analysis or verification of your findings. These can be found on the EPO homepage.

- Free products
 - European Patent Register
 - European Publication Server
 - Espacenet
 - EP full-text search
- Subscription products
 - Global Patent Index
 - EP Bulletin search

1.7 Suggested reading on patent statistics

For a thorough introduction to patent statistics, you are recommended to consult the "OECD Patent Statistics Manual 2009"

(http://www.oecd.org/document/29/0,3343,en_2649_34451_42168029_1_1_1_1,00.html).

1.8 Correct citation of PATSTAT; copyright and trademark

If you publish analyses based on this statistical database, please cite the source of the data including the name of the current version, e. g. 'PATSTAT - 2016 Autumn Edition'

The copyright to this database as distributed by the EPO remains with the EPO. "PATSTAT" is a registered trademark.

1.9 Disclaimer

The data in the PATSTAT databases is based on other EPO databases and on data provided to EPO on a voluntary basis by national and supranational patent authorities. EPO actively seeks to create and maintain a high quality data basis for PATSTAT but cannot assume any legal liability or responsibility for the accuracy or completeness of the database.

In case legal certainty, accurate or complete data is needed, EPO strongly suggests contacting the competent patent authorities. Almost all national and regional patent offices offer online Internet access to their registers. These should be used to conduct in-depth research on a single national or regional patent granting authority.

EPO would appreciate if users of PATSTAT reported deficiencies to patstat@epo.org so that appropriate measures may be taken to correct the deficiencies and to develop PATSTAT further.

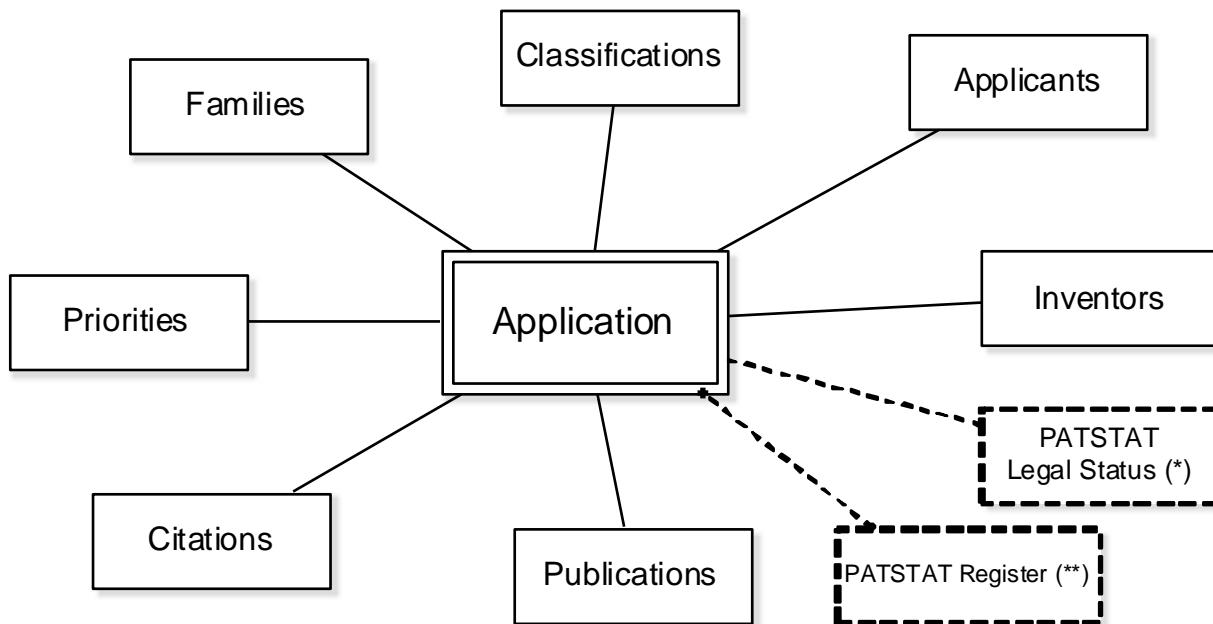
1.10 Help desk, discussion forum, feedback

A help desk is available at patstat@epo.org. The discussion forum can be found in <http://forums.epo.org/patstat/> .

Your feedback is very valuable and welcome. Please report any errors or suggestions for improvement to patstat@epo.org.

2 DOMAIN MODEL

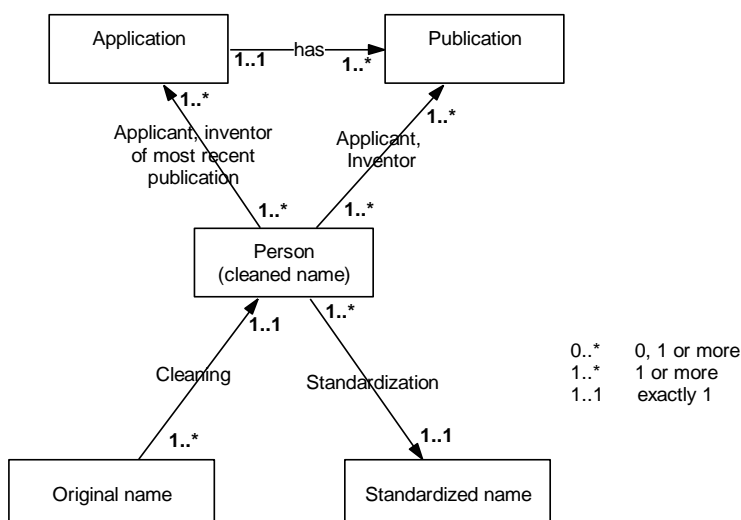
2.1 Overview



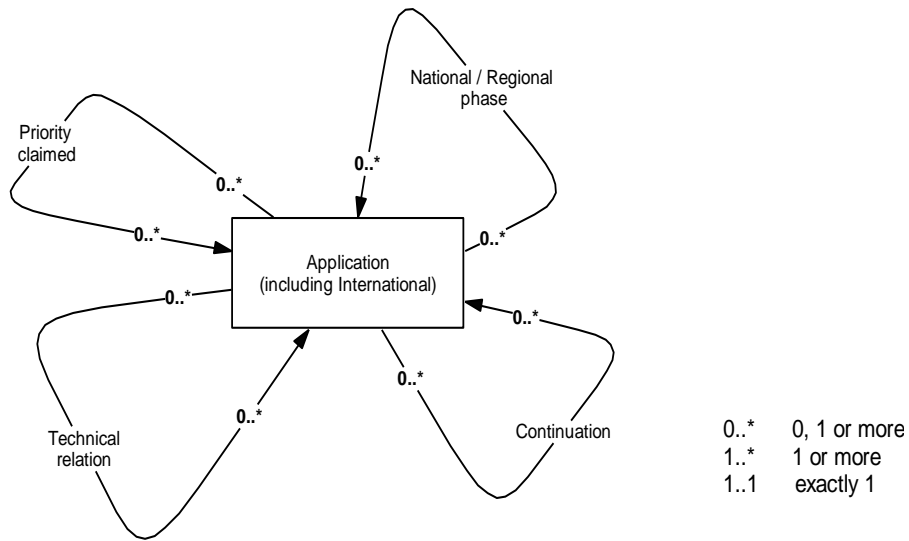
(*) This table must be purchased separately (Product 14.24.1: See tab "Conditions" [here](#))

(**) These tables must be purchased separately (Product 14.24.2; See tab "Conditions" [here](#)). They are described in a separate document (EP Register for PATSTAT Data Catalog)

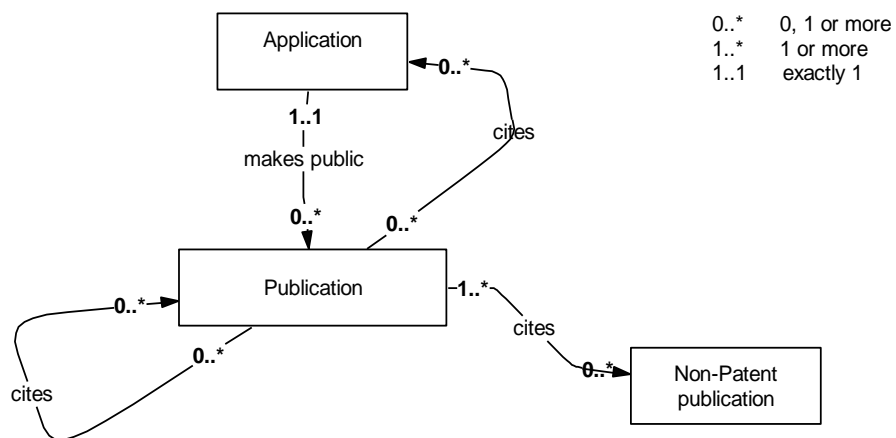
2.2 Applicants and inventors



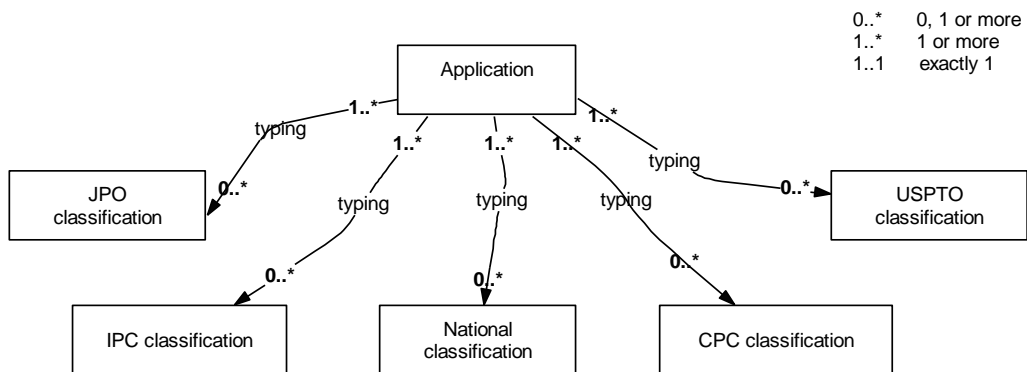
2.3 Priorities, national / regional phases, continuations, technical relations



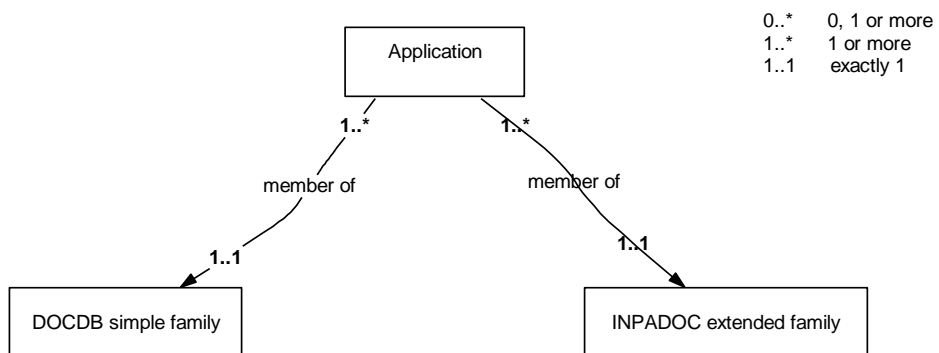
2.4 Citations and publications



2.5 Classifications



2.6 Families



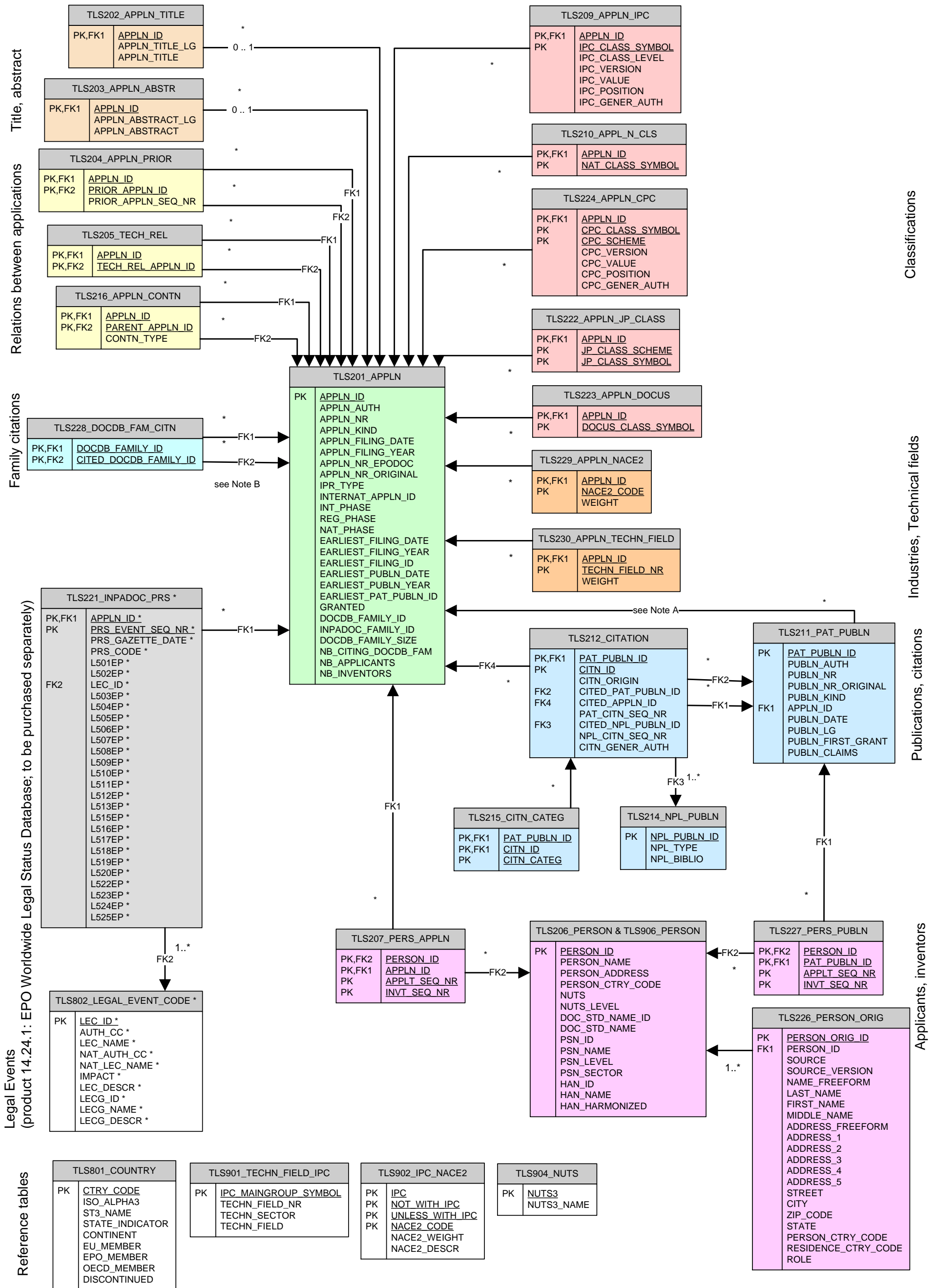
3 LOGICAL MODEL

3.1 Table naming convention

All table names in PATSTAT are of the form **TL**Snnn_xxxx

- TLS fixed prefix
- nnn unique number; the range of the number indicates its purpose;
200 range: data table;
800 range: reference table with data primarily managed by the EPO;
900 range: reference table with data primarily from external sources;
- _ underscore
- xxxx one or more words indicating the content of the table

3.2 Logical model diagram



Legend:

0..1 cardinality

* cardinality 0 ... n

PK This attribute is (part of) the Primary Key

FKn This attribute is (part of) a Foreign Key FK;

Underlined attributes constitute the Primary Key.

Note A:

Depending on their number range, some applications will or will not have publications. See section 4.4 "Application replenishment".

Applications with APPLN_ID	Number of Publications	cf. Range in section 4.4
0 - 900 000 000	1 or more	1
900 000 001 - 930 000 000	0	2
930 000 001 - 960 000 000	1 or more	3
960 000 001 - 999 999 999	0	4

Note B:

Both attributes in table TLS228_DOCDB_FAM_CITN must be linked to the attribute DOCDB_FAMILY_ID (and not APPLN_ID) of table TLS201_APPLN.

4 DESIGN PRINCIPLES

4.1 Handling of double quotes and line breaks

Double quotes (") are consistently replaced by single quotes (') in the data. This makes importing the data file, where double quotes are used to delimit text but will not appear within a text, easier.

Line breaking sequences (CR, LF) within strings are replaced by " \n ". This sequence e.g. occurs in the abstract text (attribute APPLN_ABSTRACT of table TLS203_APPLN_ABSTR) or the bibliographic data of non-patent literature (attribute NPL_BIBLIO in table TLS214_NPL_PUBLN).

4.2 Handling of missing or unknown values

It is a fact that for several documents, usually old ones, we are missing data, e.g. filing dates. In relational databases missing values are usually represented by NULL values, but these are hard to exchange in csv format.

PATSTAT data does not contain any NULL values and in fact all attributes may be defined as NOT NULL. Depending on the domain, PATSTAT represents missing values like this:

- Missing **date** values are represented as '9999-12-31'.
- Missing **string** values are represented as zero length strings (like "") or as fixed length strings containing spaces.
- Missing **numerical** values are represented as number zero.

It is important to understand that this date 9999-12-31, in your research findings, means 'unknown date'.

So if you write a query for all patents published after 2008-01-01, you will get a far larger number than you expect - you will also get all the 9999-12-31 ones. So remember to write something like

```
'..where PUBLN_DATE > '2008-01-01' and PUBLN_DATE < '9999-12-31' .
```

4.3 Surrogate database keys

4.3.1 Pro and cons

A database 'key' is a shorthand reference to an entity (e. g. a publication, an application or a person) in a database. They uniquely identify such an entity. The key is just a number, which is meaningless to the human user, because it does not correspond to any number in the business world, like an application number or publication number. Therefore these surrogate keys are also called *technical identifiers*.

In PATSTAT you can easily recognize these technical identifiers because their names end with a *_ID* (not to be confused with names ending with *_NR*). There are several such technical identifiers, like:

- APPLN_ID
- INTERNAT_APPLN_ID
- PRIOR_APPLN_ID
- TECH_REL_APPLN_ID
- PERSON_ID
- DOC_STD_NAME_ID
- PAT_PUBLN_ID
- CITN_ID
- NPL_PUBLN_ID
- CITED_PAT_PUBLN_ID
- PARENT_APPLN_ID
- DOCDB_FAMILY_ID
- INPADOC_FAMILY_ID
- LEC_ID

The **advantage** of a surrogate key is that it is generally easier and more efficient to identify entities or to join tables with a surrogate key than with business identifiers.

As an example: The PUBLN_ID 20194422 identifies the application FR833507D with kind code A. Using business identifiers, you will need 3 attributes to uniquely identify this application: PUBLN_AUTH = FR, PUBLN_NR=833507D and APPLN_KIND=A. Also, from a technical point of view, the database can store and manage surrogate keys more efficiently, which results in smaller database sizes and faster queries.

The **disadvantage** of a surrogate key is twofold:

- The identifier itself does not have any business meaning. You will always have to retrieve additional business data to be able to read your result.
- The identifier can and will change between two editions of PATSTAT. While the French patent application mentioned above will always have the same business identifiers "FR", "833507D" and "A", its surrogate key will change, e. g. from the April 2008 edition to the October 2008 edition. So if you want to exchange data between different PATSTAT editions, be sure to align the data not via surrogate keys but via business attributes.

There are a few notable exceptions: For example, since April 2011, the surrogate key for applications, the APPLN_ID, will remain stable. Also, the PAT_PUBLN_ID, the surrogate key for publications, does not change between different PASTAT editions. The list of stable IDs and mode information can be found in the next section 4.3.2 "Stable IDs".

4.3.2 Stable IDs

The following attributes will not change between different PATSTAT editions, although in rare cases there may be exceptions. Therefore these attributes can be used to link data from different PATSTAT editions. Moreover, the attributes APPLN_ID, PAT_PUBLN_ID and DOCDB_FAMILY_ID are taken directly from DOCDB, so they can also be easily be linked to other databases which are based on DOCDB.

- APPLN_ID (since April 2011 edition, within Range 1 (= not replenished applications))
- DOCDB_FAMILY_ID
- PERSON_ID (from Oct 2013 edition onwards)
- PERSON_ORIG_ID (from Oct 2013 edition onwards)
- PAT_PUBLN_ID (from 2014 Autumn Edition onwards, within Range 1 (= not replenished publications))

Exceptions to this stability assurance:

- APPLN_ID (within Range 1):
These values are taken from DOCDB, so all restrictions of DOCDB apply:
In instances where an application-reference has been re-keyed – technically speaking – the value of the APPLN_ID will remain stable and unchanged. There may be situations however – particularly as a result of on-line intellectual intervention – where this cannot be guaranteed.
- PAT_PUBLN_ID (within Range 1):
These values are taken from DOCDB. The value of the PAT_PUBLN_ID is guaranteed to be stable, including in any event where the publication-identifier is corrected – also when the publication kind code has been corrected.
- DOCDB_FAMILY_ID:
These values are taken from DOCDB, so all restrictions of DOCDB apply:
The family-identifier is unique within the database, once used it will never be re-used, but its value cannot be guaranteed to be stable.
- PERSON_ID, PERSON_ORIG_ID:
These attributes are the unique keys of person tables TLS206_PERSON and TLS226_PERSON_ORIG. These tables are also supposed to have no duplicates in their non-key values.
However, in exceptional cases, e. g. due to data cleaning, duplicates may occur. These duplicates might be removed in later PATSTAT editions. This will result in IDs which have been available in older editions, but not in newer editions. However, once used, these IDs will never be re-used.

4.4 Application replenishment

Artificial applications are added to PATSTAT to manage doubt about applications which have not been captured in the DOCDB database from which PATSTAT is built. There are several cases:

- Application replenishment for priorities
- Application replenishment from citations
 - Applications originating from cited publications
 - Applications originating from cited applications

4.4.1 Application replenishment for priorities

By “priority” we here mean not only “Paris Convention priority”, but also other types of priorities which link one application to a “prior” application. The various types of priorities are stored in separate tables:

- TLS201_APPLN
An PCT application in its regional/national phase contains in its attribute INTERNAT_APPLN_ID the APPLN_ID of its original PCT application
- TLS204_APPLN_PRIOR
contains Paris Convention priorities
- TLS205_Tech_REL
contains links between technically equivalent applications
- TLS216_APPLN_CONTN
contains various relations like continuations, divisional applications, ...

There are cases where an application is claimed as priority, but this application is not known to DOCDB. Then we nevertheless assume that this prior application does really exist, although for some reason it is not in DOCDB. Therefore, we will create an artificial prior application in PATSTAT.

This can mean for example that you might find an application in table TLS201_APPLN, but not in EPO's search engine Espacenet as an application. However, you will find it in Espacenet if you search for it as a priority document.

Typically, these artificial applications are applications which have been withdrawn or abandoned before publication, but which the applicant has used as a priority, or in America, for continuation.

In more technical terms: If in the DOCDB backfile the application which is claimed as a priority in priority-claims for `data-format="docdb"` has no corresponding application-reference in DOCDB, then an artificial application must be created.

The example shows how the attributes of these artificial applications are populated:

- APPLN_AUTH is set to `<country>US</country>` from `priority-claim`
- APPLN_NR is set to `sequence="1"` from `priority-claim` and
- APPLN_KIND is set to `<kind>A</kind>` from `priority-claim`,
all for `data-format="docdb"` .
- APPLN_FILING_DATE is set to `<date>20040802</date>` from `priority-claim`
- APPLN_ID: Allocate a unique value incrementally, starting at 900 000 001.

4.4.2 Application replenishment for citations

There are two categories of replenished applications originating from citations:

- a) Applications originating from cited publications
- b) Applications originating from cited applications

4.4.2.1 a) Applications originating from cited publications

There are cited publications for which there is no publication reference in DOCDB. This includes cited patents which were extracted from Non Patent Literature NPL citations. In this case an artificial publication is created in PATSTAT (see section 4.5 Publication

replenishment). And we also create a matching application (see this section), because every publication must be assigned to an application.

The following business rules are applied:

Check if the cited publication has a publication-reference in DOCDB. If not, then create an artificial publication and an artificial application. The attribution of the artificial application is:

- APPLN_AUTH identical to the PUBLN_AUTH of the cited publication.
- APPLN_NR identical to the PUBLN_NR of the cited publication.
- APPLN_KIND = 'D2'.
- APPLN_FILING_DATE = '9999-12-31'
- IPR_TYPE = 'PI'
- APPLN_ID: Allocate a unique value incrementally, starting at 930 000 001.

4.4.2.2 b) Applications originating from cited applications

There are cited *applications* (in contrast to case a) publications) for which there is no application reference in DOCDB. In this case an artificial application is created in PATSTAT.

The following business rules are applied:

Check if the cited application has an application-reference in DOCDB and if not, then create an artificial application. The attribution of the artificial application is:

- APPLN_AUTH identical to the APPLN_AUTH of the cited application
- APPLN_NR identical to the APPLN_NR of the cited application
- APPLN_KIND identical to the APPLN_KIND of the cited application; if not given then use "D3". Note that in 2014 Autumn Edition there was no occurrence of 'D3'.
- APPLN_FILING_DATE identical to the APPLN_FILING_DATE of the cited application, if not given then assign '9999-12-31':

If the same artificial application is cited more than once and with different application filing dates, then the earliest application filing date will be replenished. (Note: This logic minimizes the replenishment with the default date '9999-12-31').

- APPLN_ID: Allocate a unique value incrementally, starting at 960 000 001.

4.4.3 Allocating the APPLN_ID

When collecting all applications, priorities and cited documents for all of the publications registered in DOCDB, it is important to keep them separate.

Once all of the application-references, publication-references, priority-claims and cited-references have been collected, it is possible to start allocating the surrogate key values for APPLN_ID and PUBLN_ID for the artificial applications and publications.

Starting with the April 2011 edition, the DOCDB "doc-id" unique and stable identifier has been used to populate APPLN_ID for non-replenished applications instead of creating an own surrogate key. This attribute remains the same across PATSTAT editions and always refers to the same combination of application authority, application number and application kind. It has a numeric value of max. 9 digits.

There are for ranges of replenished artificial applications in PATSTAT.
 (Note: range 1 is for non-replenished applications, also called non-artificial applications;
 these applications are registered in DOCDB and their APPLN_ID is below 900 000 000):

Range 2:

Artificial applications created in PATSTAT for priorities applications which are not registered in DOCDB. They are not stable across PATSTAT editions.

Range: 900 000 001 to 930 000 000

Range 3:

Artificial applications created in PATSTAT for applications originating from cited publications not recorded in DOCDB. They are not stable across PATSTAT editions.

Range 930 000 001 to 960 000 000, kind code "D2".

Range 4:

Artificial applications created in PATSTAT for applications originating from cited applications not recorded in DOCDB. They are not stable across PATSTAT editions.

Range 960 000 001 to 999 999 999, using the kind code "D3" if the citation given has no kind code.

PATSTAT edition	Number of applications in DOCDB with a subsequent publication	Ranges of APPLN_ID for artificial applications		
		Range 2: PATSTAT applications created from unpublished DOCDB priorities	Range 3: PATSTAT applications created from cited publications with no publication in DOCDB	Range 4: PATSTAT applications created from cited applications with no application in DOCDB
2016 Autumn	78 351 200	900 000 001- 907 615 814	930 000 001 – 931 830 595	960 000 001 – 960 015 376
2016 Spring	76 504 845	900 000 001 - 907 526 532	930 000 001 - 931 744 552	960 000 001 - 960 014 259
2015 Aut. Amend	74 469 830	900 000 001 - 907 427 548	930 000 001 - 931 738 471	960 000 001 - 960 013 617
2015 Autumn	74 469 830	900 000 001 - 907 427 548	930 000 001 - 931 728 217	960 000 001 - 960 012 596
2015 Spring	72 642 820	900 000 001 - 907 335 048	930 000 001 - 931 721 184	960 000 001 - 960 013 317
2014 Autumn	71 081 761	900 000 001 - 907 230 282	930 000 001 - 931 700 084	960 000 001 - 960 013 768
2014 Spring	69 410 835	900 000 001 - 907 140 127	930 000 001 - 931 724 340	960 000 001 - 960 013 546
2013 Oct	67 766 435	900 000 001 - 907 099 488	930 000 001 - 931 714 237	960 000 001 - 960 014 115
2013 April	66 012 696	900 000 001 - 906 989 695	930 000 001 - 931 755 005	960 000 001 - 960 014 651
2012 Oct	64 571 194	900 000 001 - 906 913 465	907 000 001 - 908 677 881	909 000 001 - 909 014 510
2012 April	63 280 409	900 000 001 - 906 826 996	907 000 001 - 908 669 845	909 000 001 - 909 014 916
2011 Oct	61 570 794	900 000 001 - 906 561 807	907 000 001 - 908 550 321	909 000 001 - 909 010 181
2011 April	60 312 074	900 000 001 - 906 476 936	907 000 001 - 908 692 290	not applicable
2010 Oct	58 713 013	59 000 001 - 65 239 596	66 000 001 - 67 274 345	not applicable
2010 April	57 505 125	58 000 000 - 63 983 731	64 000 001 - 65 252 476	not applicable
2009 Sept	56 420 849	57 000 001 - 62 913 743	63 000 001 - 64 260 712	not applicable
2009 April	55 517 602	56 000 001 - 61 951 472	62 000 001 - 63 239 563	not applicable
2008 Sept	54 371 495	55 000 001 - 60 883 933	61 000 001 - 62 241 942	not applicable
2008 April	53 357 975	54 000 001 - 59 803 164	60 000 001 - 61 238 598	not applicable
2007 Oct	52 389 958	53 000 001 - 58 670 414	59 000 001 - 60 232 649	not applicable
2007 April	?	52 000 001 - 57 616 300	58 000 001 - 60 447 086	not applicable
2006 Sept	?	?	?	not applicable

2006 April	?	50 000 001 - 55 527 619	56 000 001 - 58 541 387	not applicable
------------	---	-------------------------	-------------------------	----------------

Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	20-07-2005	First version
J. Rollinson	June 2009	Source changed to XML DOCDB
D. Lingua	19-02-2010	Inserted comment on n. of D2s
D. Lingua	31-03-2011	Changes due to introduction of "doc-id"
D. Lingua	04-08-2011	Introduction of cited applications
M. Kracker	15-03-2013	Introduction of fixed ranges; clarifications
M. Kracker	01-04-2014	Clarifications

4.5 Publication replenishment

4.5.1 Publications

The EPO maintains a database called DOCDB (also known as Patent Information Resource) covering over 90 countries. The database contains patent documents and utility model documents which have been published or laid open to public inspection. These documents are a 'snapshot' of the status of an application at various stages in the lifecycle of the processing of the application according to the law of the relevant Office. Typically the contents of the application are published sometime after the priority date, in the EPO this is 18 months. If a search report is available, it is published at the same time. At later stages in the lifecycle, such as grant, the contents of the application are published again, possibly in amended form. The different publication events in the lifecycle of the processing of an application are distinguished by the system of Kind of Publication Codes as laid down in the publication kind code concordance list (<http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/regular.html>) for databases within the EPO in column "DOCDB".

4.5.2 Publication replenishment for citations

When a document is cited, it is checked whether this document is already in the database by comparing the patent authority (country), the document number and the document kind code.

However in roughly 2% of the cited documents in table TLS212_CITATION there is no corresponding publication entry in the table of published documents TLS211_PAT_PUBLN. This means that we cannot be 100% certain which document is intended to be cited.

Even if a cited publication is not known to DOCDB, we assume this document does really exist because it has been cited. Therefore, in these cases we introduce artificial publications in table TLS211_PAT_PUBLN. The attribution of an artificial publication is:

- PUBLN_AUTH, PUBLN_NR and PUBLN_KIND are taken from the citation
- PUBLN_DATE is assigned '9999-12-31', if no publication date is given.

- PUBLN_ID: Allocate a unique value incrementally, starting at 900 000 001. The PUBLN_ID number range 900 000 001 to 999 999 999 is exclusively reserved for artificial publications.

We also create artificial applications to match these artificial publications (see section 4.4.2.1a) Applications originating from cited publications (Range 3).

PATSTAT edition	Number of Publications in DOCDB	Range of PAT_PUBLN_ID for artificial publications: PATSTAT publications created from DOCDB cited publications with no publication in DOCDB
2016 Autumn	98 592 257	900 000 001 - 901 899 315
2016 Spring	96 044 918	900 000 001 - 901 805 460
2015 Autumn Amended	93 276 814	900 000 001 - 901 795 268
2015 Autumn	93 276 814	900 000 001 - 901 784 222
2015 Spring	90 812 863	900 000 001 - 901 775 950
2014 Autumn	88 725 979	900 000 001 - 901 752 404
2014 Spring	86 430 793	900 000 001 - 901 724 340
2013 Oct	84 019 544	900 000 001 - 901 714 237
2013 April	81 694 203	900 000 001 - 901 755 005
2012 Oct	80 883 905	81 000 001 - 82 677 881
2012 April	79 049 630	80 000 001 - 81 669 845
2011 Oct	76 817 848	77 000 001 - 78 550 321
2011 April	74 274 345	75 000 001 - 76 692 290
2010 Oct	72 887 199	73 000 001 - 74 274 345
2010 April	71 217 622	72 000 001 - 73 252 476
2009 Sept	69 711 942	70 000 001 - 71 260 712
2009 April	68 453 166	69 000 001 - 70 239 563
2008 Sept	66 946 928	67 000 001 - 68 241 942
2008 April	65 493 394	66 000 001 - 67 238 598
2007 Oct	64 132 954	65 000 001 - 66 232 649
2007 April	?	63 000 001 - 65 447 086
2006 Sept	?	61 000 001 - 63 541 387
2006 April	?	60 000 000 – 62 426 270

Modification history		
Author of update	Date of update	Explanation of update
M. Kracker	15-03-2013	First version

4.6 Relation types (Paris Convention priorities, continuations, etc.)

Applications may be linked in various ways to other applications. This section describes how to distinguish 6 cases of these link types. 4 out of these 6 cases are stored in PATSTAT.

Relation Type	Explanation	stored in PATSTAT table
case # 1	self-priority: ignored	--
case # 2	Paris Convention priority	TLS204_APPLN_PRIOR
case # 3	national/regional phase of international application	TLS201_APPLN
case # 4	change of IPR-type claimed: ignored	--
case # 5	technical relation (see note below)	TLS205_TECH_REL
case # 6	domestic continuation (e. g. divisional, US continuations, ...)	TLS216_CONTN

Note for case # 5 (Technical relations):

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 priority-linkage-type, also known as Link Method Indicator LMI, to "T" for 'Technical'.

The "T" indicator has allowed extracting 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, especially before 1990, do not have the indicator set to "T", thus appearing in PATSTAT as a Paris convention priority.

4.6.1 Rules

This section describes the rules to infer the relation type (also known as LMI or Link Method Indicator or Linkage Type) from DOCDB XML data.

Note that LMI (Link Method Indicator) is represented in DOCDB XML as `<priority-linkage-type>` element.

This decision tree is applied to `<priority-claim data-format="docdb">` sections of DOCDB:

Is the content of element <code><document-id></code> in <code><priority-claim data-format="docdb"></code> identical to the content of element <code><document-id></code> in <code><application-reference></code> (differences in the <code><date></code> can be ignored)?		
Y	case # 1 Note: <code><priority-linkage-type></code> will have value "Z"	
N	What is the value of <code><priority-linkage-type></code> ?	
	it is missing	case # 2
	W, w	case # 3
	A, a	Is <code><kind> = 'W'</code> ?
	Y	Is this the only priority claimed?
		Y case # 3
		N case # 2

	N	case # 4
	I	case # 2
	U	case # 4
	T	case # 5
	other	case # 6

Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	03-05-2005	First version
R. Heijna	13-07-2005	Continuation type added
R. Heijna	22-07-2005	Table continuation types separated
J. Rollinson	19-09-2008	for LMI=A , APPL_KIND='W' changed to PRIO_KIND='W'
D. Lingua	08-10-2012	Added rule for LMI=A, but APPL_KIND='W'
M. Kracker	26-03-1013	Restructuring; no change in logic

4.6.2 Continuation types

Note that LMI (Link Method Indicator) is represented in DOCDB XML as `<priority-linkage-type>` element.

This table lists several values of `<priority-linkage-type>` as they may occur for certain offices.

Section 4.6.1 "Rules" defines how these `<priority-linkage-type>` elements are mapped to relation types. This table goes one step further and defines for each case #6 (= domestic continuation) the detailed continuation type, which is stored in attribute `CONTN_TYPE` of table `TLS216_APPLN_CONTN` (see the respective table and attribute description).

APPLN_AUTH	LMI	Description	CONTN_TYPE
AT	A	CITED APPLICATION CHANGED FROM PATENT TO UTILITY	-
AT	U	CITED APPLICATION CHANGED FROM UTILITY TO PATENT	-
AU	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
AU	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
BA	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
CA	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
CA	4	PRIOR APPLICATION CLAIMED FOR A CONTINUATION IN PART	CIP
CA	5	CLAIMED APPLICATION IS A SUPPLEMENTARY DISCLOSURE	SUP
CH	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
CH	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
CS	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
CZ	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
DE	1	DOMESTIC PRIOR CLAIMED FOR PATENT	INN
DE	2	DOMESTIC PRIORITY CLAIMED FOR UTILITY MODEL	INN
DE	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
DE	1	PATENT APPLICATION CLAIMED FOR UTILITY MODEL	-
DK	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
EP	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
ES	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
FI	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
FR	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
GB	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
HK	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
HU	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
IE	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
IE	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV

IE	C	COGNATE APPLICATION	CGT
IL	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
IL	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
IN	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
IN	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
IN	C	COGNATE APPLICATION	CGT
JP	1	DOMESTIC PRIORITY	INN
JP	A	CITED APPLICATION CHANGED FROM PATENT TO UTILITY	-
JP	U	CITED APPLICATION CHANGED FROM UTILITY TO PATENT	-
KR	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
KR	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
KR	A	CITED APPLICATION CHANGED FROM PATENT TO UTILITY	-
KR	U	CITED APPLICATION CHANGED FROM UTILITY TO PATENT	-
LU	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
LV	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
MX	A	CITED APPLICATION CHANGED FROM PATENT TO UTILITY	-
MX	U	CITED APPLICATION CHANGED FROM UTILITY TO PATENT	-
NL	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
NO	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
NO	C	COGNATE APPLICATION	CGT
NZ	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
NZ	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
NZ	C	COGNATE APPLICATION	CGT
PH	1	PRIOR APPLICATION CLAIMED FOR A CONTINUATION	CON
PH	2	PRIOR APPLICATION CLAIMED FOR A CONTINUATION IN PART	CIP
PH	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
PL	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
RU, SU	6	DOMESTIC PRIORITY	INN
TW	0	PRIOR APPLICATION CLAIMED FOR AN ADDITION	ADD
US	1	PRIOR APPLICATION CLAIMED FOR A CONTINUATION	CON
US	2	PRIOR APPLICATION CLAIMED FOR A CONTINUATION IN PART	CIP
US	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
US	4	PRIOR APPLICATION CLAIMED FOR A SUBSTITUTE	SBS
US	5	CLAIMED APPLICATION IS ORIGINAL REISSUE SERIAL NUMBER	REI
US	B	ABANDONED APPLICATION CLAIMED FOR A CONTINUATION	CON
US	C	ABANDONED APPLICATION CLAIMED FOR A CONTINUATION IN PART	CIP

US	D	ABANDONED APPLICATION CLAIMED FOR A DIVISION	DIV
US	R	REQUEST FOR REEXAMINATION NUMBER	REI
YU	3	PRIOR APPLICATION CLAIMED FOR A DIVISION	DIV
YU	6	DOMESTIC PRIORITY	INN

Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	22-07-2005	Table continuation types separated
R. Heijna	26-09-2005	Definition based on application authority
M. Kracker	27-06-2013	Table sort order changed

5 TABLE DESCRIPTION

5.1 TLS201_APPLN: Application

Contains the key bibliographical data elements relevant to identify the patent application. Most of the elements in this table can be found on the first page of a printed patent document. E. g.: application authority, application number and application filing date. From a database structure point of view, this table is very important because it links to many other database tables via the application ID attribute.

Note: The following attributes have been renamed in the 2015 Spring Edition:

- PRIOR_EARLIEST_DATE → EARLIEST_FILING_DATE
- PRIOR_EARLIEST_YEAR → EARLIEST_FILING_YEAR
- PUBLN_EARLIEST_DATE → EARLIEST_PUBLN_DATE
- PUBLN_EARLIEST_YEAR → EARLIEST_PUBLN_YEAR

TLS201_APPLN	
Technical identifier (Primary key)	
APPLN_ID	
Business identifiers	
APPLN_AUTH	
APPLN_NR	
APPLN_KIND	
APPLN_FILING_DATE	
APPLN_FILING_YEAR	
APPLN_NR_EPODOC	
APPLN_NR_ORIGINAL	
IPR_TYPE	
INTERNAT_APPLN_ID	
Data about the route of the application	
INT_PHASE	
REG_PHASE	
NAT_PHASE	
Data from priorities	
EARLIEST_FILING_DATE	
EARLIEST_FILING_YEAR	
EARLIEST_FILING_ID	
Data from publications	
EARLIEST_PUBLN_DATE	
EARLIEST_PUBLN_YEAR	
EARLIEST_PAT_PUBLN_ID	
Data derived from publications	
GRANTED	
Family data	
DOCDB_FAMILY_ID	
INPADOC_FAMILY_ID	
DOCDB_FAMILY_SIZE	
NB_CITING_DOCDB_FAM	

Aggregated data	
NB_APPLICANTS	
NB_INVENTORS	
PRIMARY KEY	APPLN_ID
Alternate Key	APPLN_AUTH, APPLN_NR, APPLN_KIND
Business rules	<p>Artificial applications:</p> <p>See section 4.4 "Application replenishment."</p> <p>International Applications:</p> <p>The APPLN_ID refers to the national/regional application; the INTERNAT_APPLN_ID refers to the international application on which it is based. These two foreign keys (applications) are different ones as the international and subsequent national application numbers are different (this is even the case if international office and national/regional office are the same. There is a 1:n relationship as one international application is followed by many national/regional ones in the designated states but one national/regional application relates to one international application only in PATSTAT.</p> <p>Only international applications designating the Authority of the related National /Regional application and for which the latter is published with an INID-code in the 30-series (WIPO ST.9) is included in this table. The case to be taken into account is case #3 from section 4.6 "Relation Types".</p> <p>Known Duplicates</p> <p>Some applications are stored twice in DOCDB and therefore also in PATSTAT. You will have to consider this when you are counting applications. The rule of thumb is:</p> <ul style="list-style-type: none"> • Duplicates on APPLN_AUTH and APPLN_NR - one APPLN_KIND 'A', the other APPLN_KIND 'T': application-identifiers refer to one and the same application • Duplicates on APPLN_AUTH and APPLN_NR - one APPLN_KIND 'A', the other APPLN_KIND 'D': <ul style="list-style-type: none"> ○ for APPLN_AUTH(s) AT, AU, BG, NL and SE: application-identifiers refer to one and the same application ○ for all other APPLN_AUTH(s): application-identifiers refer to two separate applications • Duplicates on APPLN_AUTH and APPLN_NR - one APPLN_KIND 'A' and the APPLN_KIND(s) 'K', 'L', 'M' or 'N' : application-identifiers refer to two separate applications <p>As another rule of thumb: If in doubt which filing to consider for</p>

	counting, count only those which have a publication.	
Comments	A first filing, i.e. an application not claiming the priority of any other application, can be identified by its <i>absence</i> from table TLS204_APPLN_PRIOR. Also, attribute INTERNAT_APPLN_ID in table TLS201_APPLN must be 0 to exclude national / regional phases of PCT applications.	
Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	07-09-2005	First version
R. Heijna	07-10-2005	Continuations broken out
J. Rollinson	02-07-2009	added comment
M. Kracker	26-03-2013	added business rule for known duplicates
M. Kracker	01-10-2013	added PATSTAT Online extension attributes
M. Kracker	01-10-2015	Renaming and re-ordering several attributes. The tables TLS218_DOCDB_FAM and TLS219_INPADOC_FAM have been integrated. The language attributes have been moved to TLS202_APPLN_TITLE and TLS203_APPLN_ABSTR. Several attributes which have been available only in PATSTAT Online are now available in PATSTAT Raw Data as well.
M. Kracker	01-04-2016	New attributes APPLN_NR_ORIGINAL, INT_PHASE, REG_PHASE and NAT_PHASE added.

5.2 TLS202_APPLN_TITLE: Application title

Contains the English title of the application when available. In case there is no English title available, another language will be used if available.

TLS202_APPLN_TITLE			
	APPLN_ID		
	APPLN_TITLE_LG		
	APPLN_TITLE		
	PRIMARY KEY	APPLN_ID	
	FOREIGN KEY	APPLN_ID	REFERENCES TLS201_APPLN (APPLN_ID)
Business rules	Multiple titles may be published for any application, but only one title will be stored in PATSTAT, according to these rules (first applicable rule is applied) <ol style="list-style-type: none"> 1. most recent (according to publication date) title in English 2. most recent title in language of publication 3. most recent title in any other language 		
Comments	n/a		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
M. Kracker	01-10-2015	Attribute APPLN_TITLE_LG moved from table TLS201_APPLN to here	

5.3 TLS203_APPLN_ABSTR: Application abstract

English language abstract when available. If there is no abstract in English, then the most recent abstract in another language is loaded.

TLS203_APPLN_ABSTR			
APPLN_ID			
APPLN_ABSTRACT_LG			
APPLN_ABSTRACT			
PRIMARY KEY	APPLN_ID		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Business rules	<p>Multiple abstracts may be published for any application, but only one abstract will be stored in PATSTAT, according to these rules (first applicable rule is applied)</p> <ol style="list-style-type: none"> 1. most recent (according to publication date) abstract in English 2. most recent abstract in language of publication 3. most recent abstract in any other language 		
Comments	n/a		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
M. Kracker	01-10-2015	Attribute APPLN_ABSTRACT_LG moved from table TLS201_APPLN to here	

5.4 TLS204_APPLN_PRIOR: Paris convention priority

This table contains the Paris Convention priorities of an application.

TLS204_APPLN_PRIOR			
APPLN_ID			
PRIOR_APPLN_ID			
PRIOR_APPLN_SEQ_NR			
PRIMARY KEY	APPLN_ID, PRIOR_APPLN_ID		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
FOREIGN KEY	PRIOR_APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Business rules	<p>APPLN_ID refers to the claiming application; PRIOR_APPLN_ID refers to the application of which the priority is claimed. These two foreign keys (applications) should be different ones, i.e. there is no "self-priority".</p> <p>There is a n:m relationship; multiple priorities may be claimed by one application and one priority may be claimed by multiple applications.</p> <p>Only "pure" priorities i.e. those according to the Paris Convention and published with an INID-code in the 30-series (WIPO ST.9) are included in this table. The case to be taken into account is case # 2 from section 4.6 "Relation Types".</p>		
Comments	<p>PCT applications (published with an INID-code in the 80-series; cf. WIPO ST.9) are no Paris Convention priorities, so they are not included in this table.</p> <p>A first filing, i.e. an application not claiming the priority of any other application, can be identified by its <i>absence</i> from this table Also, attribute INTERNAT_APPLN_ID in table TLS201_APPLN must be 0 to exclude national / regional phases of PCT applications.</p>		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
J. Rollinson	07-07-2009	added comment	
M. Kracker	15-10-2014	changed comment	

5.5 TLS205_TECH_REL: Technical relation

Technical relations are "priority-like" relations between applications which have been detected by EPO examiners but which have not been published by a patent office. From a statistical point of view you should consider them equal to the priority and continuation relations established in TLS204_APPLN_PRIOR and in TLS216_APPLN_CONTN.

TLS205_TECH_REL			
APPLN_ID			
TECH_REL_APPLN_ID			
PRIMARY KEY		APPLN_ID, TECH_REL_APPLN_ID	
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
FOREIGN KEY	TECH_REL_APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Business rules	The case to be taken into account is case # 5 from section 4.6 "Relation Types".		
Comments	<p>Technical relations are symmetric. When an application A and an application B are technically related, then B is also technically related to A. Consequently, if this table has a record A – B, then it will also contain a record B – A.</p> <p>Technical relations are a technical solution to connect old applications, which do not have priority information, into families. These relations are entered when detected by examiners or the EPO bibliographic data experts and no other priority-like relation exists between the applications.</p> <p>Most technical priorities are from FR, US, GB and DE applications, where large old collections, also from before 1900, exist. From 2008 onwards – with the introduction of the patent family building business rules – technical linking is very much the exception.</p> <p>There can however be no guarantee of completeness. This relation is also not published by Patent Offices. You can consider these technical relations as a priority-like relationship.</p>		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
M. Kracker	15.10.2014	Updated comment	
M. Kracker	15.10.2016	Updated comment	

5.6 TLS206_PERSON / TLS906_PERSON: Person

Table that contains the key data on applicants and inventors such as: the person name, the address and the country of residence (which is not necessarily the nationality). Several types of names are available:

- The original name as delivered by the offices
- The name as standardized by the EPO (DOCDB standardized name)
- The PATSTAT standardized name (created by the Catholic Univ. of Leuven / Belgium)
- The name as standardized by the OECD (OECD Harmonized Applicant Name)

TLS206_PERSON / TLS906_PERSON	
PERSON_ID	
Name and address delivered by the offices	
PERSON_NAME	
PERSON_ADDRESS	
PERSON_CTRY_CODE	
Region	
NUTS	In TLS906_PERSON only
NUTS_LEVEL	In TLS906_PERSON only
DOCDB standardized name	
DOC_STD_NAME_ID	
DOC_STD_NAME	
PATSTAT standardized name	
PSN_ID	
PSN_NAME	
PSN_LEVEL	
PSN_SECTOR	
OECD HAN name	
HAN_ID	In TLS906_PERSON only
HAN_NAME	In TLS906_PERSON only
HAN_HARMONIZED	In TLS906_PERSON only
PRIMARY KEY	PERSON_ID
Alternate Key	PERSON_NAME, PERSON_ADDRESS, PERSON_CTRY_CODE
FOREIGN KEY	REFERENCES
Business rules	<p>One and the same person may be recorded in different places in the source files. For some applications the inventor and the applicant may be the same person. Also applicants/inventors may occur in multiple applications.</p> <p>Where the name, address and country of a person in different places in the source files are identical (by case insensitive comparison), they are stored in this table only once. It is very likely although not absolutely sure that one entry in this table represents one (and not more) person in real life. On the other hand it is quite possible that a single person is represented by multiple entries of this table due to variations in name or address or changes of name and address. Several name harmonization efforts try to reduce this ambiguity.</p>
Comments	Persons are the legal or physical persons that have a relation

	<p>with the patent granting procedure. Currently included are 2 roles a person may have: applicants and inventors.</p> <p>TLS906_PERSON contains the same data as TLS206_PERSON plus some additional attributes. TLS906_PERSON has been introduced to explicitly indicate that this table contains data from external sources (see section 3.1 "Table naming convention").</p> <p>To keep its data model simple and consistent with earlier versions, PATSTAT Online uses only TLS206_PERSON, but with all the data of TLS906_PERSON.</p>	
Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	07-09-2005	First version
J. Rollinson	October 2008	corrected column order
J. Rollinson	07-07-2009	extended comments
D. Lingua	08-03-2010	Added comment on US person data
D. Lingua	11-10-2011	Mention of OECD Working Group deleted
M. Kracker	01-10-2013	De-duplication rules changed: Comparison is now case-insensitive; Records with empty fields will be de-duplicated, too. Added PATSTAT Online extension attributes; Added comments on "see applicant" values
M. Kracker	01-04-2014	Removed comments on "see applicant" values
M. Kracker	15-10-2014	Comment updated
M. Kracker	01-04-2015	Attribute DOC_STD_NAME_ID_REPLENISHED has been removed; Order of attributes changed; Comment changed; TLS906_PERSON introduced;
M. Kracker	01-04-2016	Former EEE-PPAT names are now called <u>PATSTAT Standardized Names</u> and have been renamed to PSN_... . Attribute HRM_L1 has been removed.
M. Kracker	01-10-2016	Attributes NUTS and NUTS_LEVEL added

5.7 TLS207_PERS_APPLN: Link between Person and Application

Links the applicants and inventors of the most recent publication to an application. This table can be used to count the number of applicants/inventors per application or to select applicants/inventors by setting the applicant/inventor sequence number.

TLS207_PERS_APPLN			
PERSON_ID			
APPLN_ID			
APPLT_SEQ_NR			
INVT_SEQ_NR			
PRIMARY KEY	PERSON_ID, APPLN_ID, APPLT_SEQ_NR, INVT_SEQ_NR		
FOREIGN KEY	PERSON_ID	REFERENCES	TLS206_PERSON (PERSON_ID)
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Comments	Conceptually, the combination of PERSON_ID and APPLN_ID should be unique. In practice, due to duplicates in the source data also the attributes APPLT_SEQ_NR and INVT_SEQ_NR must be part of the Primary Key.		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
J. Rollinson	18-04-2006	Keys changed.	
M. Kracker	03-07-2013	Clarification of description; Primary Key changed, comment added	

5.8 TLS209_APPLN_IPC: International Patent Classification

The table contains all international patent classifications linked to the applications. The set of classifications linked to a single application is a de-duplicated merge of all classifications of the various publication instances linked to the specific application. Additionally only the latest version of the IPC classifications is used. This means that the user does not have to worry about reclassifications because older applications will always be classified according to the latest IPC version.

TLS209_APPLN_IPC			
	APPLN_ID		
	IPC_CLASS_SYMBOL		
	IPC_CLASS_LEVEL		
	IPC_VERSION		
	IPC_VALUE		
	IPC_POSITION		
	IPC_GENER_AUTH		
	PRIMARY KEY	APPLN_ID, IPC_CLASS_SYMBOL	
	FOREIGN KEY	APPLN_ID	REFERENCES TLS201_APPLN (APPLN_ID)
Business rules	<p>In case the aggregation of the IPC symbols of all publications of a given application contains multiple IPCs with the same IPC_CLASS_SYMBOL, a 2 step de-duplication is performed:</p> <ol style="list-style-type: none"> 1. For a given IPC symbol, only the highest IPC_CLASS_LEVEL is considered: Level A takes precedence over level C, and level C takes precedence over level S. 2. If there still exist multiple IPCs with the same IPC_CLASS_SYMBOL and IPC_CLASS_LEVEL, then the IPC from the latest publication takes precedence. 		
Comments	<p>Information on classification according to the International Patent Classification (IPC) can be found in http://www.wipo.int/classifications/ipc/en/ . The PATSTAT database only contains IPC8 symbols; therefore PATSTAT users do not need to worry about previous IPC classification schemes IPC 1 to 7, when doing statistical analysis based on IPC codes. IPC 1 to 7 have all been reclassified to the latest IPC 8 version.</p> <p>DOCDB contains the MCD Master Classification Database. The MCD has IPC symbols allocated to over 90% of the documents in DOCDB; the remaining 10% older documents are unlikely ever to be classified.</p>		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
R. Heijna	31-10-2005	IPC_GENER_AUTH added	
R. Heijna	31-10-2005	generating office added	
J. Rollinson	18-04-2006	Business rules and contents updated	
J. Rollinson	Aug 2007	IPC Advanced added	

J. Rollinson	Oct 2008	corrected column order
D. Lingua	02-07-2009	Text revised
D. Lingua	21-04-2010	Inserted warning on Core symbols
D. Lingua	05-04-2011	Adapted text due to Core discontinuation
M. Kracker	01-10-2013	Added PATSTAT Online extension attributes
M. Kracker	15-10-2014	Business rules clarified. Comment updated.
M. Kracker	10-10-2015	Removed pre-computed and redundant attributes IPC_SUBCLASS_SYMBOL and TECHN_FIELD_NR.

5.9 TLS210_APPLN_N_CLS: National classification

Some countries (GB, CH, CA, DE, FR, SE, ...) use national patent classification schemes beside the IPC. This table is a list of the national classifications linked to the respective national applications.

TLS210_APPLN_N_CLS			
	APPLN_ID		
	NAT_CLASS_SYMBOL		
	PRIMARY KEY	APPLN_ID, NAT_CLASS_SYMBOL	
	FOREIGN KEY	APPLN_ID	REFERENCES TLS201_APPLN (APPLN_ID)
Business rules	Classes can be present in DOCDB for all publication levels of an application. In PATSTAT these are re-grouped per application. Per application a national class symbol is present only once in PATSTAT unless the content of the source-field is unstructured in which case no de-duplication can be performed. Coverage is weak, particularly since US national classification migrated to table TLS223_APPLN_DOCUS in October 2011. Only a minority of applications in PATSTAT have a national class allotted.		
Comments Rules	Dedicated tables TLS222_APPLN_JP_CLASS and TLS223_APPLN_DOCUS exist for JP and US national classification symbols.		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
R. Heijna	10-03-2006	Business rules extended	
J. Rollinson	07-07-2009	extended comment	
D. Lingua	19-02-2010	Inserted comment	
D. Lingua	04-08-2011	Eliminated national US and JP classification symbols from table	

5.10 TLS211_PAT_PUBLN: Patent publication

Contains the key bibliographical data elements relevant to identify patent publications. These elements can be found on the first page of printed patent documents. For example: publication authority, publication number, publication kind and publication date. This table is directly linked to the TLS201_APPLN table via the appln_id, a surrogate key that groups all the data elements from a single patent application. 2 important extra elements that cannot be found on a patent publication have been added:

- a) The PUBLN_FIRST_GRANT:
indication that this publication was the first indication of a patent grant
- b) PUBLN_CLAIMS:
number of claims (only available for a number of publishing authorities).

TLS211_PAT_PUBLN			
PAT_PUBLN_ID			
PUBLN_AUTH			
PUBLN_NR			
PUBLN_NR_ORIGINAL			
PUBLN_KIND			
APPLN_ID			
PUBLN_DATE			
PUBLN_LG			
PUBLN_FIRST_GRANT			
PUBLN_CLAIMS			
PRIMARY KEY	PAT_PUBLN_ID		
Alternate Key	PUBLN_AUTH, PUBLN_NR, PUBLN_KIND, PUBLN_DATE		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Business rules	This table also includes publications that are not present in DOCDB as a publication-reference but that are in references cited as cited documents, or as corresponding documents. In these cases the publication is artificial. Only the key data of the publication are known, such as authority, publication number and kind code, but not publication date, title or abstract or person data.		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
R. Heijna	30-09-2005	Business rule changed (-> "last")	
R. Heijna	20-04-2006	Business rule extended	
D. Lingua	23-02-2009	PUBLN_FIRST_GRANT element added	
J. Rollinson	2-07-2009	changed from DOCDB to DOCDB XML source	
D. Lingua	04-08-2011	Element PUBLN_CLAIMS added	
M. Kracker	01-04-2014	Correction of comment and foreign key: Multiple occurrences of publications with the same kind code are allowed.	
M. Kracker	01-04-2016	New attribute PUBLN_NR_ORIGINAL.	

5.11 TLS212_CITATION: Citation

Establishes the links between publications, applications and non-patent literature documents with regards to citations. Forward and backward citations are defined as well as the citation generating authority (e.g. search authority) and the procedural step in which the citation was created (e.g. search report or opposition procedure).

TLS212_CITATION			
PAT_PUBLN_ID			
CITN_ID			
CITN_ORIGIN			
Patent literature			
CITED_PAT_PUBLN_ID			
CITED_APPLN_ID			
PAT_CITN_SEQ_NR			
Non-patent literature			
CITED_NPL_PUBLN_ID			
NPL_CITN_SEQ_NR			
Other attributes			
CITN_GENER_AUTH			
PRIMARY KEY	PAT_PUBLN_ID, CITN_ID		
FOREIGN KEY	PAT_PUBLN_ID	REFERENCES	TLS211_PAT_PUBLN (PAT_PUBLN_ID)
FOREIGN KEY	CITED_PAT_PUBLN_ID	REFERENCES	TLS211_PAT_PUBLN (PAT_PUBLN_ID)
FOREIGN KEY	CITED_APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
FOREIGN KEY	CITED_NPL_PUBLN_ID	REFERENCES	TLS214_NPL_PUBLN (NPL_PUBLN_ID)
Business rules	<ul style="list-style-type: none"> • PAT_PUBLN_ID refers to the <i>citing</i> publication. • CITED_PAT_PUBLN_ID refers to a <i>publication</i> being cited. • CITED_APPLN_ID refers to an <i>application</i> being cited. This cited application is <i>not</i> related to a publication cited by CITED_PAT_PUBLN_ID, but is a valid citation on its own. • CITED_NPL_PUBLN_ID refers to a non-patent-literature being cited, which in turn may contain "hidden" references to patent publications. <p>The two foreign keys for patent publications (PAT_PUBLN_ID and CITED_PAT_PUBLN_ID) should be different, i.e. there is no "self-citation".</p> <p>Citations can represent a n:m relationship between publications: multiple publications may be cited in one publication and one publication may be cited by multiple</p>		

others.

3 cases can be distinguished:

- a) Patent citation, which is either a citation of a patent *publication* or patent *application*
- b) Non-Patent Literature citation
- c) Non-Patent Literature citation which refers to a patent *publication* or a patent *application*

These 3 cases can be distinguished by the content of 5 attributes:

	a) Patent citation of a		b) NPL citation	c) NPL citation which refers to a	
	patent publication	patent application		patent publication	patent application
CITED_PAT_PUBLN_ID	>0	=0	=0	>0	=0
CITED_APPLN_ID	=0	>0	=0	=0	>0
PAT_CITN_SEQ_NR	>0	>0	=0	=0	=0
CITED_NPL_PUBLN_ID	=0	=0	>0	>0	>0
NPL_CITN_SEQ_NR	=0	=0	>0	>0	>0

Explanation of case c):
 Patent publication citations *extracted from Non-Patent Literature* are also included, because the entry in table TLS212_CITATION will have the PAT_PUBLN_ID / APPLN_ID of the patent publication / patent application which is 'hidden' in the NPL citation stored in the column CITED_PAT_PUBLN_ID / CITED_APPLN_ID. For these citations the NPL_CITN_SEQ_NR of the relevant NPL-citations from which the patent citation was extracted is filled.

Comments

One publication can cite another publication or application multiple times, if the citation origin (see attribute CITN_ORIGIN) is different. E.g., the applicant and the examiner might cite the same publication, which would result in 2 records, with CITN_ORIG being “APP” resp. “EXA”. If you want to avoid double counting, make sure to count distinct citations only.

A search report for a Euro-PCT application will in most cases *not* repeat the citations which are already in the international search report for the PCT application (cf. “Guidelines for Examination in the EPO”, section X-9.1.4). Consequently, when doing citation analysis of Euro-PCT applications, one must also include the citation of the corresponding PCT application, which can be identified via the attribute INTERNAT_APPLN_ID of the Euro-PCT.
 As this automatic addition of attributes is done by tools like

	<p>Espacenet and EP Register, one cannot directly compare the citation shown by these tools with PATSTAT data.</p> <p>Regular information on citations is available to the EPO from the national patent authorities of the following countries: AP, AT, AU, BE, CZ, DE, EA, EP, ES, FR, GB, IT, JP, NL, SG, US, and WO.</p> <p>Batches of citations are present mainly for: BG, CH, DK, GR, KR, LU, and TR.</p> <p>For a complete and up-to-date coverage information see "Overview of citation data in REFI" (http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/regular.html)</p> <p>Before April 2016 due to a limitation in DOCDB, which is PATSTAT's source database, the number of citations was limited to 99 citations per publication and citation phase (CITN_ORIGIN), unless the citations were provided as "rich" citation. This affected the citations of less than 0.1% of the publications. Almost all of these missing citations were applicant citations of US publications.</p>
--	--

Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	07-09-2005	First version
R. Heijna	13-09-2005	Alternate key added
R. Heijna	20-10-2005	Business rules extended Primary key redefined
R. Heijna	17-11-2005	Alternate key removed
R. Heijna	21-11-2005	Citation model upgraded
J. Rollinson	18-04-2006	Implementation rules added
D. Lingua	04-08-2011	Added elements CITED_APPLN_ID and CITN_GENER_AUTH
J Rollinson	30-04-2009	Patents hidden in NPL are now included in the NPL row in tls212_citation; they no longer have their own row.
J. Rollinson	02-07-2009	removed 5/6 as secret citations are not in DOCDB XML
J. Rollinson	02-07-2009	added WO reference
D. Lingua	13-07-2009	Updated citation information
D. Lingua	04-08-2011	Added elements CITED_APPLN_ID and CITN_GENER_AUTH
D. Lingua	18-04-2012	Updated comment information
M. Kracker	01-12-2013	Reordering of attributes. More detailed business rules
M. Kracker	15-10-2014	Comment added on how to avoid double counts of citations
M. Kracker	15-10-2015	Amended description of business rules
M. Kracker	01-12-2015	Extended case c description of

		business rules; added comment on DOCDBs 99 citations limit
M. Kracker	01-04-2016	Attribute NPL_PUBLN_ID renamed to CITED_NPL_PUBLN_ID. Business rules and comment amended.

5.12 TLS214_NPL_PUBLN: Non patent literature publication

Contains bibliographical information on non-patent literature documents. Typical information contains title and author of the article, title of the publication, an abstract, the publication date, an ECLA classification and sometimes an ISBN, ISSN or DOI number.

TLS214_NPL_PUBLN		
	NPL_PUBLN_ID	
	NPL_TYPE	
	NPL_BIBLIO	
	PRIMARY KEY	NPL_PUBLN_ID
Business rules	From the 2016 Spring Edition onwards NPL citations which contain only strings like "none" or "See also references of WOxxxxxx" are removed to not distort citations counts.	
Comments	The literature which is identified by this description is likely to be copyrighted.	
Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	07-09-2005	First version
R. Heijna	13-09-2005	Primary key redefined
R. Heijna	21-11-2005	Citation model upgraded
J. Rollinson	18-04-2006	Implementation rules added.
J. Rollinson	02-07-2009	added comment
M. Kracker	01-04-2016	New attribute NPL_TYPE. Business rule added.

5.13 TLS215_CITN_CATEG: Citation category

For most citations introduced during the search (citation origin is SEA), a citation category is added to the specific citation. Regular used citation categories are: X, Y and A. For example: category "X" is applicable where a document is such that when taken alone, a claimed invention cannot be considered novel or cannot be considered to involve an inventive step.

TLS215_CITN_CATEG			
	PAT_PUBLN_ID		
	CITN_ID		
	CITN_CATEG		
	PRIMARY KEY	PAT_PUBLN_ID, CITN_ID, CITN_CATEG	
	FOREIGN KEY	(PAT_PUBLN_ID, CITN_ID)	REFERENCES TLS212_CITATION (PAT_PUBLN_ID, CITN_ID)
Business rules	The CITN_ID is a sequence number allocated to each citation made by a single document. The CITN_CATEG is the category of the citation as mentioned in search reports, e.g. X, Y, A, D, P.		
Comments			
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	07-09-2005	First version	
R. Heijna	07-10-2005	Primary key definition	
R. Heijna	21-11-2005	Citation model upgraded	
J. Rollinson	18-04-2006	Implementation rules added.	
J. Rollinson	2-07-2009	added business rule	

5.14 TLS216_APPLN_CONTN: Application continuation

In a similar way as the TLS204_APPLN_PRIOR establishes the priority links between applications, the links between parent and child applications for various types relations such as: continuation (in part), divisional applications, internal priorities are defined via the TLS216_APPLN_CONTN table. Continuation (in part) is generally only applicable to US patent applications. This table should be considered as a priority-like relationship similar to the TLS204_APPLN_PRIOR table.

TLS216_APPLN_CONTN			
APPLN_ID			
PARENT_APPLN_ID			
CONTN_TYPE			
PRIMARY KEY		APPLN_ID, PARENT_APPLN_ID	
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
FOREIGN KEY	PARENT_APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Business rules	<p>APPLN_ID refers to the continuation application; the PARENT_APPLN_ID refers to the application of which the APPLN_ID is a continuation.</p> <p>The two foreign keys (applications) should be different ones, i.e. there is no "self-continuation".</p> <p>There is a n:m relationship so a parent application may have multiple continuations and a continuation can have more than one parent.</p> <p>Only earlier applications for which a continuation is filed with the same authority (domestic) and for which the continuation is published with an INID-code in the 60-series (WIPO ST.9) are included in this table (plus inner priority, INID (23) as used by DE). The case to be taken into account is case # 6 from section 4.6 "Relation Types".</p>		
Comments	Continuations are e. g. divisional applications, additions, continuations in part, ...		
Modification history			
Author of update	Date of update	Explanation of update	
R. Heijna	22-09-2005	First version	

5.15 TLS221_INPADOC_PRS: Legal event **PATSTAT Legal Status**

This table holds the INPADOC Patent Register Service (PRS) data, which contains information on legal events that occurred during the life of a patent, either before or after grant. Typical events are: request for examination, payment of renewal fees, lapse of the patent, change of ownership, withdrawal of the application, patents entering the national phase, patents which have been opposed or revoked, etc.

Events regarding priorities, application filing, coming into force, or publications are not included in this table. These events can, however, be derived from other PATSTAT tables.

TLS221_INPADOC_PRS			
APPLN_ID		PATSTAT Legal Status	
PRS_EVENT_SEQ_NR		PATSTAT Legal Status	
PRS_GAZETTE_DATE		PATSTAT Legal Status	
PRS_CODE		PATSTAT Legal Status	
L501EP		PATSTAT Legal Status	
L502EP		PATSTAT Legal Status	
LEC_ID		PATSTAT Legal Status	
L503EP		PATSTAT Legal Status	
L504EP		PATSTAT Legal Status	
L505EP		PATSTAT Legal Status	
L506EP		PATSTAT Legal Status	
L507EP		PATSTAT Legal Status	
L508EP		PATSTAT Legal Status	
L509EP		PATSTAT Legal Status	
L510EP		PATSTAT Legal Status	
L511EP		PATSTAT Legal Status	
L512EP		PATSTAT Legal Status	
L513EP		PATSTAT Legal Status	
L515EP		PATSTAT Legal Status	
L516EP		PATSTAT Legal Status	
L517EP		PATSTAT Legal Status	
L518EP		PATSTAT Legal Status	
L519EP		PATSTAT Legal Status	
L520EP		PATSTAT Legal Status	
L522EP		PATSTAT Legal Status	
L523EP		PATSTAT Legal Status	
L524EP		PATSTAT Legal Status	
L525EP		PATSTAT Legal Status	
PRIMARY KEY	APPLN_ID, PRS_EVENT_SEQ_NR		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
FOREIGN KEY	LEC_ID	REFERENCES	TLS802_LEGAL_EVENT_CODE (LEC_ID)

Business rules	<p>An entry with PRS_CODE = "PGFP" (Post Grant Fees Paid) indicates that the annual fees have been paid in the national phase of an EP application. Because this type of event typically repeats each year for each EP member state as long as it is valid in this member state, only the <i>last</i> PGFP event for each member state is recorded. Example: as soon as the 9th annual fee payment event for the FR national phase of an EP patent is recorded, the 8th annual fee payment event is removed from this table.</p>	
Comments	<p>Legal event data like "post grant" information for European patents in the national phase ("validation" data of EP patents) is available in the PRS database, also known as the INPADOC worldwide legal status database.</p> <p>This table must be purchased separately. PATSTAT customers who wish to obtain this table should subscribe to product 14.24.1 to receive PATSTAT Legal Status. (see http://www.epo.org/searching-for-patents/business/patstat.html , Tab "Conditions")</p> <p>The PATSTAT database model is also centred on the DOCDB filing application identifier. Therefore it is very easy to combine the two sets of data.</p> <p>The elements listed above are described in the Worldwide Legal Status Manual "Technical description of the EPO legal status data (Manual)" ; http://www.epo.org/searching-for-patents/legal/inpadoc.html , Tab "Further information".</p> <p>Via the attribute LEC_ID each legal event can be linked to the reference table TLS802_LEGAL_EVENT_CODE. This reference table contains additional information about each type of legal event, e. g. a description, the group of the legal event or its impact on the life of the patent.</p>	
Modification history		
Author of update	Date of update	Explanation of update
D. Lingua	27-07-2010	First version
D. Lingua	20-07-2012	Changed reference to product 14.24.1
M. Kracker	22-04-2013	Changed reference to manual
M. Kracker	01-10-2013	Explicitly listed all attributes
M. Kracker	01-11-2013	Added implementation rules
M. Kracker	01-04-2014	Attribute LEC_ID added; Comment extended.

5.16 TLS222_APPLN_JP_CLASS: Japanese classification

FI and F-terms linked to JP application (only):

FI (File Index) has been developed to expand IPC in some technical fields. FI consists of an IPC symbol and an IPC-subdivision symbol and/or file discrimination symbol added to the IPC symbol.

F-TERMS (File Forming Terms) re-classify or further segment each specific technical field of IPC from a variety of viewpoints (i.e., objective, application, structure, material, manufacturing process, processing, etc.).

TLS222_APPLN_JP_CLASS			
APPLN_ID			
JP_CLASS_SCHEME			
JP_CLASS_SYMBOL			
PRIMARY KEY	APPLN_ID, JP_CLASS_SCHEME, JP_CLASS_SYMBOL		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Business rules	The Japanese Classification schemes FI and FTERM, included in this table, are used by the Japanese Patent Office for carrying out patent application searches. The FI scheme is built on top of the International Patent Classification system (IPC), and is constantly being revised and updated. The FTERM scheme contains technical terms attributed from multiple viewpoints to facilitate computerised retrieval of patent documents. For more details there is an interesting wiki page here .		
Comments	n/a		
Modification history			
Author of update	Date of update	Explanation of update	
D. Lingua	04-08-2011	First version	
D. Lingua	11-10-2011	Updated link	

5.17 TLS223_APPLN_DOCUS: US classification

USPC codes linked to US applications (only):

US patent classification codes are typically expressed in the following format "482/1". The first number, 482, represents the class of invention. The number following the slash is the subclass of invention within the class. There are about 450 classes of invention and about 150 000 subclasses of invention in the USPC.

TLS223_APPLN_DOCUS			
	APPLN_ID		
	DOCUS_CLASS_SYMBOL		
	PRIMARY KEY	APPLN_ID, DOCUS_CLASS_SYMBOL	
	FOREIGN KEY	APPLN_ID	REFERENCES TLS201_APPLN (APPLN_ID)
Business rules	The USPTO Classification scheme DOCUS, included in this table, and is used by the USPTO for carrying out patent application searches. The DOCUS scheme has a completely different structure than the International Patent Classification system (IPC), and is constantly being revised and updated. For more details follow this link (http://www.uspto.gov/patents/resources/classification/help.jsp#5).		
Comments	With the introduction of the Cooperative Patent Classification CPC on 2013-01-01 (http://www.cooperativepatentclassification.org/index.html), the DOCUS scheme will be superseded by the new CPC scheme in the near future. The DOCUS coverage in PATSTAT has not been updated in the last editions.		
Implementation rules	n/a		
Modification history			
Author of update	Date of update	Explanation of update	
D. Lingua	04-08-2011	First version	

5.18 TLS224_APPLN_CPC: Cooperative Patent Classification

The table contains all cooperative patent classifications linked to the applications. The set of classifications linked to a single application is a de-duplicated merge of all classifications of the various publication instances linked to the specific application.

From a statistical point of view it is important to remember that CPC codes are propagated to all members of the same DOCDB family (simple family).

TLS224_APPLN_CPC			
APPLN_ID			
CPC_CLASS_SYMBOL			
CPC_SCHEME			
CPC_VERSION			
CPC_VALUE			
CPC_POSITION			
CPC_GENER_AUTH			
PRIMARY KEY	APPLN_ID, CPC_CLASS_SYMBOL, CPC_SCHEME		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
Business rules	<p>CPC classes can be present in DOCDB for all publication levels of an application. However, in PATSTAT these are aggregated and de-duplicated at application level. For example, frequently there are some CPC symbols assigned to a published application which are also assigned to the published grant. Nevertheless a CPC symbol will be stored in PATSTAT only once.</p> <p>In case the aggregation of the CPC symbols of all publications of a given application contains multiple CPCs with the same CPC_CLASS_SYMBOL and CPC_SCHEME, then the CPC from the latest publication takes precedence.</p>		
Comments	<p>Information on classification according to the Cooperative Patent Classification (CPC) can be found in http://www.epo.org/searching-for-patents/helpful-resources/first-time-here/classification/cpc.html .</p> <p>In the scheme CPC (but not in the scheme CPCNO) all applications of the same DOCDB simple family have the same CPC symbols assigned.</p>		
Modification history			
Author of update	Date of update	Explanation of update	
M. Kracker	07-03-2012	First version	
M. Kracker	15-05-2012	Primary key extended by CPC_SCHEME	
M. Kracker	01-10-2013	Added PATSTAT Online extension attribute	
M. Kracker	15-10-2014	Business rules clarified. Comment updated.	
M. Kracker	01-10-2015	Removed pre-computed and redundant attribute PC_MAINGROUP_SYMBOL.	

5.19 TLS226_PERSON_ORIG: Unmodified person data

This table is best suited for detailed analysis of person data.

A row contains the name and address of a person (applicant and/or inventors; physical person or legal person). The data is taken from various data sources. It is kept in the "original" form, i.e. the data has not been cleaned, aggregated or otherwise modified. Depending on the data structure of each data source, not all attributes of this table are populated for every person.

Each row has one corresponding row in TLS206_PERSON. In TLS206_PERSON the data has been cleaned and unified and its table structure has been simplified and harmonized.

TLS226_PERSON_ORIG			
Identifier and metadata attributes			
PERSON_ORIG_ID			
PERSON_ID			
SOURCE			
SOURCE_VERSION			
Name attributes			
NAME_FREEFORM			
LAST_NAME			
FIRST_NAME			
MIDDLE_NAME			
Address attributes			
ADDRESS_FREEFORM			
ADDRESS_1			
ADDRESS_2			
ADDRESS_3			
ADDRESS_4			
ADDRESS_5			
STREET			
CITY			
ZIP_CODE			
STATE			
PERSON_CTRY_CODE			
RESIDENCE_CTRY_CODE			
Other attributes			
ROLE			
PRIMARY KEY	PERSON_ORIG_ID		
FOREIGN KEY	PERSON_ID	REFERENCES	TLS206_PERSON (PERSON_ID)
Business rules	<p>The table below explains which data source can populate which name and address attribute.</p> <p>Rows where all attributes (except the primary key PERSON_ORIG_ID) are identical are de-duplicated.</p>		
Comments	<p>This table contains just name and address data. It cannot make reliable statements about persons in the real world.</p> <ul style="list-style-type: none"> • It (quite likely) may be the case that 2 rows in the table represent one and the same person in the real world, due to variations of name or address data. • It may also be the (rare?) case that 2 persons in the real 		

	world are represented by the same row in this table, due to incomplete data.
--	--

	Data Source		
	DOCDB	EP (Register)	USPTO Backfile; USPTO Frontfile DTD v4.2 - v4.5
NAME_FREEFORM	✓	✓	
LAST_NAME			✓
FIRST_NAME			✓
MIDDLE_NAME			✓
ADDRESS_FREEFORM	✓		
ADDRESS_1		✓	✓
ADDRESS_2		✓	✓
ADDRESS_3		✓	✓
ADDRESS_4		✓	
ADDRESS_5		✓	
STREET			✓
CITY			✓
ZIP_CODE			✓
STATE			✓
PERSON_CTRY_CODE	✓	✓	✓
RESIDENCE_CTRY_CODE			✓ (only inventors)
ROLE			✓ (only applicants)

Modification history		
Author of update	Date of update	Explanation of update
M. Kracker	23-07-2013	First version
M. Kracker	01-10-2015	Data source for EP Register addresses are ADDRESS_1 to ADDRESS_5
M. Kracker	01-04-2016	Data source for USPTO addresses could also be ADDRESS_1 to ADDRESS_3.

5.20 TLS227_PERS_PUBLN: Link between person and publication

This table links each publication to its applicants and inventors. This can be used to analyse the changes of applicants / inventors at the times of their publication.

TLS227_PERS_PUBLN			
PERSON_ID			
PAT_PUBLN_ID			
APPLT_SEQ_NR			
INVT_SEQ_NR			
PRIMARY KEY	PERSON_ID, PAT_PUBLN_ID, APPLT_SEQ_NR, INVT_SEQ_NR		
FOREIGN KEY	PERSON_ID	REFERENCES	TLS206_PERSON (PERSON_ID)
FOREIGN KEY	PAT_PUBLN_ID	REFERENCES	TLS211_PAT_PUBLN (PAT_PUBLN_ID)
Comments	Conceptually, the combination of PERSON_ID and APPLN_ID should be unique. In practice, due to duplicates in the source data also the attributes APPLT_SEQ_NR and INVT_SEQ_NR must be part of the Primary Key.		
Modification history			
Author of update	Date of update	Explanation of update	
M. Kracker	01-10-2013	First version	

5.21 TLS228_DOCDB_FAM_CITN: Citation between DOCDB families

This table contains one entry for each pair of DOCDB simple families, where one member of a family cites at least one member of another family.

That means if multiple publications of one family cite one or multiple publication(s) / application(s) of another family, then this is counted as one citation between these 2 families.

TLS228_DOCDB_FAM_CITN			
DOCDB_FAMILY_ID			
CITED_DOCDB_FAMILY_ID			
PRIMARY KEY	DOCDB_FAMILY_ID, CITED_DOCDB_FAMILY_ID		
FOREIGN KEY	DOCDB_FAMILY_ID	REFERENCES	TLS201_APPLN (DOCDB_FAMILY_ID)
FOREIGN KEY	CITED_DOCDB_FAMILY_ID	REFERENCES	TLS201_APPLN(DOC DB_FAMILY_ID)
Business rules	n/a		
Comments	Cited publications (the typical case) and well as cited applications (this is also possible) are considered when computing this table.		
Modification history			
Author of update	Date of update	Explanation of update	
M. Kracker	01-10-2013	First version	
M. Kracker	01-04-2015	Formerly this table was called DOCDB_FAMILY_CITATION and was only available in PATSTAT Online. The order of the 2 columns has been reversed and one column has been renamed.	
M. Kracker	01-10-2015	Foreign Keys now link to table TLS201_APPLN attribute DOCDB_FAMILY_ID	

5.22 TLS229_APPLN_NACE2: NACE2 industry classification

This table tells to which degree an application belongs to one or more industries.

TLS229_APPLN_NACE2			
	APPLN_ID		
	NACE2_CODE		
	WEIGHT		
PRIMARY KEY	APPLN_ID, NACE2_CODE		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
FOREIGN KEY	NACE2_CODE	REFERENCES	TLS902_IPC_NACE2 (NACE2_CODE)
Business rules	n/a		
Comments	This table is computed based on the reference table TLS902_IPC_NACE2 and the IPCs of an application. Consequently, applications without IPCs are not assigned to NACE2 codes.		
Modification history			
Author of update	Date of update	Explanation of update	
M. Kracker	01-04-2015	First version	

5.23 TLS230_APPLN_TECHN_FIELD: Classification by technical field

This table tells to which degree an application belongs to one or more technical fields.

TLS230_APPLN_TECHN_FIELD			
	APPLN_ID		
	TECHN_FIELD_NR		
	WEIGHT		
PRIMARY KEY	APPLN_ID, TECHN_FIELD_NR		
FOREIGN KEY	APPLN_ID	REFERENCES	TLS201_APPLN (APPLN_ID)
FOREIGN KEY	TECHN_FIELD_NR	REFERENCES	TLS901_TECHN_FIELD_IPC (TECHN_FIELD_NR)
Business rules	n/a		
Comments	This table is computed based on the reference table TLS901_TECHN_FIELD_IPC and the IPCs of an application. Consequently, applications without IPCs are not assigned to technical fields.		
Modification history			
Author of update	Date of update	Explanation of update	
M. Kracker	01-10-2015	First version	

5.24 TLS801_COUNTRY: Reference table of country codes

Contains information about states and IP organisations which were listed in WIPO standard ST.3, e.g. their (short) name and whether they are member of the EU, the EPO or the OECD.

TLS801_COUNTRY		
	CTRY_CODE	
	ISO_ALPHA3	
	ST3_NAME	
	STATE_INDICATOR	
	CONTINENT	
	EU_MEMBER	
	EPO_MEMBER	
	OECD_MEMBER	
	DISCONTINUED	
PRIMARY KEY	CTRY_CODE	
FOREIGN KEY	Via the CTRY_CODE attribute this table can be joined with any table which contains an attribute with ST.3 country codes	
Business rules	n/a	
Comments	This table is based on WIPO standard ST.3 with additional public information. It is manually maintained by the EPO.	
Modification history		
Author of update	Date of update	Explanation of update
M. Kracker	01-04-2014	First version
M. Kracker	01-04-2015	Addition of column ISO_ALPHA3

5.25 TLS802_LEGAL_EVENT_CODE: Reference table of legal event codes

PATSTAT Legal Status

Contains all legal event codes (LEC) which are used in EPO's worldwide legal status database. Formerly this data was called PRS legal status data. For every code it is indicated whether it has a positive, negative or neutral impact on the life of the patent, and to which legal event code group (LECG) it belongs. Legal events which are not used any more since 2003 may not be included in this list.

TLS802_LEGAL_EVENT_CODE		
Primary Key		
LEC_ID		
Alternate Key, consisting of the following 4 attributes		
AUTH_CC		
LEC_NAME		
NAT_AUTH_CC		
NAT_LEC_NAME		
Attributes of the Legal Event Code		
IMPACT		
LEC_DESCR		
Legal Event Code Group		
LECG_ID		
LECG_NAME		
LECG_DESCR		
PRIMARY KEY	LEC_ID	
ALTERNATE KEY	AUTH_CC, LEC_NAME, NAT_AUTH_CC, NAT_LEC_NAME	
Business rules	n/a	
Comments	<p>This table corresponds to the download "Categorisation of recently used legal status codes" at http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/regular.html .</p> <p>It contains only codes which are used in or after 2003-01-01. Older codes which occur more than about 500 times are also included. These codes cover more than 98% of all legal events of TLS221_INPDOC_PRS.</p> <p>Via the attribute LEC_ID this table can be easily linked to table TLS221_INPADOC_PRS.</p>	
Modification history		
Author of update	Date of update	Explanation of update
M. Kracker	01-04-2014	First version

5.26 TLS901_TECHN_FIELD_IPC: Mapping between technology fields and IPC

A reference table which contains the mapping between 35 technology fields and the much more detailed IPC classification. These technology fields allow for the easy grouping of applications based on technology. The same technology fields are used by EPO and WIPO for their statistics.

TLS901_TECHN_FIELD_IPC			
	IPC_MAINGROUP_SYMBOL		
	TECHN_FIELD_NR		
	TECHN_SECTOR		
	TECHN_FIELD		
PRIMARY KEY	IPC_MAINGROUP_SYMBOL		
FOREIGN KEY	Via the IPC_MAINGROUP_SYMBOL attribute this table can be joined with any table which contains a compatible IPC or CPC attribute		
FOREIGN KEY	TECHN_FIELD_NR	REFERENCES	TLS209_APPLN_IPC (TECHN_FIELD_NR)
Business rules	n/a		
Comments	<p>The content of this table is derived from http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/xls/ipc_technology.xls .</p> <p>More information on this technology classification can be found in "Concept of a Technology Classification for Country Comparisons" by Ulrich Schmoch, July 2008; http://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_41/ipc_ce_41_5-annex1.pdf</p>		
Modification history			
Author of update	Date of update	Explanation of update	
M. Kracker	?	First version	
M. Kracker	01-10-2015	Order of attributes changed	

5.27 TLS902_IPC_NACE2: Mapping between IPC and industrial sectors

A reference table which contains the mapping between the IPC classification and the NACE2 codes for industrial sectors. The industrial sectors allow for the grouping of applications based on the industry.

TLS902_IPC_NACE2		
	IPC	
	NOT_WITH_IPC	
	UNLESS_WITH_IPC	
	NACE2_CODE	
	NACE2_WEIGHT	
	NACE2_DESCR	
PRIMARY KEY	IPC, NOT_WITH_IPC, UNLESS_WITH_IPC, NACE2_CODE	
FOREIGN KEY	Via the attributes IPC, NOT_WITH_IPC and UNLESS_WITH_IPC this table can be matched with any table which contains a compatible IPC or CPC attribute	
Business rules	n/a	
Comments	<p>NACE2 is the Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008) (Nomenclature statistique des activités économiques dans la Communauté européenne). It serves a similar purpose than the SIC (Standard Industrial Classification) and the NAICS (North American Industry Classification System).</p> <p>This concordance table maps IPC sub classes / IPC main groups to the first 2-4 digits of the hierarchical NACE code. All NACE codes can be found in https://circabc.europa.eu/d/a/workspace/SpacesStore/1d3a4661-f37d-4cbe-9998-8bd5966f1c62/CONCORDANCE%20IPC%20V8_NACE%20Rev%202%2020150630.xlsx</p> <p>The data on which this table is based is provided by EUROSTAT in co-operation with KU Leven / Belgium.</p> <p>The data and the methodology to create them is described in https://circabc.europa.eu/sd/a/daab2c8e-a29f-421d-8f51-657a6744ea20/IPC_NACE2_Version2%200_20150630.pdf</p>	
Modification history		
Author of update	Date of update	Explanation of update
M. Kracker	01-04-2015	First version
M. Kracker	01-04-2016	Links in comment updated, because a version of the mapping became available.

5.28 TLS904_NUTS: NUTS regional codes

NUTS (Nomenclature of Territorial Units for Statistics) is a European Union standard for referencing the subdivisions of countries for statistical purposes. This reference table contains the regions of the NUTS 3 level.

TLS904_NUTS			
	NUTS3		
	NUTS3_NAME		
PRIMARY KEY	NUTS3		
FOREIGN KEY	NUTS3	REFERENCES	TLS906_PERSON (NUTS)
Business rules	n/a		
Comments	<p>This table contains NUTS 3 level information of NUTS version 2013.</p> <p>This table is likely to be extended in the future editions of PATSTAT.</p>		
Modification history			
Author of update	Date of update	Explanation of update	
M. Kracker	01-10-2016	First version	

5.29 TLS906_PERSON: Person

See table TLS206_PERSON

6 ATTRIBUTE DESCRIPTION

6.1 Explanation of attribute description

Descriptor	Content	
Name	User name of the field, e.g. "Application number"	
Also Known As	Alternative user names of the field, e.g. "Dossier number" in case of EP applications	
Description	Explanatory description of the field, e.g. "Numeric part of the identification of the application"	
Domain	Description of the domain of values. Depending on the database management system you will use to manage this database, the appropriate data types must be chosen (e. g. nchar, nvarchar, date, integer, ...).	
Default value	The default value from the domain of values, if applicable	
Source database	Name of the database that contains the original data, e.g. "DOCDB".	
Source field name	Name of the field in the source database, e.g. "APPLT_SEQ_NR". This section may also contain instructions for EPO's IT supplier on how to process the data.	
Source sub-field identifier	If necessary: Additional information to identify the source data.	
Comments	Any further comments as deemed necessary	
Modification history		
Author of update	Date of update	Explanation of update
R. Heijna	03-11-2004	First version
D. Lingua	14-07-2009	Preferred caption, Actuality and Source codes deleted from the table
M. Kracker	15-03-2013	Domain description does not depend on a specific DBMS

6.2 ADDRESS_1, ADDRESS_2, ADDRESS_3, ADDRESS_4, ADDRESS_5

Name: Address line 1, Address line 2, Address line 3, Address line 4, Address line 5

Also Known As: address

Description: First / Second / Third / Forth / Fifth address line of a person

Domain: string up to 500 characters

Default value: empty

Source database: EP Register

Source field name

```
<parties>
  <applicants change-gazette-num="2000/29">
    <applicant app-type="applicant" designation="all" sequence="1">
      <addressbook>
        <name>Seidel, Helmut</name>
        <address>
          <address-1>Fliederstrasse 19</address-1>
          <address-2>65396 Walluf</address-2>
          <country>DE</country>
        </address>
      </addressbook>
      <nationality>
        <country/>
      </nationality>
      <residence>
        <country/>
      </residence>
    </applicant>
  </applicants>
  <inventors change-gazette-num="2000/29">
    <inventor sequence="01">
      <addressbook>
        <name>Franta, Georg</name>
        <address>
          <address-1>Ulrich-Rapp-Strasse 18</address-1>
          <address-2>87634 Obergünzburg</address-2>
          <country>DE</country>
        </address>
      </addressbook>
    </inventor>
    <inventor sequence="02">
      <addressbook>
        <name>Dojan, Viktor</name>
        <address>
          <address-1>Ludwig-Strecker-Strasse 5</address-1>
          <address-2>55129 Mainz</address-2>
          <country>DE</country>
        </address>
      </addressbook>
    </inventor>
  </inventors>
```

Comments

The postal code and the city typically are in the last address line which is populated with data.

In PATSTAT Online due to data privacy reasons, the PERSON_ADDRESS has been emptied for all persons who might be a natural person (e. g. all inventors, or where the SECTOR attribute contains "INDIVIDUAL" or "UNKNOWN" or is empty.)

Modification history

Author of update - Date of update - Explanation of update

M. Kracker – 2015-10-01 – First version

6.3 ADDRESS_FREEFORM

Name: Full address in a single string

Also Known As: n/a

Description: Contains the full address in case the address is not available in structured form, where street, city, zip code, ... are in different fields.

Domain: Up to 1000 characters

Default value: empty string

Source database: DOCDB

Source field name

```
<inventors>
  <inventor sequence="1" data-format="docdb">
    <inventor-name>
      <name>STACY N SMITH</name>
    </inventor-name>
    <residence>
      <country>US</country>
    </residence>
  </inventor>
  <inventor sequence="1" data-format="docdba">
    <inventor-name>
      <name>STACY N. SMITH</name>
    </inventor-name>
    <address>
      <text>305 Cottonwood Lane, NC 27540 Holly Springs,
UNITED STATES OF AMERICA (USA)</text>
    </address>
  </inventor>
  <inventor sequence="1" data-format="original">
    <inventor-name>
      <name>Stacy N. Smith</name>
    </inventor-name>
  </inventor>
</inventors>

<applicants>
  <applicant sequence="1" data-format="docdb">
    <applicant-name>
      <name>ERICSSON INC</name>
    </applicant-name>
    <residence>
      <country>US</country>
    </residence>
  </applicant>
  <applicant sequence="1" data-format="docdba">
    <applicant-name>
      <name>ERICSSON INC.</name>
    </applicant-name>
    <address>
      <text>7001 Development Drive, 27709-3969 Research
Triangle Park,UNITED STATES OF AMERICA (USA)</text>
    </address>
  </applicant>
  <applicant sequence="1" data-format="original">
    <applicant-name>
      <name>Ericsson Inc.</name>
    </applicant-name>
  </applicant>
</applicants>
```

Source sub-field identifier

data-format="docdba"

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.4 APPLN_ABSTRACT

Name: Abstract of application

Also Known As: n/a

Description: Abstract of the application

Domain: Up to 12 000 characters

Default value: n/a

Source database: DOCDB

Source field name

```
<abstract lang="EN" data-format="docdba" abstractsource="National Office">
```

```
<p>There is provided a floating surgical cannula. A method of forming a surgical cannula by inserting a floating surgical cannula at a location in need of surgery is provided.</p>
```

```
</abstract>
```

```
<abstract lang="FR" country=WO doc-number="2005000001" kind="A2" date="20050106" data-format="docdba" abstractsource="National Office">
```

```
<p>L'invention concerne une canule chirurgicale flottante. L'invention concerne également un procédé de formation d'une canule chirurgicale qui consiste à introduire une canule chirurgicale flottante dans une zone à opérer.</p>
```

Source sub-field identifier

```
data-format="docdba"
```

Comments

Only one of possibly multiple abstracts is stored. See description of table TLS203_APPLN_ABSTRACT for details.

The average size of abstracts is 854 characters; maximum size is 9992 (as of Oct 2013).

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 13-05-2005 - First version

R. Heijna - 26-09-2005 - Oldest -> youngest

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 04-08-2011 - Addition of the PAJs

D. Lingua - 26-04-2012 - Eliminate comment on PAJs

M. Kracker - 26-03-2013 - Move comments to TLS203_APPLN_ABSTR table description

M. Kracker - 01-10-2013 - Increase suggested domain

6.5 APPLN_ABSTRACT_LG

Name: Language of abstract of application

Also Known As: n/a

Description: Language of the abstract of the application selected for and loaded in PATSTAT

Domain: 2 ASCII characters, according to ISO language codes (ISO 639-1) or spaces

Default value: spaces

Source database: DOCDB

Source field name

```
<abstract lang="EN" data-format="docdb" abstractsource="National Office">
```

```
    <p>There is provided a floating surgical cannula. A method of forming a surgical cannula by inserting a floating surgical cannula at a location in need of surgery is provided.</p>
```

Source sub-field identifier

```
data-format="docdb"
```

Comments

Use the value of this attribute for the abstract stored in the table TLS203_APPLN_ABSTR.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 13-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.6 APPLN_AUTH

Name: Application Authority

Also Known As: Country, State. Receiving Office in case of PCT application

Description: Patent Authority where the National, International or Regional application was filed

Domain: Up to 2 ASCII characters (A-Z), according to [WIPO ST.3](#) and including RH (South Rhodesia)

Default value: n/a

Source database: DOCDB

Source field name

1) Source for the standard applications:

```
<application-reference data-format="docdb" is-representative="N">
  <document-id>
    <country>DE</country>
    <doc-number>10331291</doc-number>
    <kind>A</kind>
    <date>20030710</date>
```

2) For priorities in DOCDB for which there is no application registered in DOCDB, use the authority (country) of the priority:

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>AE</country>
      <doc-number>4000</doc-number>
      <kind>A</kind>
      <date>20000529</date>
    </document-id>
    <priority-active-indicator>Y</priority-active-indicator>
```

3) For artificial applications which were created for all artificial publications which were themselves artificially created for those cited publications, where the cited publications are not registered in DOCDB as publications: use the authority (country) of the cited publication:

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>AM</country>
        <doc-number>199</doc-number>
        <kind>A</kind>
```

4) For artificial applications which were created for applications originating from cited applications not recorded in DOCDB: use the authority (country) as cited in the cited application:

```
<citation srep-phase="APP" sequence="1">
  <patcit num="2" dnum="US19420613452A" dnum-type="application number">
    <document-id>
      <country>US</country>
      <doc-number>19420613452</doc-number>
      <kind>A</kind>
```

```
                <date>00000000</date>
            </document-id>
        </patcit>
</citation>
```

Source sub-field identifier

data-format="docdb"

Comments

Not to be confused with *country of origin* which is the country of the applicant.

Note that for PCT applications, the APPLN_AUTH is the authority where it was filed, not 'WO'. E.g. PCT/EP2007/063304 has been filed at the EPO and therefor has APPLN_AUTH of 'EP' and APPLN_KIND of 'W'.

See section 4.4 "Application replenishment".

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 16-11-2004 - First version

R. Heijna - 22-04-2005 - Source extended

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 15-03-2013 - Added artificial cited applications

M. Kracker - 15-05-2013 - Added exception to domain

M. Kracker - 01-04-2015 - Added 'RH' to domain

6.7 APPLN_FILING_DATE

Name: Application filing date

Also Known As: Date of receipt

Description: Date on which the application was physically received at the Patent Authority

Domain: Date (up to 9999-12-31)

Default value: 9999-12-31

Source database: DOCDB

Source field name

1) Standard applications:

```
<application-reference data-format="docdb" is-representative="N">
  <document-id>
    <country>DE</country>
    <doc-number>10331291</doc-number>
    <kind>A</kind>
    <date>20030710</date>
```

2) Artificial applications from priorities:

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>DE</country>
      <doc-number>10331291</doc-number>
      <kind>A</kind>
      <date>20030710</date>
    </document-id>
  <priority-active-indicator>Y</priority-active-indicator>
```

We assume that all priorities are accurately recorded in DOCDB. If a priority reference does not appear as an application reference, then in PATSTAT we create an artificial application with the authority (country), number kind and date of the priority. See APPLN_ID for the rules for creating the APPLN_ID for these artificial applications. See rules for processing PRIOR_APPLN_SEQ_NR.

3) Artificial applications from citations:

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>US</country>
        <doc-number>3380531</doc-number>
        <kind>A</kind>
```

We assume that all cited references are publications. If a cited reference does not appear as a publication-reference, then in PATSTAT we create an artificial publication. See rules in element PUBLN_NR. We also create an artificial application, using the same country and number as the artificial publication, but we give an APPLN_FILING_DATE of 9999-12-31 and an APPLN_KIND of 'D2'.

4) For artificial applications which were created for applications originating from cited applications not recorded in DOCDB: use the application filing date as cited in the cited application, if not given or invalid then assign '9999-12-31':

```
<citation srep-phase="APP" sequence="1">
  <patcit num="2" dnum="US19420613452A" dnum-type="application number">
    <document-id>
      <country>US</country>
```

```
        <doc-number>19420613452</doc-number>
        <kind>A</kind>
        <date>00000000</date>
    </document-id>
</patcit>
</citation>
```

Source sub-field identifier

data-format="docdb"

Comments

The legal filing date i.e. the date on which the legal protection starts may differ from the Physical filing date. In case of a Divisional Application for instance the legal filing date is the one valid for the parent application which is earlier. It can also be later, e.g. when certain formal requirements are fulfilled later than the physical filing.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 06-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 15-03-2013 - Added artificial cited applications

6.8 APPLN_FILING_YEAR

Name: Year of the application filing date

Also Known As: n/a

Description:

Domain: 4 digits in the form yyyy (e. g. 2015)

Default value: n/a

Source database: PATSTAT

Source field name: Derived from attribute APPLN_FILING_DATE of table TLS201_APPLN:

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2015 – Computation explained

6.9 APPLN_ID

Name: Application identification

Also Known As: n/a

Description: Surrogate key: Technical unique identifier without any business meaning

Domain: Number 0 ... 999 999 999

Default value: n/a

Source database: DOCDB (range 1), PATSTAT (ranges 2, 3, 4)

Source field name:

For range 1 (see below for definition of ranges):

```
<application-reference is-representative="YES" doc-id="11607218" data-  
format="docdb">  
  <document-id>  
    <country>DE</country>  
    <doc-number>8909720</doc-number>  
    <kind>U</kind>  
    <date>19890812</date>  
  </document-id>
```

For ranges 2, 3 and 4: APPLN_ID is set as described in section 4.4 "Application replenishment".

Source sub-field identifier: n/a

Source codes:

For range 1:

```
<application-reference is-representative="YES" doc-id="11607218" data-  
format="docdb">
```

Comments

Previous to the April 2011 edition, a sequential number unique for each unique combination of the elements in the candidate primary key was attributed. The actual number had no particular meaning and would change from one edition to the next.

Starting with the April 2011 edition, the DOCDB "doc-id" unique and stable identifier has been used for all applications found in DOCDB (but not the number ranges 2, 3 & 4 below) to populate APPLN_ID instead of creating a PATSTAT-edition-specific surrogate key. DOCDB attribute "doc-id" contains a stable and unique identifier that will allow for linking up a number of EPO raw data products through the application in a reliable way. However, in exceptional cases some values of APPLN_ID might change even in number range 1 (see below). For details see section 4.3.2 "Stable IDs".

There are 4 ranges of APPLN_ID:

Range 1: 1 to 900 000 000.

This range covers the filed applications which have a related publication in DOCDB. This range 1 is unique but not sequential (there are gaps in the sequence due to loading techniques). This attribute remains the same across PATSTAT editions and always refers to the same combination of application authority, application number and application kind. In case an application is corrected, i.e. the application number and/or kind are changed, then it gets a new APPLN_ID. This is the only reason why a set of data (e.g. person names, publications) can relate to different APPLN_IDs across PATSTAT editions.

Range 2: from 900 000 001 to 930 000 000.

This range covers the artificial applications which are created in PATSTAT for prior applications, claimed as priorities, which do not have an application-reference in DOCDB.

The actual numbers in range 2 have no particular meaning and will change from one edition to the next.

Range 3: 930 000 001 to 960 000 000.

This range 3 covers the artificial filing applications with kind code D2 which are created in PATSTAT for those artificial publications which are also created in PATSTAT because these *publications* are cited, but do not have a publication-reference in DOCDB.

The actual numbers in this range have no particular meaning and will change from one edition to the next.

Range 4: 960 000 001 to 999 999 999

This range 4 covers the artificial filing applications with kind code D3 which are created in PATSTAT because these *applications* are cited.

The actual numbers in this range have no particular meaning and will change from one edition to the next.

See also section 4.4 "Application replenishment".

Note: For reasons of database consistency, there must be a dummy application with an APPLN_ID value of 0.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 15-04-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 31-03-2011 - Introduction of DOCDB unique stable identifier "doc-id"

M. Kracker - 15-03-2013 - Introduction of Range 4

6.10 APPLN_KIND

Name: Kind of Application

Also Known As: n/a

Description: Specification of the kind of application

Domain: Up to 2 ASCII characters:

- A patent
- U utility model
- W PCT application (in the international phase)
- T used by some offices (e. g. AT, DE, DK, ES, GR, HR, PL, PT, SI, SM, TR) for applications which are "translations" of granted PCT or EP applications
- P provisional application (US only)
- F design patent
- V plant patent
- D2, D3 artificial applications (see section 4.4 "Application replenishment")
- Other "exotic" kind codes:

See DOCDB User Documentation(<http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/manuals.html>), section 22.1 "'Exotic' Kind-codes"

Default value: n/a

Source database: DOCDB

Source field name

1) Source for the standard applications:

```
<application-reference data-format="docdb" is-representative="N">
  <document-id>
    <country>DE</country>
    <doc-number>10331291</doc-number>
    <kind>A</kind>
    <date>20030710</date>
```

2) For priorities in DOCDB for which there is no application registered in DOCDB, use the authority (country) of the priority:

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>AE</country>
      <doc-number>4000</doc-number>
      <kind>A</kind>
      <date>20000529</date>
    </document-id>
    <priority-active-indicator>Y</priority-active-indicator>
```

3) For artificial applications which were created for all artificial publications which were themselves artificially created for those cited publications, where the cited publications are not registered in DOCDB as publications: use the kind code "D2":

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>AM</country>
        <doc-number>199</doc-number>
```

<kind>A</kind>

4) For artificial applications which were created for applications originating from cited applications not recorded in DOCDB: use the application kind as cited in the cited application; if not given then use "D3". Note that in 2014 Autumn Edition there was no occurrence of 'D3':

```
<citation srep-phase="APP" sequence="1">
  <patcit num="2" dnum="US19420613452A" dnum-type="application number">
    <document-id>
      <country>US</country>
      <doc-number>19420613452</doc-number>
      <kind>A</kind>
      <date>00000000</date>
    </document-id>
  </patcit>
</citation>
```

Source sub-field identifier

data-format="docdb"

Source sub-field identifier

n/a

Comments

Warning: Please consider that the application kind code landscape can be at times complicated (see also table description TLS201_APPLN in this document).

E.g. for German applications the kind codes "D" or "T" have been used at times followed by the kind code "A" for one and the same application number thus making difficult a proper count of filings. PATSTAT users must consult the [DOCDB application & priority concordance documents](#), and the [DOCDB User Documentation](#) to avoid misinterpretation of the data.

Hint: If APPLN_KIND = "W", then the attribute APPLN_AUTH indicates the Receiving Office, which is the office where this PCT application has been filed.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 01-12-2004 - First version

R. Heijna - 21-04-2005 - Domain redefined, source extended

J. Rollinson - 18-04-2006 - Source codes extended

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 13-04-2012 - Added warning on usage of kind codes

M. Kracker - 15-03-2013 - Added artificial cited applications

M. Kracker - 01-11-2013 - Hint added to comment

M. Kracker - 01-04-2014 – Application kind P (provisional application) added

M. Kracker - 01-10-2016 – Application kind V (plant patent) added

6.11 APPLN_NR

Name: Application number

Also Known As: "Dossier number" in case of EP applications

Description: Number issued by the Patent Authority where the National, International or Regional application was filed

Domain: Up to 15 ASCII characters

This attribute must be unique in combination with APPLN_AUTH & APPLN_KIND.

The last character is either numeric or A, D, K, T or X. The DOCDB administrators make the application numbers end with a D, T or X to create "dummy" application numbers that are present because the number is mandatory but the actual number is not known.

A - data errors

D - dummy application; the publication number is put in front of the D

K – special type of older Brazilian application (number format 11nnnnnK)

T - dummy technical priority

X - dummy pre-1970 derived priority

Default value: empty string

Source database: DOCDB

Source field name

1) Source for the standard applications:

```
<application-reference data-format="docdb" is-representative="N">
  <document-id>
    <country>DE</country>
    <doc-number>10331291</doc-number>
    <kind>A</kind>
    <date>20030710</date>
```

2) Source for the artificial applications from priorities:

We assume that all priorities are accurately recorded in DOCDB. If a priority reference does not appear as an application reference, then in PATSTAT we create an artificial application with the authority (country), number, kind and date of the priority. See APPLN_ID for the rules for creating the APPLN_ID for these artificial applications.

See rules for processing PRIOR_APPLN_SEQ_NR.

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>AE</country>
      <doc-number>4000</doc-number>
      <kind>A</kind>
      <date>20000529</date>
    </document-id>
    <priority-active-indicator>Y</priority-active-indicator>
```

3) Source for the artificial applications from citations:

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>AM</country>
        <doc-number>199</doc-number>
        <kind>A</kind>
```

If a cited document does not appear as a publication-reference in DOCDB, then in PATSTAT we create an artificial publication. See rules in element PUBLN_NR. We also create an artificial application, using the same country and number as the artificial publication, but we give an APPLN_FILING_DATE of 31-12-9999 and an APPLN_KIND of 'D2'. See also the rules for allocating the PUBLN_ID range. See rules for processing CITED_PAT_PUBLN_ID.

4) For artificial applications which were created for applications originating from cited applications not recorded in DOCDB: use the application number as cited in the cited application:

```
<citation srep-phase="APP" sequence="1">
  <patcit num="2" dnum="US19420613452A" dnum-type="application number">
    <document-id>
      <country>US</country>
      <doc-number>19420613452</doc-number>
      <kind>A</kind>
      <date>00000000</date>
    </document-id>
  </patcit>
</citation>
```

Source sub-field identifier

data-format="docdb"

Source codes

n/a

Comments

The terms "Application number" and "Dossier number" are in use for the complete identification, for example "EP99101234"

See "Application Replenishment"

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 03-11-2004 - First version

R. Heijna - 20-04-2005 - Domain identified, source extended

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 15-03-2013 - Added artificial cited applications

6.12 APPLN_NR_EPODOC

Name: Application number in EPODOC format

Also Known As: EPODOC application number

Description: Number in EPODOC format (containing letters and digits) which, if present - will uniquely identify an application. The number is created by the EPO based on the DOCDB application number, application authority and application kind.

Domain: Up to 20 ASCII characters (typically, 13 - 14 characters)

Explanation of the format, according to Annex XI of the "Exchange Format" document of DOCDB, version 2.4.3 from 01.01.2013

Basic structure of application and priority-numbers in data-format="epodoc" is:

- country
- number
 - ccyy - century/year derived from application- or priority-date
 - nnnnnn - serial number, leading zeroes when required
- kind-code, when kind-code not = 'A'

Extended structure for a number of countries:

- country ["WO" when kind-code in data-format="docdb" is "W"]
- number
 - ccyy : century/year derived from application- or priority-date
 - xx : "other data"
 - nnnnn : serial number, leading zeroes when required
- kind-code, when kind-code not = 'A'

"Other data" may be:

- regional office, e.g. 'MI' when country = 'IT' and regional office = Milan
- filing country, e.g. 'US' when country = 'WO' and filing country = US
- ...

Length of the concatenated string is generally fixed at 13 characters or 14 when the kind-code is appended. Strings exceeding a total of 13 or 14 may occur, when the number of significant digits exceeds the number of digits reserved for the serial number, e.g. DE.

A special format applies to numbers that in data-format="docdb" have been suffixed with letters 'D' or 'T' or 'X':

- country
 - 'D' or 'T' or 'X'
 - number
 - kind-code, when kind-code not = 'A'
- Default value:** empty (if not provided by DOCDB due to formatting issues)

Source database: DOCDB

Source field name

1) Source for the standard (= non-artificial) applications:

```
</application-reference>
  <application-reference data-format="epodoc">
    <document-id>
      <doc-number>US20070859929</doc-number>
    </document-id>
  </application-reference>
```

2) For all artificial applications the attribute APPLN_NR_EPODOC will contain an empty string.

Source sub-field identifier

data-format="epodoc"

Source codes

n/a

Comments

The number is almost unique. For technical reasons there currently (Oct 2013) are a few hundred applications with non-unique values in APPLN_NR_EPODOC.

This attribute is useful to easily look up details on an application in Espacenet, which also uses the EPODOC application number to identify an application. You can either

- enter the attribute (e. g. DE20051040258) into the search mask of Espacenet or
- construct a URL like e. g. <http://worldwide.espacenet.com/DE20051040258> to directly see the search result

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.13 APPLN_NR_ORIGINAL

Name: Application number in original format

Also Known As: Original application number

Description: Application number in original format as provided by the supplier. It is assumed that the number is as printed on the respective publications.

Typically these numbers do not contain the country code. In about 10% of the applications no original application number is known.

Domain: Up to 100 characters

Default value: empty

Source database: DOCDB

Source field name

1) Source for the standard (= non-artificial) applications:

```
<exch:application-reference data-format="original">
  <document-id>
    <doc-number>11137814</doc-number>
  </document-id>
</exch:application-reference>
```

If DOCDB does not provide an original application number in any of the publications of an application, then APPLN_NR_ORIGINAL will contain an empty string.

If DOCDB provides multiple conflicting original application numbers for the same application, then only one (= any of the conflicting) original application numbers should be stored. (Note: This is supposed to not happen, but may still occur due to data errors)

EP publications published after 2013-03-13, the application number is published in DOCDB with a check digit, i.e. 04801606.7. For sake of consistency with previous original application numbers, the check digit is removed in PATSTAT.

2) For all artificial applications the attribute APPLN_NR_ORIGINAL will contain an empty string.

Source sub-field identifier

data-format="original"

Source codes

n/a

Comments

This attribute is useful to combine application data of PATSTAT with other databases which also contains the original application number.

The original application number is not necessarily unique within the same APPLN_AUTH and the same APPLN_KIND (e.g. for patents and utility models). For example, the offices of US, JP, FR, CH, CS, IT, SU seem to have re-used their application numbers at least in some periods of time.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2016 - First version

6.14 APPLN_TITLE

Name: Title of application

Also Known As: n/a

Description: Title of the application

Domain: Up to 3.000 characters

Default value: n/a

Source database: DOCDB

Source field name:

```
<invention-title lang="EN" data-format="docdba"> SURGICAL  
CANNULA</invention-title>
```

Source sub-field identifier

```
data-format="docdba"
```

Comments

Only one of possibly multiple abstracts is stored. See description of table TLS203_APPLN_ABSTRACT for details.

The average size of titles is 53 characters; maximum size is 3000 for a Brazilian document (as of April 2013).

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 13-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.15 APPLN_TITLE_LG

Name: Language of title of application

Also Known As: n/a

Description: Language of the title of the application selected for and loaded in PATSTAT

Domain: 2 ASCII characters, according to ISO language codes (ISO 639-1) or spaces

Default value: spaces

Source database: DOCDB

Source field name

```
<invention-title lang="EN" data-format="docdb"> SURGICAL  
CANNULA</invention-title>
```

Source sub-field identifier

data-format="docdb"

Comments

Use the value of this attribute for the title stored in the table TLS201_APPLN_TITLE.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 13-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.16 APPLT_SEQ_NR

Name: Sequence number of applicant

Also Known As: n/a

Description: Number indicating the place in the list of applicants in the application

Domain: Number 0 ... about 250

Default value: 0

Source database: DOCDB

Source field name

- 1) EPO Register for EP patent applications
- 2) OECD patents database for US data post 1976-01-01 up to and including November 15th 2005 for Published Grants.
- 3) PATSTAT weekly file extracts from USPTO website for Published Grants from November 22nd 2005 until today; Published Applications from September 29th 2005 to today inclusive.
- 4) DOCDB Applicant sequence number for USPTO Published Applications from March 1st 2001 to September 22nd 2005 from DOCDB, data-format="docdba".
- 5) all other Applicant Sequence numbers come from DOCDB, data-format="docdba".

```
<applicant sequence="1" data-format="docdba">
  <applicant-name>
    <name>THE JOHNS HOPKINS UNIVERSITY MACDONALD, ALEX
BRUCE</name>
  </applicant-name>
</applicant>
<applicant sequence="2" data-format="docdba">
  <applicant-name>
    <name>AN, LING LING UNIVERSITY OF MASSACHUSETTS, A
PUBLIC INSTITUTION OF HIGHER EDUCATION OF THE COMMONWEALTH
OF MASSACHUSETTS,</name>
  </applicant-name>
</applicant>
```

Source sub-field identifier

sequence="1" data-format="docdba"

Comments

An entry with a value 1 to n represents an applicant; an entry with the value 0 does not represent an applicant, but another person (e.g. an inventor). It is possible that there are applications where no applicants are known.

Consequently, adding the condition "APPLT_SEQ_NR > 0" to the WHERE clause in a query retrieves only those persons from TLS207_PERS_APPLN or TLS227_PERS_PUBLN which are applicants.

Likewise, adding the condition "APPLT_SEQ_NR > 0 AND INVT_SEQ_NR > 0 " retrieves only persons which for a certain application are applicants as well as inventors.

For US data:

Documents published after 1976-01-01: The sequence number is designed to represent the sequence in which Applicants appear on the documents. In this database, this is accurate

for the first-named applicant. For the second- or later- named applicants, the sequence number in this database has been arbitrarily given.

For all US documents published before 1976-01-01, where the data was taken from DOCDB, the sequence numbers are believed to be correct.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 21-12-2004 - First version

R. Heijna - 07-07-2005 - Value zero for the physical model

J. Rollinson - 18-04-2006 - US data comment added

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-10-2013 - Changed source from EPO Bulletin to EPO Register;
changed domain

M. Kracker - 01-10-2015 - Changed comment

6.17 AUTH_CC **PATSTAT Legal Status**

Name: Authority country code.

Also Known As: n/a

Description: Refers to the authority (regional or national) of the IPR.

Part of the Alternate Key. To identify the type of a legal event, it has to be combined with the other 3 attributes of the Alternate Key.

Domain: 2 ASCII characters (according to WIPO standard ST.3)

Default value: n/a

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.18 CITED_APPLN_ID

Name: Identification of cited application

Also Known As: n/a

Description: Surrogate key of the application that is cited

Domain: Number 0 ... 999 999 999; see also attribute APPLN_ID

Default value: 0

Source database: DOCDB, PATSTAT

Source field name

1) Direct patent application citations:

If `citation srep-phase="APP"`, indicating this citation was done by the applicant, then `<citation><patcit>` may contain either a reference to a cited *publication* or a reference to a cited *application*. If `<pat-cit dnum-type="application number">`, then use `country`, `doc-number` and `kind` in `<patcit><document-id>` to find the corresponding APPLN_ID for this application - via APPLN_AUTH, APPLN_NR and APPLN_KIND. The value of APPLN_ID for this application is the CITED_APPLN_ID.

APPLN_DATE is taken from the date in `<patcit><document-id>`. If the date is not given in `<document-id>`, then 9999-12-31 is to be used.

If there is no corresponding application in table TLS201_APPLN in PATSTAT, then create an artificial application in table TLS201_APPLN. See section 4.4.2 "Application replenishment for citations".

Usage Example EP 2305027 A2:

```
<citation srep-phase="APP" sequence="46">
  <patcit num="1" dnum="US46600890A" dnum-type="application number">
    <document-id>
      <country>US</country>
      <doc-number>46600890</doc-number>
      <kind>A</kind>
      <date>19900112</date>
    </document-id>
  </patcit>
```

2) Patent application cited within Non Patent Literature citations:

Here CITED_APPLN_ID refers to a patent application id which has been extracted from a Non Patent Literature citation. In a row in table TLS212_CITATION, you will find these columns:

- PAT_PUBLN_ID
- CITN_ID
- CITN_ORIGIN
- CITED_APPLN_ID
- PAT_CITN_SEQ_NR
- NPL_PUBLN_ID
- NPL_CITN_SEQ_NR

If the NPL_PUBLN_ID is greater than 0, and if that NPL citation refers to a corresponding patent application, then CITED_PAT_PUBLN_ID will hold the value of the APPLN_ID of the corresponding document.

If the corresponding document cannot be found as a publication-reference in DOCDB, then create an artificial cited application for it (see case 1) above).

Citation of patent applications can only occur in a `<nplcit>` without `<source-doc>` element. In this case the cited patent document is the document defined in the element `<document-id>` of the first `<patcit>` element in corresponding-docs in nplcit in citation in references-cited in DOCDB.

There may be 0, 1 or more `<corresponding-docs>`. Be sure to take the first `<patcit>` element - if there is any - while ignoring all `<nplcit>` elements.

Note that this first `<patcit>` element can be a reference to a patent *publication* (then see section 6.21 "CITED_PAT_PUBLN_ID") or to a patent *application*, depending on the `dnum-type` attribute of the `<patcit>` element.

Usage Example:

```
<citation cited-phase="APP" sequence="15">
  <nplcit num="1" npl-type="a">
    <text>HAMMOND S, DUNKI-JACOBS R, HARDY R, TOPKA T: 'Architecture and
    Operation of a Systolic Sparse Matrix Engine' PROCEEDINGS OF THE THIRD SIAM
    CONFERENCE ON PARALLEL PROCESSING FOR SCIENTIFIC COMPUTING 1998, pages 419 -
    423</text>
  </nplcit>
  <corresponding-docs>
    <patcit num="15" dnum="US87354004A" dnum-type="application number">
      <document-id doc-id="53925313">
        <country>US</country>
        <doc-number>87354004</doc-number>
        <kind>A</kind>
        <date>20040621</date>
      </document-id>
    </patcit>
  </corresponding-docs>
</citation>
```

Source sub-field identifier: n/a

Comments

Not only applications can be cited, but – much more typically - publications as well.

Note: Cited publications (see CITED_PAT_PUBLN_ID) and *not* related to cited applications (see CITED_APPLN_ID).

In the 2015 Autumn Edition there no replenished applications having kind code "D3" occurred (see section 4.4.3 "Allocating the APPLN_ID").

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 04-08-2011 - First version

D. Lingua - 26-04-2012 - Comment on "D3" kind code added

M. Kracker - 01-04-2015 – Clarification added in comment

M. Kracker - 01-12-2015 – Patent applications can also be cited from within NPL citations

6.19 CITED_DOCDB_FAMILY_ID

Name: ID of the cited DOCDB simple family

Also Known As: n/a

Description: Uniquely identifies the cited family. The ID has no business meaning.

Domain: Number 1 ... 999 999 999

Default value: n/a

Source database: PATSTAT

Source field name: Derived from the publication information (TLS211_PAT_PUBLN), citation information (TLS212_CITATION) and DOCDB family information (TLS218_DOCDB_FAM)

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.20 CITED_NPL_PUBLN_ID

Name: Identification of cited non-patent literature

Also Known As: n/a

Description: Surrogate key for Non-Patent Literature publications which has been cited

Domain: Number 0 ... 999 999 999; same as attribute NPL_PUBLN_ID

Default value: 0

Source database: DOCDB, PATSTAT

Source field name:

See attribute NPL_PUBLN_ID

Comments

These numbers are not allocated sequentially. Only the "replenished" surrogate keys starting from 950 000 001 are allocated sequentially.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2016 – Name changed (was: NPL_PUBLN_ID)

6.21 CITED_PAT_PUBLN_ID

Name: Identification of cited patent publication

Also Known As: n/a

Description: Surrogate key of the publication that is cited

Domain: Number 0 ... 999 999 999; see also attribute PAT_PUBLN_ID

Default value: 0

Source database: DOCDB, PATSTAT

Source field name

1) Direct patent publication citations:

With `country`, `doc-number` and `kind` in `document-id` in `patcit` in `citation` in `references-cited` in DOCDB the corresponding publication in PAT_PUBLN in PATSTAT is determined (via PUBLN_AUTH, PUBLN_NR and PUBLN_KIND). The value of PAT_PUBLN_ID for this application is the CITED_PUBLN_ID.

If there is no corresponding publication in PAT_PUBLN in PATSTAT, an artificial publication in table TLS211_PAT_PUBLN is to be created. Besides the key-elements, PUBLN_DATE is filled from `date` in `document-id` in `references-cited`, if it is present.

An artificial application must then be created as well, with APPLN_AUTH equal PUBLN_AUTH, APPLN_NR equal PUBLN_NR and APPLN_KIND equal 'D2'. The APPLN_FILING_DATE is the same as the PUBLN_DATE for the corresponding artificial publication.

A corresponding surrogate key APPLN_ID must also be created, in the range of ids for artificial applications for artificial cited publications.

2) Patent publications cited within Non Patent Literature citations:

Here CITED_PAT_PUBLN_ID refers to a patent publication id which has been extracted from a Non Patent Literature citation. In a row in table TLS212_CITATION, you will find these columns:

- PAT_PUBLN_ID
- CITN_ID
- CITN_ORIGIN
- CITED_PAT_PUBLN_ID
- PAT_CITN_SEQ_NR
- NPL_PUBLN_ID
- NPL_CITN_SEQ_NR

If the NPL_PUBLN_ID is greater than 0, and if that NPL citation refers to a corresponding patent document, then CITED_PAT_PUBLN_ID will hold the value of the PAT_PUBLN_ID of the corresponding document.

If the corresponding document cannot be found as a publication-reference in DOCDB, then create an artificial cited publication for it (see case 1) above).

Two cases have to be distinguished:

2a) `<nplcit>` with `<source-doc>` element

In this case the cited patent document is the document defined in the element `<document-id>` in `source-doc` in `nplcit` in `citation` in `references-cited` in `DOCDB`. There is at most one `<source-doc>` element.

The `<source-doc>` element will always contain one reference to a patent *publication*, and never a reference to a patent *application*.

Usage Example:

```
<references-cited>
  ...
  <citation cited-phase="SEA" cited-date="20110103" srep-office="EP"
    sequence="4">
    <nplcit num="1" npl-type="d" extracted-xp="002391653">
      <text>DATABASE WPI Week 200577, Derwent Publications Ltd.,
London, GB; AN 2005-752331, XP002391653</text>
      <online>
        <edition>0</edition>
        <vid>2005</vid>
        <ino>77</ino>
        <absno>2005-752331</absno>
      </online>
      <source-doc>
        <document-id>
          <country>JP</country>
          <doc-number>2005281133</doc-number>
          <kind>A</kind>
        </document-id>
      </source-doc>
    </nplcit>
  </citation>
</references-cited>
```

2b) `<nplcit>` without `<source-doc>` element

In this case the cited patent document is the document defined in the element `<document-id>` of the **first** `<patcit>` element in `corresponding-docs` in `nplcit` in `citation` in `references-cited` in `DOCDB`. There may be 0, 1 or more `<corresponding-docs>`. Be sure to take the **first** `<patcit>` element - if there is any - while ignoring all `<nplcit>` elements.

Note that this first `<patcit>` element can be a reference to a patent *publication* or to a patent *application* (in the latter case see section 6.18 "CITED_APPLN_ID"), depending on the `dnum-type` attribute of the `<patcit>` element.

Usage Example:

```
<citation cited-phase="SEA" cited-date="20120620" srep-office="EP" sequence="1">
  <nplcit num="1" npl-type="e" extracted-xp="002678021">
    <text>DATABASE Geneseq [online] 12 June 2008 (2008-06-12), "Rice
cDNA-encoded protein SEQ ID No 31047.", XP002678021, retrieved from EBI accession
no. GSP:AQD37188 Database accession no. AQD37188</text>
```

```

    <online>
      <online-title>Rice cDNA-encoded protein SEQ ID No
31047.</online-title>
      <hosttitle>Geneseq</hosttitle>
      <imprint>
        <name/>
        <pubdate>20080612</pubdate>
      </imprint>
      <hostno>GSP:AQD37188</hostno>
      <absno>AQD37188</absno>
      <avail>EBI</avail>
    </online>
  </nplcit>
  <rel-passage>
    <passage>
      <sequence>.</sequence>
      <pp>A</pp>
    </passage>
    <passage>shows 100% identity to present SEQ ID NO:246 (protein) and
corresponding polynucleotide shows 100 % identity to SEQ ID NO:7 over 458
nucleotides</passage>
    <category>X</category>
    <rel-claims>1-10</rel-claims>
  </rel-passage>
  <category>X</category>
  <corresponding-docs>
    <nplcit num="2" npl-type="e" extracted-xp="002678022">
      <text>DATABASE EMBL [online] 15 July 2011 (2011-07-15), "JP
2005185101-A/2581: Full length cDNA of plant and the use thereof.", XP002678022,
retrieved from EBI accession no. EM_PAT:HV067703 Database accession no.
HV067703</text>
      <online>
        <online-title>JP 2005185101-A/2581: Full length cDNA of
plant and the use thereof.</online-title>
        <hosttitle>EMBL</hosttitle>
        <imprint>
          <name/>
          <pubdate>20110715</pubdate>
        </imprint>
        <hostno>EM_PAT:HV067703</hostno>
        <absno>HV067703</absno>
        <avail>EBI</avail>
      </online>
    </nplcit>
  </corresponding-docs>
  <corresponding-docs>
    <patcit num="1" dnum="US2006123505A1" dnum-type="publication number">
      <document-id doc-id="282734157">
        <country>US</country>
        <doc-number>2006123505</doc-number>
        <kind>A1</kind>
        <name>KIKUCHI SHOSHI [JP], et al</name>
        <date>20060608</date>
      </document-id>
    </patcit>
  </corresponding-docs>
  <corresponding-docs>
    <patcit num="2" dnum="JP2005185101A" dnum-type="publication number">
      <document-id doc-id="282734156">
        <country>JP</country>
        <doc-number>2005185101</doc-number>

```

```
        <kind>A</kind>
        <name>NAT INST OF AGROBIO SCIENCES, et al</name>
        <date>20050714</date>
    </document-id>
</patcit>
</corresponding-docs>
</citation>
```

Source sub-field identifier

n/a

Comments

No self-citing is allowed, so ignore any cited documents which are the same as the publication-reference. In this respect, ignore the Kind Code for EP publications. I.e. if EP1000000B1 cites EP1000000 with any kind code (including EP1000000 with no kind code), then ignore this citation.

An aggregate count of publications in PATSTAT will result in a higher count than in DOCDB, due to the inclusion of these artificial publications in PATSTAT. The difference is usually at the publication kind code level, as the cited kind code is incomplete or missing. For example, publication EP1000000A in PATSTAT is artificial, it does not exist in DOCDB - the correct kind code is A1, e.g. EP1000000A1

See also the rules in the description of table TLS212_CITATION.

Note: Cited publications (see CITED_PAT_PUBLN_ID) are *not* related to cited applications (see CITED_APPLN_ID).

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

R. Heijna - 15-11-2005 - Special EP rule removed

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-04-2015 – Clarification added in comment

M. Kracker - 01-12-2015 – New processing rules for citations within NPL citations (case 2))

6.22 CITN_CATEG

Name: Category of the citation

Also Known As: n/a

Description: Category of the citation as mentioned in Search Reports

Domain: 1 character (X, I, Y,A,D,E,P,L,R,T,O)

Default value: n/a

Source database: DOCDB

Source field name

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>US</country>
        <doc-number>3380531</doc-number>
        <kind>A</kind>
      </document-id>
    </patcit>
    <category>A</category>
  </citation>
  <citation srep-phase="SEA" sequence="2">
    <nplcit>
      <text> REGLEMENT (CEE) n° 2377/90 DU CO anima & PL-163415-A and
XP1234456</text>
    </nplcit>
    <corresponding-docs>
      <document-id>
        <country>PL</country>
        <doc-number>163415</doc-number>
        <kind>A</kind>
      </document-id>
      <refno>XP001234456</refno>
    </corresponding-docs>
    <category>XPD</category>
  </citation>
</references-cited>
```

Source sub-field identifier

n/a

Comments

All categories for one citation are stored together in one field in DOCDB. They have to be disentangled in order to obtain multiple single categories.

Only when CITN_ORIGIN is SEA, ISR, SUPP and PRS (= citations introduced during search, International Search Report, Supplementary Search Report or pre-search) categories may – but need not - occur; in general only the search examiners give these categories. For some countries (e. g. US, JP, but also other countries) no categories are available.

See Annex XIV of the [DOCDB User Documentation](#) for an explanation of the meaning of the categories.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 06-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 11-10-2010 - Added rules on SEA categories

D. Lingua - 26-04-2012 - Added category "I"

M.Kracker – 01.04.2014 – Clarified comment

M.Kracker – 01.04.2015 – Comment changes: Restrictions on number of categories per application has been lifted

M.Kracker – 01.04.2015 – Comment changes: Categories may also occur in the PRS phase.

6.23 CITN_GENER_AUTH

Name: Identification of International Search Authority (ISA) for PCT search reports (incl. supplementary search reports)

Also Known As: n/a

Description: Country code identifying the patent authority performing the International Search Report

Domain: 2 characters (A-Z), according to [WIPO ST.3](#) or spaces

Default value: spaces

Source database: DOCDB

Source field name < citation srep-phase="ISR" srep-office="AT" sequence="1">

Usage Example:

```
<references-cited>
  <citation srep-phase="ISR" srep-office="AT" sequence="1">
    <patcit num="1" dnum="US4996335A" dnum-type="publication number">
      <document-id>
        <country>US</country>
        <doc-number>4996335</doc-number>
        <kind>A</kind>
        <date>19910226</date>
      </document-id>
    </patcit>
    <category>X</category>
  </citation>
  <citation srep-phase="ISR" srep-office="AT" sequence="2">
    <patcit num="2" dnum="BE889563A1" dnum-type="publication number">
      <document-id>
        <country>BE</country>
        <doc-number>889563</doc-number>
        <kind>A1</kind>
        <date>19811103</date>
      </document-id>
    </patcit>
    <category>X</category>
  </citation>
  <citation srep-phase="ISR" srep-office="AT" sequence="3">
    <nplcit num="1">
      <text>DATABASE CAPLUS [Online] LITTLE R.J. ET AL.: 'Soft drugs
based on hydrocortisone: the inactive metabolite approach and its application to
steroidal antiinflammatory agents', XP009000424 Retrieved from STN Database
accession no. (1999:408919) & PHARMACEUTICAL RESEARCH vol. 16, no. 6, 1999,
pages 961 - 967</text>
    </nplcit>
    <corresponding-docs>
      <refno>009000424</refno>
    </corresponding-docs>
    <category>X</category>
  </citation>
</references-cited>
```

Source sub-field identifier: n/a

Comments

The column CITN_GENER_AUTH will only be populated where CITN_ORIGIN is SEA, ISR, SUP or PRS.

Before the 2016 Autumn Edition this information was quite messy, but the data has been cleaned up now.

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 04-08-2011 - First version

M. Kracker - 01-10-2016 – Comment changed

6.24 CITN_ID

Name: Citation identification

Also Known As: n/a

Description: Number distinguishing the citations in one citing document (patent publication)

Domain: Number 1 .. about 1100

Default value: n/a

Source database: Computed from PATSTAT. It is a sequential number for each citation within one citing patent publication. The numbering starts with 1.

Source field name: n/a

Source sub-field identifier: n/a

Comments

The number does not bear a particular meaning; it is just a running number among all citations in one citing document.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 21-11-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-12-2015 – Domain and processing instructions changed

6.25 CITN_ORIGIN

Name: Origin of the citation

Also Known As: Citation phase

Description: Provenance of the citation

Domain: 3 ASCII character code

The code indicates the origin of the citation:

APP	citations introduced by the applicant
SEA	citations introduced during search (from Search Report)
ISR	citations from the International Search Report
SUP	citations from the Supplementary Search Report
PRS	"PRe-Search" citations (available before official publication; only for US applications)
EXA	citations introduced during examination
OPP	citations introduced during opposition (citations by opponent published with a European Patent Specification (EP-B2))
APL	citations introduced when filed for appeal by applicant / proprietor / patentee
FOP	citations introduced when filed opposition by any third party after the publication of a European Patent Specification (EP- B1)
TPO	citations introduced because of Third Party Observations (Art 115 EPC)
CH2	citations introduced during the Chapter 2 phase of the PCT

Default value: n/a

Source database: DOCDB

Source field name

```
<citation srep-phase="SEA" sequence="1">
  <patcit num="1" dnum="WO9505670A1" dnum-type="publication number">
    <document-id>
      <country>WO</country>
      <doc-number>9504670</doc-number>
      <kind>A1</kind>
    </document-id>
  </patcit>
  <category>Y</category>
</citation>
<citation srep-phase="SEA" sequence="2">
  <patcit num="2" dnum="DE4135041A1" dnum-type="publication number">
    <document-id>
      <country>DE</country>
      <doc-number>4135041</doc-number>
      <kind>A1</kind>
    </document-id>
  </patcit>
  <category>Y</category>
</citation>
<citation srep-phase="SEA" sequence="3">
  <patcit num="3" dnum="FR2730035A1" dnum-type="publication number">
    <document-id>
      <country>FR</country>
      <doc-number>2730035</doc-number>
      <kind>A1</kind>
    </document-id>
```

```

    </patcit>
    <category>Y</category>
</citation>
<citation srep-phase="APP" sequence="1">
  <patcit num="1" dnum="DE4007646A1" dnum-type="publication number">
    <document-id>
      <country>DE</country>
      <doc-number>4007646</doc-number>
      <kind>A1</kind>
    </document-id>
  </patcit>
</citation>
<citation srep-phase="APP" sequence="2">
  <patcit num="2" dnum="JP4241100A" dnum-type="publication number">
    <document-id>
      <country>JP</country>
      <doc-number>4241100</doc-number>
      <kind>A</kind>
    </document-id>
  </patcit>
</citation>
<citation srep-phase="APP" sequence="3">
  <patcit num="3" dnum="JP7044800A" dnum-type="publication number">
    <document-id>
      <country>JP</country>
      <doc-number>7044800</doc-number>
      <kind>A</kind>
    </document-id>
  </patcit>
</citation>

```

Comments

Look at the value of the element `srep-phase` to get the value of `CITN_ORIGIN`. Element `<patcit>` may contain cited publications or cited applications. Cited applications only when `srep-phase="APP"`.

The table "Overview of citation data in the EPO's citation database (REFI)" in <http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/regular.html> provides a full list of origins available for a given authority.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 06-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 14-06-2010 - Introduced new citation origins

D. Lingua - 08-10-2012 - Introduced new citation origin PRS

M. Kracker - 01.10.2013 - Changed domain; Introduced new citation origins; clarification of codes

M. Kracker - 01.10.2015 - Code 115 changed to code TPO

6.26 CITY

Name: City part of the address

Also Known As: n/a

Description: Contains the city part of the address

Domain: Up to 200 characters

Default value: empty string

Source database: USPTO data of published applications and published grants

Source field name:

<addressbook> <address> <city>

Source sub-field identifier: n/a

Comments:

n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-10-2015 – Removed source “EP Register data”;
cf. attributes ADDRESS_1, ..., ADDRESS_5

6.27 CONTINENT

Name: Continent

Also Known As: n/a

Description: Name of the continent (in English) in which a state is located (according to Wikipedia)

Domain: Up to 25 ASCII characters:

Only populated if the attribute STATE_INDICATOR is 'Y':

May have one of these 6 values:

- Africa
- America
- Asia
- Australia and Oceania
- Europe
- Europe/Asia

Note that the Russian Federation, Soviet Union and Turkey have the value 'Europe/Asia'

Default value: empty

Source database: based on Wikipedia

Source field name: n/a

Source sub-field identifier: n/a

Comments:

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.28 CONTN_TYPE

Name: Continuation type

Also Known As: n/a

Description: The type of continuation describing what relation the later application has to the earlier application. In DOCDB, this is known as the type of linkage between applications and priorities.

Domain: 3 ASCII characters

ADD - Addition
CON - Continuation
CGT - Cognate
CIP - Continuation in part
DIV - Division
INN - Internal priority
REI - Re-issue
SBS - Substitute
SUP - Supplementary disclosure
spaces - unknown

Default value: n/a

Source database: DOCDB

Source field name

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb" status="A">
    <country>US</country>
    <doc-number>90976604</doc-number>
    <kind>A</kind>
    <date>20040802</date>
    <priority-active-indicator>Y</priority-activeindicator>
  </priority-claim>
  <priority-claim sequence="1" data-format="epodoc">
    <doc-number>US20040909766</doc-number>
  </priority-claim>
  <priority-claim sequence="2" data-format="docdb" status="A">
    <country>US</country>
    <doc-number>9885602</doc-number>
    <kind>A</kind>
    <date>20020314</date>
    <priority-linkage-type>3</priority-linkage-type>
    <priority-active-indicator>N</priority-activeindicator>
  </priority-claim>
  <priority-claim sequence="2" data-format="epodoc">
    <doc-number>US20020098856</doc-number>
  </priority-claim>
  <priority-claim sequence="1" data-format="original">
    <doc-number>9885602</doc-number>
  </priority-claim>
</priority-claims>
```

From the application publication authority code (APPLN_AUTH) and the `priority-linkage-type` the continuation type is determined from the table in section 4.6.2 Continuation types.

Note:

a) if there is no element <priority-linkage-type>, then put spaces in CONTN_TYPE.

b) if there is no matching entry in the table, then put spaces in CONTN_TYPE.

Note that before 1991, the EPO did not record the so called "linkage type" of priority numbers, that is the EPO did not record which kind of relation a given priority number has (Paris Union priority, continuation, division, etc.). Data in this element prior to 1991 is thus not reliable.

Source sub-field identifier

data-format="docdb"

Comments

n/a

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

R. Heijna - 13-07-2005 - Domain adapted

J. Rollinson - 14-02-2008 - pre 1991 US fact.

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.29 CPC_CLASS_SYMBOL

Name: CPC classification symbol

Also Known As: CPC class, CPC classification, CPC symbol

Description: Classification symbol according to the Cooperative Patent Classification

Domain: Up to 19 characters (A-Z, 0-9, /, space);

All values which are allowed by the CPC;

Corresponds to position 1 - 19 (i.e. section, class, subclass, main group, subgroup) of the 50 character long text string as defined by [WIPO ST.8](#) with trailing spaces removed.

Examples: A61K
H04Q 7/32
C07K 14/00
C07D 405/06
H01M2220/20

Note that spaces may be required on position 5-7, because the slash "/" is always on the 9th position. For more details see the table below.

Default value: n/a

Source database: DOCDB

Source field name:

```
<patent-classifications>
  <patent-classification sequence="1">
    <classification-scheme office="EP" scheme="CPC">
      <date>20130101</date>
    </classification-scheme>
    <classification-symbol>B60V 1/16 </classification-symbol>
    <classification-value>I</classification-value>
    <classification-status>B</classification-status>
    <classification-data-source>H</classification-data-source>
    <action-date>
      <date>20130101</date>
    </action-date>
  </patent-classification>
</patent-classifications>
```

Source sub-field identifier

without trailing spaces

Recording of IPC (CPC is compatible to IPC) is described in [WIPO ST.8](#):

For the recording of IPC symbols on machine-readable records a field of 50 positions should be allotted for each symbol, the 50 positions of the field to be used as follows:

symbol, the 50 positions of the field to be used as follows: <i>Position(s)</i>	<i>Content</i>	<i>Values</i>
1	Section	A,...,H
2,3	Class	01,...,99
4	Subclass	A,...,Z
5 to 8	Main Group (right aligned)	1,...,9999, blank
9	Separating character	/ ("Slash")
10 to 15	Subgroup (left aligned)	00,...,999999, blank
16 to 19	For future use	4 blanks

20 to 27	Version indicator	YYYYMMDD date format
28	Classification level	C,A,S
29	First or later position of symbol	F,L
30	Classification value (invention or additional)	I,A
31 to 38	Action date	YYYYMMDD date format
39	Original or reclassified data	B,R,V,D
40	Source of classification data	H,M,G
41-42	Generating office	AA,...,ZZ (ST.3)
43-50	For future use	8 blanks

For each symbol, be sure to take the corresponding values of CPC_GENER_AUTH, CPC_VERSION, CPC_POSITION and CPC_VALUE from the same `patent_classification` element.

Comments

See the description of table TLS224_APPLN_CPC on how the CPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 13-03-2013 - First version

M. Kracker - 15-10-2014 - Comments updated

M. Kracker - 01-04-2016 – Examples showing the correct format have been added

6.30 CPC_GENER_AUTH

Name: CPC generating authority

Also Known As: n/a

Description: Patent office that classified the application with a CPC symbol

Domain: up to 2 characters (A-Z) or spaces;

- empty/spaces (when scheme is CPC, i.e. EP / US are assigning the CPC symbols)

- values according to [WIPO ST.3](#) (when scheme is CPCNO)

Default value: n/a

Source database: DOCDB

```
<patent-classification sequence="2">
  <classification-scheme office="EP" scheme="CPCNO">
    <date>20130101</date>
  </classification-scheme>
  <classification-symbol>B60V 1/16 </classification-symbol>
  <classification-value>I</classification-value>
  <classification-status>B</classification-status>
  <classification-data-source>H</classification-data-source>
  <generating-office>GB</generating-office>
  <action-date>
    <date>20130101</date>
  </action-date>
</patent-classification>
```

Source field name

```
<generating-office>GB</generating-office>
```

This field is only used for scheme "CPCNO".

This field is not used with scheme "CPC".

Source sub-field identifier: n/a

Comments

See the description of table TLS224_APPLN_CPC on how the CPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT

Within CPC_SCHEME = 'CPCNO', only a single authority may assign CPC symbols to an application. This should be the authority of this application.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 13-03-2013 - First version

M. Kracker - 15-10-2014 - Comments updated

6.31 CPC_POSITION

Name: First or later position of CPC symbol

Also Known As: n/a

Description: Indicates the position of the class symbol in the sequence of classes that form the classification. First / later indications are only available for CPC symbols allocated by the EPO or USPTO.

Domain: 1 character; F = first, L = later, space = unidentified

Default value: space

Source database: DOCDB

```
<patent-classifications>
  <patent-classification sequence="1">
    <classification-scheme office="EP" scheme="CPC">
      <date>20130101</date>
    </classification-scheme>
    <classification-symbol>B60V 1/16 </classification-symbol>
    <classification-value>I</classification-value>
    <symbol_position>L</symbol_position>
    <classification-status>B</classification-status>
    <classification-data-source>H</classification-data-source>
    <action-date>
      <date>20130101</date>
    </action-date>
  </patent-classification>
```

Source field name:

```
<symbol_position>L</symbol_position>
```

This field is only available for scheme "CPC".

This field is not used with scheme "CPCNO".

Source sub-field identifier: n/a

Comments

The following facts are asserted by DOCDB:

- only one CPC allocated to a given patent family will be identified to have symbol-position = "F" (first)
- CPC symbol identified by symbol-position = "F" (first) will always have classification-value = "I" (invention)
- the most recent CPC symbol allocated by the USPTO will be identified "first"
- failing the presence of a USPTO allocated CPC that can be identified "first", the most recent CPC symbol allocated by the EPO will be identified "first"
- all other CPC symbols allocated to a given patent family - whether USPTO or EPO, whether invention or additional - will have symbol-position = "L" (later)

For patent authorities where the law entails the concept of "first class", the first class symbol in a list of class symbols is the main class. For other authorities, like the EPO, there is no meaning in the position - classes may be quoted in alphabetical order for instance. Some researchers use a weighting technique to analyse by CPC.

See the description of table TLS224_APPLN_CPC on how the CPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 13-03-2013 - First version

M. Kracker - 15-10-2015 – Comment updated

6.32 CPC_SCHEME

Name: Classification scheme

Also Known As: n/a

Description:

The two schemes are:

CPC - CPC symbol allocated by the EPO or the USPTO

CPCNO - CPC symbol allocated by the National Office

Domain: up to 5 ASCII characters; CPC or CPCNO

Default value: n/a

Source database: DOCDB

Source field name

```
<patent-classification sequence="1">
  <classification-scheme office="EP" scheme="CPC">
    <date>20130101</date>
  </classification-scheme>
  <classification-symbol>G06F 17/30233 </classification-symbol>
  <symbol-position>F</symbol-position>
  <classification-value>I</classification-value>
  <classification-status>B</classification-status>
  <classification-data-source>H</classification-data-source>
  <action-date>
    <date>20130101</date>
  </action-date>
</patent-classification>
<patent-classification sequence="2">
  <classification-scheme office="EP" scheme="CPCNO">
    <date>20130101</date>
  </classification-scheme>
  <classification-symbol>G06F 9/06 </classification-symbol>
  <classification-value>I</classification-value>
  <classification-status>B</classification-status>
  <classification-data-source>H</classification-data-source>
  <generating-office>GB</generating-office>
  <action-date>
    <date>20130101</date>
  </action-date>
</patent-classification>
```

Source sub-field identifier: n/a

Comments

Comments

See the description of table TLS224_APPLN_CPC on how the CPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 13-03-2013 - First version

M. Kracker - 15-10-2015 – Comment added

6.33 CPC_VALUE

Name: Classification value

Also Known As: Invention / Additional

Description: Indication of the value of the classification i.e. is the class symbol relating to the invention or to aspects not related to the invention (but in the application).

Domain: 1 character; I=Invention A=Additional (Non-invention)

Default value: n/a

Source database: DOCDB

Source field name:

```
<patent-classifications>
  <patent-classification sequence="1">
    <classification-scheme office="EP" scheme="CPC">
      <date>20130101</date>
    </classification-scheme>
    <classification-symbol>B60V 1/16 </classification-symbol>
    <classification-value>I</classification-value>
    <classification-status>B</classification-status>
    <classification-data-source>H</classification-data-source>
    <action-date>
      <date>20130101</date>
    </action-date>
  </patent-classification>
</patent-classifications>
```

Source sub-field identifier: n/a

Comments

See the description of table TLS224_APPLN_CPC on how the CPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 13-03-2013 - First version

M. Kracker - 15-10-2015 – Comment updated

6.34 CPC_VERSION

Name: CPC version

Also Known As: n/a

Description: Version of the CPC

Domain: Date between '2013-01-01' and current date

Default value: n/a

Source database: DOCDB

Source field name:

```
<patent-classifications>
  <patent-classification sequence="1">
    <classification-scheme office="EP" scheme="CPC">
      <date>20130101</date>
    </classification-scheme>
    <classification-symbol>B60V 1/16 </classification-symbol>
    <classification-value>I</classification-value>
    <classification-status>B</classification-status>
    <classification-data-source>H</classification-data-source>
    <action-date>
      <date>20130101</date>
    </action-date>
  </patent-classification>
</patent-classifications>
```

Source sub-field identifier: n/a

Comments

See the description of table TLS224_APPLN_CPC on how the CPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 13-03-2013 - First version

M. Kracker - 15-10-2014 – Comment updated

6.35 CTRY_CODE

Name: Country code

Also Known As: cc

Description: The two letter code for the representation of states, other entities and intergovernmental organisations, as defined in WIPO standard ST.3

Domain: 2 ASCII characters

Default value: n/a

Source database: WIPO standard ST.3

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.36 DISCONTINUED

Name: Indicator whether a state or organisation no longer exists.

Also Known As: n/a

Description: Indicator whether a state or organisation no longer exists (according to WIPO standard ST.3).

Domain: 1 ASCII character: Y or space;
Y discontinued
space otherwise

Default value: n/a

Source database: [WIPO ST.3](#)

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.37 DOCDB_FAMILY_ID

Name: Identifier of a DOCDB simple family

Also Known As: DOCDB family ID; Simple family ID

Description:

Means that most probably the applications share exactly the same priorities (Paris Convention or technical relation or others) as contained in table TLS201_APPLN, TLS204_PRIOR_APPLN, TLS205_TECH_REL and TLS216_APPLN_CONTN.

Domain: Number 0 ... 999 999 999

Default value: n/a

Source database: DOCDB

Source field name

```
<exchange-document country="DE" doc-number="10331291" kind="A1" family-id="33441709" date="20050217" is-representative="Y" date-of-last-exchange="2006120611" date-of-previous-exchange="20050217" date-added-docdb="20050201" status="A">
```

In addition: For the dummy application (i.e. APPLN_ID = 0) and for artificial applications (i.e. APPLN_ID > 900 000 000) the value of the DOCDB_FAMILY_ID will be the same as the value of the APPLN_ID.

Source sub-field identifier

family-id

Comments

Every application belongs to exactly one DOCDB family. In the trivial case, an application belongs to a DOCDB family which consists of no other family members except this application itself. This is e.g. the case for all artificial applications (APPLN_ID > 900 000 000; see section 4.4).

Generally speaking, if two applications claim exactly the same prior applications as priorities (these can be e. g. Paris Convention priorities or technical relation priorities – for details see section 4.4.1 “Application replenishment for priorities”), then they are defined by the EPO as belonging to the same DOCDB simple family. The EPO reserves the right to classify an application into a particular simple family irrespective of this general rule - the EPO does this by creating artificial priorities for an application or by ignoring certain priorities (declaring them “inactive”) for the purpose of family building.

The simplified definition of the DOCDB family is that all their priorities must be the same. DOCDB family members generally refer to the same invention.

The simple family is also at times used to attribute automatically the same CPC classification symbols and other attributes to their family members.

As a general rule, the value of the DOCDB_FAMILY_ID will not change. It will be the same across editions of DOCDB and PATSTAT. However, corrections to priority numbers or changes in the priority pictures (priority numbers changing from active to inactive or vice-versa) might lead to a change in the family-ID of a given publication. See also section 4.3.2 “Stable IDs”.

Modification history

Author of update - Date of update - Explanation of update

J. Rollinson - 13-03-2008 - First version

D. Lingua - 14-05-2008 - Revised text

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 02-10-2013 - Extended for use in table TLS201_APPLN (PATSTAT Online Extension)

M. Kracker - 01-04-2015 – Revised comments; attribute in PATSTAT Online deprecated

M. Kracker - 01-10-2015 – Revised comments;

M. Kracker - 01-10-2016 – Revised comments;

6.38 DOCDB_FAMILY_SIZE

Name: Size of DOCDB simple family

Also Known As: n/a

Description: Size of DOCDB simple family of a given application

Domain: Number 1 ... about 1.000

Default value: n/a

Source database: PATSTAT

Source field name: Derived from table TLS201_APPLN

Source sub-field identifier: n/a

Comments:

A family size of 1 means that the application is the only member in that family.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-10-2015 – Source changed to TLS201_APPLN

M. Kracker - 01-10-2016 – Change of Domain. Minimum value is 1 (was: 0)

6.39 DOC_STD_NAME

Name: Standardised name as recorded in DOCDB

Also Known As: n/a

Description: Standard name attributed to applicant and inventor names for inclusion in DOCDB.

Domain: Up to 500 characters

Most names are only up to 30 characters in length.

Default value: n/a

Source database: DOCDB

Source field name

```
<applicants>
  <applicant sequence="1" data-format="docdb" status="A">
    <applicant-name>
      <name>MACDONALD ALEX BRUCE</name>
    </applicant-name>
    <residence>
      <country>US</country>
    </residence>
  </applicant>

<inventors>
  <inventor sequence="1" data-format="docdb">
    <inventor-name>
      <name>MACDONALD ALEX BRUCE</name>
    </inventor-name>
    <residence>
      <country>US</country>
    </residence>
  </inventor>
```

Source sub-field identifier

data-format="docdb"

Comments

It is not 100% certain that the DOCDB standardised names are always linked with the correct person name, in particular if the person information came from a source other than DOCDB. This is especially true for names in USPTO patents. The reason is that the matching algorithm which merges the different sources relies that the names are being listed in the same sequence in all data sources (DOCDB and others), which is sometimes not the case.

In case DOCDB does not provide a DOCDB standardized name, this attribute will contain the same data as the attribute PERSON_NAME.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 15-04-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-10-2013 - Added comments about use in PATSTAT Online

M. Kracker - 15-10-2014 – Comment updated; comment added to domain

M. Kracker – 01-04-2015 – Comment amended

M. Kracker – 01-12-2015 – Length of domain extended

6.40 DOC_STD_NAME_ID

Name: ID for the DOCDB standardized name

Also Known As: n/a

Description: Surrogate key based on the elements in the candidate primary key chosen

Domain: Number 1 ... 999 999 999

Default value: n/a

Source database: PATSTAT

Source field name: DOC_STD_NAME

Source sub-field identifier: n/a

Comments

Unique number for each DOC_STD_NAME.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 15-04-2005 - First version

M. Kracker - 15-10-2014 – Comment amended

M. Kracker – 01-04-2015 – Stability assertion removed; default value introduced

6.41 DOCUS_CLASS_SYMBOL

Name: Symbols defined within the DOCUS classification scheme of the USPTO

Also Known As: n/a

Description: The DOCUS scheme consists of USPTO classification symbols which are now superseded by CPC (Cooperative Patent Classification).

Domain: Up to 50 ASCII characters (as of April 2013, all symbols have a length between 8 and 19 characters)

Default value: n/a

Source database: DOCDB

Source field name

```
<patent-classifications>
  <patent-classification sequence="1">
    <classification-scheme office="US" scheme="DOCUS"/>
    <classification-symbol>081058000X</classification-symbol>
```

Source sub-field identifier

n/a

Comments

These classifications are being stored in DOCDB as supplied by the National Office without inspection of the contents. The EPO does not hold any responsibility for content, format or validity.

See also the [DOCDB User Documentation](#), Annex VI.

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 04-08-2011 - First version

6.42 EARLIEST_FILING_DATE

Name: Date of the earliest filing

Also Known As: n/a

Description: The earliest date of the filing dates of the application itself, its international application, its Paris Convention priority applications, the applications with which it is related via technical relations and its application continuations.

Only directly related applications are considered; this is unlike the INPADOC family, where applications might also be indirectly related.

Domain: Date (up to 9999-12-31)

Default value: 9999-12-31

Source database: PATSTAT

Source field name:

It is the APPLN_FILING_DATE of the earliest filing (see attribute EARLIEST_FILING_ID)

Source sub-field identifier: n/a

Comments:

For more details see EARLIEST_FILING_ID

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2014 – Description clarified

M. Kracker - 01-10-2015 – Name of attribute has changed (was PRIOR_EARLIEST_DATE)

6.43 EARLIEST_FILING_ID

Name: Application ID of the earliest filing

Also Known As: First filing

Description: The ID of the earliest application, considering the application itself, its international application, its Paris Convention priority applications, the applications with which it is related via technical relations and its application continuations.

Only directly related applications are considered; this is unlike the INPADOC family, where applications might also be indirectly related.

Domain: Number 0 ... 999 999 999;

Surrogate key: Technical unique identifier without any business meaning

Default value: n/a

Source database: PATSTAT

Source field name: Derived from the tables

- | | |
|----------------------|---|
| - TLS201_APPLN | self-priority |
| - TLS201_APPLN | PCT application (= international application) |
| - TLS204_APPLN_PRIOR | Paris Convention priority |
| - TLS205_TECH_REL | technical relations |
| - TLS216_APPLN_CONTN | application continuations |

Source sub-field identifier: n/a

Comments:

If multiple applications have been filed on the earliest filing date, then conceptually any of these applications can be regarded as the earliest application. Nevertheless, the logic to determine the application which has been filed first is like this:

1. If there is a PCT application which was filed on the earliest application date, then the APPLN_ID of this PCT application is taken as the EARLIEST_FILING_ID.
2. Else: If there are 1 or more Paris convention priorities which were filed on the earliest application date, then the Paris convention priority with the smallest APPLN_ID is taken as the EARLIEST_FILING_ID.
3. Else: the application which was filed on the earliest application date with the smallest APPLN_ID will be taken.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2015 - First version

M. Kracker - 01-04-2016 – Detailed rules for selection of earliest filed application has been added to comment.

6.44 EARLIEST_FILING_YEAR

Name: Year of the earliest filing date

Also Known As: n/a

Description: Year of the earliest filing date

Domain: 4 digits in the form yyyy (e. g. 2015)

Default value: n/a

Source database: PATSTAT

Source field name: Derived from attribute EARLIEST_FILING_DATE of table TLS201_APPLN.

It is the year component of the attribute EARLIEST_FILING_DATE.

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-10-2015 – Name of attribute has changed (was: PRIOR_EARLIEST_YEAR)

6.45 EARLIEST_PAT_PUBLN_ID

Name: ID of the earliest publication of an application

Also Known As: n/a

Description: The ID of a publication published on the earliest publication date of an application

Domain: Number 0 ... 999 999 999

Default value: 0

Source database: PATSTAT

Source field name: The earliest application date is indicated by attribute EARLIEST_PUBLN_DATE of table TLS201_APPLN. Table TLS211_PAT_PUBLN contains the publications with their ID (attribute PAT_PUBLN_ID).

Source sub-field identifier: n/a

Comments: If more than one publication is published on the same (earliest) publication date, then any one is selected.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.46 EARLIEST_PUBLN_DATE

Name: Date of earliest publication of an application

Also Known As: n/a

Description: Date of earliest publication of an application

Domain: Date (up to 9999-12-31)

Default value: 9999-12-31

Source database: PATSTAT

Source field name: Derived from table TLS211_PAT_PUBLN

It is the publication date PUBLN_DATE of the publication identified by

TLS211_PAT_PUBLN.PAT_PUBLN_ID = TLS201_APPLN.EARLIEST_PAT_PUBLN_ID .

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-10-2015 – Name of attribute has changed (was: PUBLN_EARLIEST_DATE)

6.47 EARLIEST_PUBLN_YEAR

Name: Year of the earliest publication date of an application

Also Known As: n/a

Description: Year of the earliest publication date of an application

Domain: 4 digits in the form yyyy (e. g. 2015)

Default value: n/a

Source database: PATSTAT

Source field name: Derived from attribute EARLIEST_PUBLN_DATE of table TLS201_APPLN ;

It is the year component of the attribute EARLIEST_PUBLN_DATE.

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2015 – Computation explained

M. Kracker - 01-10-2015 – Name of attribute has changed (was: PUBLN_EARLIEST_YEAR)

6.48 EPO_MEMBER

Name: Member of the European Patent Organisation

Also Known As: n/a

Description: Indicates whether this country is a member state of the EPO

Domain: 1 ASCII character: Y or space

Y If a country is member of the EPO. Only full members are considered,
no contracting states or extension states.
space otherwise

Default value: n/a

Source database: <http://www.epo.org/about-us/organisation/member-states.html>

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.49 EU_MEMBER

Name: Member of the European Community

Also Known As: n/a

Description: Indicates whether this country is a member state of the European Union

Domain: 1 ASCII character: Y or space

Y If a country is member of the EU
space otherwise

Default value: n/a

Source database: http://europa.eu/abc/european_countries/eu_members/index_de.htm

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.50 FIRST_NAME

Name: First name of a physical person

Also Known As: n/a

Description: Contains the first name of a physical person

Domain: Up to 500 characters

Default value: empty string

Source database: USPTO data of published applications and published grants

Source field name:

<addressbook> <first-name>

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2016 – Domain extended to 500 characters

6.51 GRANTED

Name: "Granted" indicator

Also Known As: n/a

Description: "1" if there exists a publication of the grant; "0" otherwise

Domain: 0 or 1

Default value: n/a

Source database: PATSTAT

Source field name: Derived from attribute PUBLN_FIRST_GRANT of table TLS211_PAT_PUBLN

Source sub-field identifier: n/a

Comments:

The same disclaimer as for attribute PUBLN_FIRST_GRANT applies:

Although the EPO has taken great care in analysing the grant information, this process is the result of interpretations and assumptions for which no responsibility whatsoever can be accepted.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2014 – Comment added

6.52 HAN_HARMONIZED

Name: Harmonization indicator for OECD HAN

Also Known As: n/a

Description: Indicates the degree of harmonization and standardization which could be achieved

Domain: Number 0 ... 2

- 0 the HAN_NAME has been replenished with the original name, because the name could not be harmonized.
- 1 the HAN_NAME has been harmonized but could not be matched with the ORBIS© database.
- 2 the HAN_NAME has been harmonized and could be matched with the ORBIS© database.

Default value: n/a

Source database: OECD HAN database

Source field name: n/a

Source sub-field identifier: n/a

Comments:

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available 3-4 months afterwards. Consequently, additions and changes introduced by the current PATSTAT edition are not harmonized in the current edition, but will be harmonized in the next edition of PATSTAT.

See also comment of attribute HAN_NAME.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2014 –Comment updated

M. Kracker - 01-04-2016 –Comment updated

6.53 HAN_ID

Name: ID of a Harmonized Applicant Name (HAN) from OECD

Also Known As: OECD HAN ID, HAN ID

Description: For each unique occurrence in attribute HAN_NAME an ID will be assigned. This will facilitate faster grouping.

Multiple rows may have the same HAN_ID, if multiple person names in table TLS206_PERSON have been harmonized into one HAN name.

Domain: Number 1 ... 999 999 999

Default value: n/a

Source database: OECD HAN database

Source field name: n/a

Source sub-field identifier: n/a

Comments:

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available 3-4 months afterwards. Consequently, additions and changes introduced by the current PATSTAT edition are not harmonized in the current edition, but will be harmonized in the next edition of PATSTAT.

See also comment of attribute HAN_NAME.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2015 – Comment updated

M. Kracker - 01-04-2015 – Comment updated

M. Kracker - 01-04-2016 –Comment updated

6.54 HAN_NAME

Name: Harmonized Applicant Name (HAN) from OECD

Also Known As: OECD HAN name, HAN name

Description: This field contains for many applicants the names as harmonized by the OECD HAN (Harmonized Applicant Name) project of the OECD.

The scope of this harmonization effort is described by the OECD as:

"The OECD HAN database, July 2014, provides groupings of patent applicant's names for the following set of countries or economies : {AR, AT, AU, BE, BR, CA, CH, CL, CN, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IL, IN, IS, IT, JP, KR, LU, MX, NL, NO, NZ, PL, PT, RU, SE, SI, SK, TR, TW, US, ZA}. The list of patents filed to the EPO, the USPTO and through the PCT is made available for each grouping of applicants. Further improvements are expected in future version, notably on the countries coverage."

The attribute is populated for all persons. Names of persons which have not been harmonized (e. g. persons who are inventors but not applicants) are just copied from the attribute PERSON_NAME.

Domain: Up to 500 characters

Default value: n/a

Source database: OECD HAN database

Source field name: n/a

Source sub-field identifier: n/a

Comments:

These names have been taken from the OECD HAN database (cf.

<http://www.oecd.org/sti/innovationinsciencetechnologyandindustry/oecdpatentdatabases.htm>).

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available 3-4 months afterwards. Consequently, additions and changes introduced by the current PATSTAT edition are not harmonized in the current edition, but will be harmonized in the next edition of PATSTAT.

Please note that the OECD HAN database is provided for research and analytical work. When publishing the results of your analysis, make sure it is quoted as: "OECD, HAN database, <Month, Year>".

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2014 – Update of description and comment

M. Kracker - 01-04-2016 –Comment updated

6.55 IMPACT **PATSTAT Legal Status**

Name: Impact of the legal event on the patent.

Also Known As: n/a

Description: Whether the legal event is perceived to have a positive, negative or neutral effect on the life of the patent. This attribute is an educated guess done by the EPO and has no legal basis.

Domain: 1 ASCII character: +, - or space

+ positive impact

- negative impact

space no impact

Default value: space

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.56 INPADOC_FAMILY_ID

Name: Identifier of an INPADOC extended priority family

Also Known As: INPADOC family ID; Extended family ID

Description: Means that the applications share a priority directly or indirectly via a third application. A 'priority' in this case means a link shown between applications as in tables TLS201_APPLN (regional/national phase of a PCT application), TLS204_APPLN_PRIOR (PARIS convention priorities), TLS205_TECH_REL (patents which have been technically linked by patent examiners on the basis of similar content) and table TLS216_APPLN_CONTN (continuations, divisions etc.).

Domain: Number 0 ... 999 999 999

Default value: n/a

Source database: This attribute is calculated during the preparation of PATSTAT data. For the dummy application (i.e. APPLN_ID = 0) and for artificial application replenished because of citations (i.e. APPLN_ID > 930 000 000) the value of the INPADOC_FAMILY_ID will be the same as the value of the APPLN_ID.

Source field name: n/a

Source sub-field identifier: n/a

Comments

Every application belongs to exactly one INPADOC family. In the trivial case, an application belongs to an INPADOC family which consists of no other family members except this application itself. This is e.g. the case for all artificial applications with an APPLN_ID > 930 000 000; see section 4.4.24.4).

Much patent research is affected by the "family" concepts. There are various definitions of how to link different patents into "families". This INPADOC extended priority family was developed by the INPADOC organisation before it was integrated into the EPO.

For the computation of the INPADOC families these tables are taken into account:

- TLS201_APPLN
A PCT application in its regional/national phase contains in its attribute INTERNAT_APPLN_ID the APPLN_ID of its original PCT application
- TLS204_appln_prior (Paris convention priorities)
- TLS205_TECH_REL (patents which have been technically linked by patent examiners on the basis of similar content)
- TLS216_appln_contn (continuations, divisions, etc.).

The EPO reserves the right to apply this rule as needed for its internal purposes.

The simplified definition of the INPADOC family is that the members relate in some way (directly or indirectly) to the "first" application. Generally an INPADOC family covers one or more DOCDB families and covers a set of related inventions.

From a statistical point of view: a large DOCDB family might indicate that the applicant seeks a wide geographical protection for the invention.

The value of the INPADOC_FAMILY_ID is not stable but will change with every edition of PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

J. Rollinson / D. Lingua - 19-09-2008 - First version

D. Lingua - 11-10-2011 - Comment on SQL queries eliminated

M. Kracker - 01-10-2013 - Change of domain

M. Kracker - 01-10-2015 – Value 0 added to domain; Change of comment and description

M. Kracker - 01-10-2016 – Change of comment

6.57 INT_PHASE

Name: Indicator whether the application *is* or *has been* in the international phase

Also Known As: n/a

Description: Indicates that an application *is* or *has been* in the international phase.

This covers all international filings at the receiving office as well as all applications based on these filings.

Domain: 1 ASCII character

Y Yes

N No

space not known (In case of uncertain interpretations; used very little or not at all)

Default value: N

Source database: PATSTAT

Source field name: Derived from table TLS201_APPLN

Y if the application has APPLN_KIND = W (i.e. international filing)
or INTERNAT_APPLN_ID > 0; (i.e. based on internat. application)

N otherwise

Source sub-field identifier: n/a

Comments:

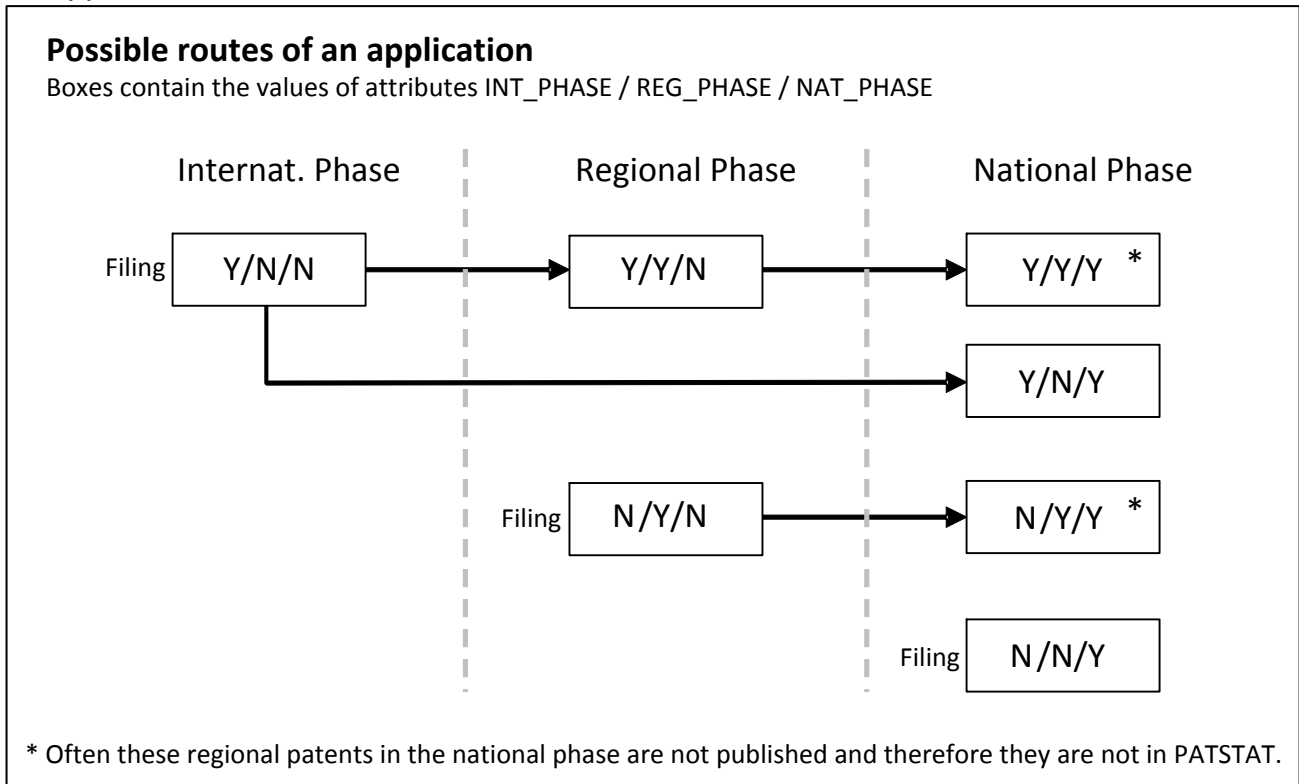
These indicators provide a somewhat simplistic approach to identify the route an application has taken. Although the EPO has taken great care in analysing the underlying data (especially publication and application kind codes), this process is the result of interpretations and assumptions for which no responsibility whatsoever can be accepted.

Please note that these indicators only help to understand applications which actually exist in PATSTAT. It does not help to answer questions like "How many EP applications are valid in country x", because not every office publishes patents which are validated / granted in their country. Consequently, there is no publication or application in PATSTAT for every granted patent.

The same will apply for the Unitary Patents, if there is no publication for that.

Especially if you want to count the patents of an office which is a member of a regional office: As argued above, it is not sufficient to consider the indicator NAT_PHASE. Depending on the office, you must also analyse the legal events of the application in the regional phase (see also table TLS221_INPADOC_PRS or the database "PATSTAT Register" for the legal events of EP applications.

An application can take one of these routes:



Application ...	Values in attributes		
	INT_PHASE	REG_PHASE	NAT_PHASE
... in the international phase	Y	N	N
... PCT in the regional phase (e. g. Euro-PCT)	Y	Y	N
... PCT via regional office, now in national phase	Y	Y	Y
... PCT in the national phase (no regional phase)	Y	N	Y
... in the regional phase (no PCT)	N	Y	N
... via regional office, now in national phase (no PCT)	N	Y	Y
... national application (no PCT)	N	N	Y

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2016 - First version

6.58 INTERNAT_APPLN_ID

Name: Application identification of the earlier PCT international application for an application.

Also Known As: n/a

Description: Technical unique identifier without any business meaning

Domain: Number 0 ... 999 999 999

Default value: 0

Source database: DOCDB, PATSTAT

Source field name

International applications designating the Authority of the related National /Regional application. The **latter** is published with an INID-code in the 80-series ([WIPO ST.9](#)). The case to be taken into account is case # 3 from section 4.6 "Relation Types".

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>US</country>
      <doc-number>0107931</doc-number>
      <kind>W</kind>
      <date>20010312</date>
    </document-id>
    <priority-linkage-type>W</priority-linkage-type>
  <priority-active-indicator>N</priority-active-indicator>
</priority-claim>
```

With

```
<document-id>
  <country>US</country>
  <doc-number>0107931</doc-number>
  <kind>W</kind>
```

in DOCDB the corresponding international application in PATSTAT is determined (via APPLN_AUTH, APPLN_NR and APPLN_KIND) and the value of INTERNAT_APPLN_ID for this national/regional application is set to the value of APPLN_ID of the international application. If there is no corresponding international application in PATSTAT it should be created, see section 4.4 "Application replenishment".

With the October 2012 edition, this additional rule has been implemented to further identify national applications originating from the PCT:

If a PCT priority is the only priority claimed then LMI = 'A' is to be considered as LMI='W', i.e. an indicator for identifying "entry into the national phase". See following usage example:

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>EP</country>
      <doc-number>2005009340</doc-number>
      <kind>W</kind>
      <date>20050830</date>
    </document-id>
    <priority-linkage-type>A</priority-linkage-type>
  </priority-claim>
```

As this is the only priority, the priority linkage type should be actually considered as being <priority-linkage-type>W</priority-linkage-type>

Source sub-field identifier

n/a

Comments

The default value 0 means this application has no earlier PCT application.

If the value of INTERNAT_APPLN_ID is > 0, then this application does have an earlier PCT application, whose APPLN_ID equals the value of INTERNAT_APPLN_ID.

Warning: the linkage type information <priority-linkage-type>W</priority-linkage-type> is key to recognising whether an application originates from a PCT filing or not. Due to an incomplete usage of the "W" linkage type indicator in DOCDB, the usage of the element INTERNAT_APPLN_ID might not be sufficient to isolate all applications originating from the PCT. For EP A publications, a claim count equal to zero is a good indicator for an application originating from the PCT, see also the warning under attribute PUBLN_CLAIMS.

Note that for some countries there will be no applications with INTERNAT_APPLN_ID >0, because for these countries the "national route" via the PCT has been closed (for a list of these countries see http://www.wipo.int/export/sites/www/pct/en/list_states.pdf, footnote 2). For example, France does not accept PCT applications to go directly to France. Instead, for the PCT application the appropriate regional office (for FR it is EP) must be designated first, and the granted regional patent may then be validated in France.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 03-05-2005 - First version

R. Heijna - 20-07-2005 - Source field definition improved

R. Heijna - 07-07-2005 - Value zero for the physical model

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 27-07-2010 - Revision of text

D. Lingua - 17-04-2011 - Warning added

D. Lingua - 08-10-2012 - Additional rule to identify Euro-PCTs

M.Kracker - 10-04-2015 – Added comment: for some countries the national phase via the PCT is closed

6.59 INVT_SEQ_NR

Name: Sequence number of inventor

Also Known As: n/a

Description: Number indicating the place in the list of inventors in the application

Domain: Number 0 ... about 250

Default value: 0

Source database:

1) EPO Register for EP patent applications
Contains the sequence numbers.

2) OECD patents database for US data post 1976-01-01 up to and including November 15th 2005 for Published Grants. This data does not contain sequence numbers, so they are allocated within PATSTAT.

3) PATSTAT weekly file extracts from USPTO website:
Published Grants from November 22nd 2005 until today;
Published Applications from September 29th 2005 to today inclusive.
This data contains the sequence numbers.

4) Inventor & Applicant names for USPTO Published Applications from March 1st 2001 to September 22nd 2005 from DOCDB, data-format="docdba". This data contains the sequence numbers.

5) all other names from DOCDB, data-format="docdba". This data contains the sequence numbers.

Source field name

```
<inventor sequence="1" data-format="docdba">
  <inventor-name>
    <name>WHITTUM- HUDSON, JUDITH A</name>
  </inventor-name>
</inventor>
<inventor sequence="2" data-format="docdba">
  <inventor-name>
    <name>MACDONALD, ALEX BRUCE, </name>
  </inventor-name>
</inventor>
<inventor sequence="3" data-format="docdba">
  <inventor-name>
    <name>AN, LING LING</name>
  </inventor-name>
</inventor>
```

Source sub-field identifier

data-format="docdba"

Comments

An entry with a value 1 to n represents an inventor, an entry with the value 0 does not represent an inventor, but another person (e. g. an applicant). It is possible that there are applications with no inventors.

Consequently, adding the condition "INVT_SEQ_NR > 0" to the WHERE clause in a query retrieves only those persons from TLS207_PERS_APPLN or TLS227_PERS_PUBLN which are inventors.

Likewise, adding the condition "APPLT_SEQ_NR > 0 AND INVT_SEQ_NR > 0 " retrieves only persons which for a certain application are applicants as well as inventors.

For US data : Documents published after 1976-01-01: For the inventors , the sequence numbers are all given arbitrarily, with the exception of the documents published after March 2005, where the sequence numbers are all correct.

For all US documents published before 1976-01-01, where the data was taken from DOCDB, the sequence numbers are believed to be correct

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 19-04-2005 - First version

R. Heijna - 07-07-2005 - Value zero for the physical model

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-10-2013 - Changed source from EPO Bulletin to EPO Register;
changed domain

6.60 IPC

Name: IPC subclass / IPC main group

Also Known As: n/a

Description: First 4 – 8 characters of an IPC symbol according to WIPO ST.3. (In most cases they are only the first 4 characters)

Domain: Up to 8 ASCII characters; Example: 'B01D', 'A61K 6'

Default value: n/a

Source database: See Eurostat's paper described in section 5.27 "TLS902_IPC_NACE2: Mapping between IPC and industrial sectors".

Source field name: n/a

Source sub-field identifier: n/a

Comments: See attribute IPC_CLASS_SYMBOL for the full IPC symbol.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2015 – Overall amendment, because the attribute has been moved from table INDUSTRY_IPC to table TLS902_IPC_NACE.

6.61 IPC_CLASS_LEVEL

Name: IPC classification level indicator

Also Known As: n/a

Description: Denotes whether an authority classified either in the full IPC, in main groups or in sub classes only.

Domain: 1 character:

A = classification in the full IPC	e.g. 'H04Q 7/32', 'C07K 14/00'
C = classification in main groups only	e.g. 'H04H 1/00', 'A61K 31/00'
S = classification in subclasses only	e. g. 'H04H', 'A61K'

Default value: n/a

Source database: DOCDB

Source field name

```
<classifications-ipcr>
  <classification-ipcr sequence="1">
    <text>A43C 11/00          20060101CFI20070118BHUS      </text>
  </classification-ipcr>
  <classification-ipcr sequence="2">
    <text>A43C 11/00          20060101AFI20070118BHUS      </text>
  </classification-ipcr>
</classifications-ipcr>
```

Source sub-field identifier

positions 28 of the source-field

```
.....12345678901234567890123456789012345678901234567890
<text>A43C 11/00          20060101CFI20070118BHUS      </text>
```

These text strings are all 50 bytes long. See [WIPO ST.8](#).

Take byte 28 as the value of IPC_CLASS_LEVEL.

Source sub-field identifier

position 28

Comments

See the description of table TLS209_APPLN_IPC on how the IPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

J. Rollinson - 27-08-2007 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 31-03-2011 - Roll up of Core symbols to Advanced

D. Lingua - 07-10-2011 - Value "S" (Symbol) has been eliminated in DOCDB

M. Kracker - 15-10-2014 - All levels A, C, S are available.

6.62 IPC_CLASS_SYMBOL

Name: IPC classification symbol (IPC 8th edition)

Also Known As: (IPC) class, (IPC) classification

Description: Classification symbol according to the International Patent Classification, eighth edition (entered into force January 1, 2006)

Domain: Up to 15 characters (A-Z, 0-9, /, space) as allowed by IPC;

Examples: A61K
 H04Q 7/32
 C07K 14/00
 C07D 405/06
 H01M2220/20

Note that spaces may be required on position 5-7, because the slash "/" is always on the 9th position. For more details see the table below.

Default value: n/a

Source database: DOCDB

Source field name:

```
<classifications-ipcr>
  <classification-ipcr sequence="1">
    <text>A43C 11/00 20060101CFI20070118BHUS </text>
  </classification-ipcr>
  <classification-ipcr sequence="2">
    <text>A43C 11/00 20060101AFI20070118BHUS </text>
  </classification-ipcr>
</classifications-ipcr>
```

Source sub-field identifier

positions 1-15 of the source-field

Source codes

```
.....12345678901234567890123456789012345678901234567890
<text>A43C 11/00 20060101CFI20070118BHUS </text>
```

These text strings are all 50 bytes long. See [WIPO ST.8](#):

For the recording of IPC symbols on machine-readable records a field of 50 positions should be allotted for each symbol, the 50 positions of the field to be used as follows:

symbol, the 50 positions of the field to be used as follows: <i>Position(s)</i>	<i>Content</i>	<i>Values</i>
1	Section	A,...,H
2,3	Class	01,...,99
4	Subclass	A,...,Z
5 to 8	Main Group (right aligned)	1,...,9999, blank
9	Separating character	/ ("Slash")
10 to 15	Subgroup (left aligned)	00,...,999999, blank
16 to 19	For future use	4 blanks
20 to 27	Version indicator	YYYYMMDD date format

28	Classification level	C,A,S
29	First or later position of symbol	F,L
30	Classification value (invention or additional)	I,A
31 to 38	Action date	YYYYMMDD date format
39	Original or reclassified data	B,R,V,D
40	Source of classification data	H,M,G
41-42	Generating office	AA,...,ZZ (ST.3)
43-50	For future use	8 blanks

Take the first 15 bytes as the value of IPC_CLASS_SYMBOL. For each symbol, be sure to take the corresponding values of IPC_GENER_AUTH, IPC_VERSION, IPC_POSITION, IPC_VALUE and IPC_CLASS_LEVEL from the same `classification-ipcr` element.

Comments

See the description of table TLS209_APPLN_IPC on how the IPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 19-04-2005 - First version

J. Rollinson - Aug 2007 - Addition of "Advanced" symbols

D. Lingua - 16-04-2009 - Amended text

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 19-02-2010 - "Source code" description modified

J. Rollinson - 01-04-2011 - Core no longer maintained by WIPO

M. Kracker - 15-10-2014 – Comment updated.

M. Kracker - 01-04-2016 – Examples showing the correct format have been added

6.63 IPC_GENER_AUTH

Name: IPC generating authority

Also Known As: n/a

Description: Patent office that generated the IPC classification of the application concerned

Domain: 2 ASCII characters (A-Z), according to [WIPO ST.3](#)

Default value: n/a

Source database: DOCDB

Source field name

```
<classifications-ipcr>
  <classification-ipcr sequence="1">
    <text>A43C 11/00 20060101CFI20070118BHUS </text>
  </classification-ipcr>
  <classification-ipcr sequence="2">
    <text>A43C 11/00 20060101AFI20070118BHUS </text>
  </classification-ipcr>
</classifications-ipcr>
```

Source sub-field identifier

Position 41-42: Generating office AA, ZZ (ST.3)

Comments

See [WIPO ST.8](#).

See the description of table TLS209_APPLN_IPC on how the IPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 31-10-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 15-05-2013 - Added exception to Domain

M. Kracker - 15-10-2014 – Comment updated; removed exception to domain.

6.64 IPC_MAINGROUP_SYMBOL

Name: IPC subclass or IPC main group

Also Known As: n/a

Description: The subclass (i.e. first 4 characters) or main group (i.e. first 8 characters) of an IPC symbol according to WIPO ST.3

Domain: 4 or 8 ASCII characters;

Examples: 'A61K', 'A61K 8', 'A61K 133'

Note: Spaces are relevant, as with all IPC or CPC symbols.

Default value: n/a

Source database:

http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/xls/ipc_technology.xls

Source field name:

Column IPC_CODE of the above mentioned Excel file (without trailing %-sign)

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.65 IPC_POSITION

Name: First or later position of symbol

Also Known As: n/a

Description: Indicates the position of the class symbol in the sequence of classes that form the classification

Domain: 1 character: F=first, L=later. space =unidentified

Default value: space

Source database: DOCDB

Source field name:

```
<classifications-ipcr>
  <classification-ipcr sequence="1">
    <text>A43C 11/00 20060101CFI20070118BHUS </text>
  </classification-ipcr>
  <classification-ipcr sequence="2">
    <text>A43C 11/00 20060101AFI20070118BHUS </text>
  </classification-ipcr>
</classifications-ipcr>
```

If there is a space in <classification-ipcr> in position 29, then record a space in PATSTAT in IPC_POSITION.

Source sub-field identifier

Position 29: First or later position of symbol F, L

Comments

See [WIPO ST.8](#) for an explanation.

For patent authorities (e. g. USPTO) where the law entails the concept of "first" class, the first class symbol in a list of class symbols is the main class. For other authorities, like the EPO, there is no meaning in the position - classes may be quoted in alphabetical order for instance. Some researchers use a weighting technique to analyse by IPC.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 19-04-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

J. Rollinson - 01-04-2011 - Core no longer maintained by WIPO

M. Kracker - 15-10-2014 – Comment updated

6.66 IPC_VALUE

Name: Classification value

Also Known As: Invention / Additional; Inventive/Non-Inventive

Description: Indication of the value of the classification i.e. is the class symbol relating to the invention or to aspects not related to the invention (but in the application).

Domain: 1 character: I=Invention, N=Additional (Non-Invention)

Default value: n/a

Source database: DOCDB

Source field name:

```
<classifications-ipcr>
  <classification-ipcr sequence="1">
    <text>A43C 11/00 20060101CFI20070118BHUS </text>
  </classification-ipcr>
  <classification-ipcr sequence="2">
    <text>A43C 11/00 20060101AFI20070118BHUS </text>
  </classification-ipcr>
</classifications-ipcr>
```

Source sub-field identifier

Position 30: Classification value (inventive or non-inventive) I, N

Comments

See [WIPO ST.8](#) for an explanation.

See the description of table TLS209_APPLN_IPC on how the IPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Invention related IPC symbols are printed bold on the front page of patent documents, according to WIPO standard ST.10/C.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 19-04-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

J. Rollinson - 01-04-2011 - Core no longer maintained by WIPO

D. Lingua - 16-08-2012 - Remark on bold prints added

M. Kracker - 15-10-2014 – Comment updated

6.67 IPC_VERSION

Name: IPC version

Also Known As: n/a

Description: Version of the IPC

Domain: Date between '2006-01-01' and current date

Default value: n/a

Source database: DOCDB

Source field name:

```
<classifications-ipcr>
  <classification-ipcr sequence="1">
    <text>A43C 11/00          20060101CFI20070118BHUS      </text>
  </classification-ipcr>
  <classification-ipcr sequence="2">
    <text>A43C 11/00          20060101AFI20070118BHUS      </text>
  </classification-ipcr>
</classifications-ipcr>
```

Source sub-field identifier

20 to 27 Version indicator YYYYMMDD date format

Comments

See [WIPO ST.8](#) for an explanation.

See the description of table TLS209_APPLN_IPC on how the IPC symbols, which are allocated in DOCDB to publications, are de-duplicated and assigned to applications in PATSTAT.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 19-04-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

J. Rollinson - 01-04-2011 - Core no longer maintained by WIPO

M. Kracker - 15-10-2014 - Comment updated

6.68 IPR_TYPE

Name: Type of Intellectual Property Right

Also Known As: n/a

Description: Type of Intellectual Property Right

Domain: 2 ASCII characters: PI, UM, DP;

PI - Patent of Invention

UM - Utility Model

DP - Design Patent

Default value: n/a

Source database: PATSTAT

Source field name: APPLN_AUTH, APPLN_KIND, PUBLN_KIND

Source sub-field identifier: n/a

Source codes

If first character of APPLN_KIND is 'U' or 'V' or 'Y' or 'Z', or

(APPLN_AUTH = 'FR' and APPL_KIND = 'A' and at least one related publications has a PUBLN_KIND = 'A3' or 'A4' or 'A7' or 'A8')

then IPR_TYPE = 'UM' for utility model

else if APPLN_KIND = 'F ' and APPLN_AUTH is not 'FR' then IPR_TYPE = 'DP' for design patent.

For all other values of APPLN_KIND, set IPR_TYPE to 'PI' for Patent of Invention. Note that in America, a Patent of Invention is known as a Utility Patent.

This rule applies to all instances of APPLN_KIND, whether it is derived from application-reference or a priority-reference.

Comments

The rule to compute utility models and design patents does cover all major, but not necessarily all cases. The rule may be improved in the future.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 12-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML.
Added Design Patent info.

M. Kracker - 01-10-2013 - Changed rule to compute the IPR_TYPE; added comment

6.69 ISO_ALPHA3

Name: 3-letter country code

Also Known As: ISO 3166 alpha-3 code

Description: The three letter code for the representation of states, as defined in standard ISO 3166

Domain: 3 ASCII letters [A-Z] (for states) or spaces (for other entities and intergovernmental organisations)

Default value: spaces

Source database:

ISO 3166 alpha-3 codes:

http://www.nationsonline.org/oneworld/country_code_list.htm

Deleted entries:

See section "Deleted Codes" in http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3

Source field name

n/a

Source sub-field identifier

n/a

Comments

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2015 - First version

6.70 JP_CLASS_SCHEME

Name: Description of the JP classification scheme

Also Known As: n/a

Description:

The two schemes for JP classification are:

FI - File Index

FTERM - File Term.

Domain: Up to 5 ASCII characters: FI, FTERM

Default value: n/a

Source database: DOCDB

Source field name

```
<patent-classifications>
  <patent-classification sequence="1">
    <classification-scheme office="JP" scheme="FI"/>
    <classification-symbol>4F21V8 /00 231</classification-symbol>
  </patent-classification>
  <patent-classification sequence="2">
    <classification-scheme office="JP" scheme="FTERM"/>
    <classification-symbol>4H129/BA20</classification-symbol>
  </patent-classification>
</patent-classifications>
```

Source sub-field identifier

n/a

Comments

These classifications are being stored in DOCDB as supplied by the National Office without inspection of the contents. The EPO does not hold any responsibility for content, format or validity.

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 04-08-2011 - First version

6.71 JP_CLASS_SYMBOL

Name: Symbols defined within the JP classification scheme

Also Known As: n/a

Description: The two schemes FI and FTERM consist of symbols, which can be up to 50 characters long.

Domain: Up to 50 characters (almost all symbols are between 10 and 18 characters long)

Default value: n/a

Source database: DOCDB

Source field name

```
<patent-classifications>
  <patent-classification sequence="1">
    <classification-scheme office="JP" scheme="FI"/>
    <classification-symbol>4F21V8 /00 231</classification-symbol>
  </patent-classification>
  <patent-classification sequence="2">
    <classification-scheme office="JP" scheme="FTERM"/>
    <classification-symbol>4H129/BA20</classification-symbol>
  </patent-classification>
</patent-classifications>
```

Source sub-field identifier

n/a

Comments

These classifications are being stored in DOCDB as supplied by the National Office without inspection of the contents. The EPO does not hold any responsibility for content, format or validity.

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 04-08-2011 - First version

6.72 L501EP **PATSTAT Legal Status**

Name: National authority which refines a regional legal event code

Also Known As: n/a

Description: The meaning of this attribute depends on the attribute PRS_CODE

- PRS_CODE = "REG":
Corresponding country code for the national legal event:
The "REG" code is used in the offices EP, WO, SU (in case a SU application results in a RU patent), GB (in case of Hong Kong) and CN (in case of Hong Kong)
- PRS_CODE = "PGFP", "PG25" or "PGRI":
Country code for which the information from the European post grant system is valid.
These 3 codes are only used in the office EP.

Domain: 2 characters (A-Z), according to [WIPO ST.3](#) or empty

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L501EP`

Source sub-field identifier: n/a

Comments:

Due to legacy data, this attribute L501EP may also contain some values for other PRS_CODES than REG, PGFP, PG25 or PGRI. In these cases the content of L501EP should be ignored.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01 - 11-2013 - Comment added

6.73 L502EP **PATSTAT Legal Status**

Name: National legal event code which refines a regional legal event code

Also Known As: n/a

Description: Contains the corresponding national legal event code if PRS_CODE = "REG", which is used by the authorities EP, WO, SU (in case a SU application results in a RU patent), GB (in case of Hong Kong) and CN (in case of Hong Kong). Is empty otherwise.

Domain: Up to 4 ASCII characters which represent a legal event code

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L502EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.74 L503EP **PATSTAT Legal Status**

Name: Document number

Also Known As: n/a

Description: Number of the national publication or application, in case a new number is introduced. Examples:

- PCT applications entering the national phase are typically assigned a new national application number
- DE, AT, ES, EE, SK and GR assign new (national) publication numbers to EP patents

Domain: Up to 20 ASCII characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L503EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.75 L504EP **PATSTAT Legal Status**

Name: Country code of document

Also Known As: n/a

Description: Country code of the national publication or application, in case a new number is introduced. See also description of attribute L503EP.

Domain: 2 characters (A-Z), according to [WIPO ST.3](#) or empty

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L504EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.76 L505EP **PATSTAT Legal Status**

Name: Publication date of document (publication)

Also Known As: n/a

Description: Publication date of the national publication or application, in case a new number is introduced. See also description of attribute L503EP.

Domain: date

Default value: 9999-12-31

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L505EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.77 L506EP **PATSTAT Legal Status**

Name: Kind code of document (publication)

Also Known As: n/a

Description: Kind code of the national publication or application, in case a new number is introduced. See also description of attribute L503EP.

Domain: Up to 2 ASCII characters or empty

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L5016EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.78 L507EP **PATSTAT Legal Status**

Name: Designated state(s)

Also Known As: n/a

Description: List of country codes of designated states

Domain: Up to 300 characters (A-Z), according to [WIPO ST.3](#);

The country codes of the designated states are concatenated without any separator.

Example: "ATBECHDEDKESFRGBGRIEITLILUMCNLPTSE"

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L507EP`

Source sub-field identifier: n/a

Comments:

Patent protection can be sought in multiple ways, either via designation through the WIPO international application or via designation through a regional office, or by filing directly a patent application in a national office

Currently, PCT and EP applications automatically designate all states - so this attribute L507EP does not carry much information. Designating all states is either expressed by an empty list or by a list of all possible designated states.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-10-2015 – Comment amended

6.79 L508EP **PATSTAT Legal Status**

Name: Extension or validation state(s)

Also Known As: n/a

Description: List of country codes of extension state or validation states. Regional groupings of national patent offices exist such as the EPO. These groups sometimes allow applicants to extend protection to non-member states.

Domain: Up to 20 characters (A-Z), according to [WIPO ST.3](#) or empty. The country codes of the extension or validation states are concatenated without any separator.

Example: "ALBAMERS"

In the huge majority, this attribute contains at most 1 country code.

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L508EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.80 L509EP **PATSTAT Legal Status**

Name: New owner

Also Known As: n/a

Description: New owner if name or address of an owner has changed; Sometimes the *old* owner is stored in the free format text field L510EP.

Domain: Up to 255 characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag L509EP

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.81 L510EP **PATSTAT Legal Status**

Name: Additional information

Also Known As: n/a

Description: Additional information in free form text relating to an event which is not covered by another attribute L5xxEP. Examples are product names for SPCs; indications whether a license is exclusive or non-exclusive; indication which data has been corrected, ...

Domain: Up to 700 characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L510EP`

Source sub-field identifier: n/a

Comments: Almost all strings have been truncated at PATSTAT's data source, the INPADOC worldwide legal status database. Only a minority (about 400 instances in the 2013 October edition) of the values exceeds 50 characters.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M.Kracker – 01-04.2014 – Comment added

6.82 L511EP **PATSTAT Legal Status**

Name: SPC number

Also Known As: n/a

Description: Application and/or publication number of Supplementary Protection Certificate

Domain: Up to 20 ASCII characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L511EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.83 L512EP **PATSTAT Legal Status**

Name: Filing date of SPC, opposition or similar

Also Known As: n/a

Description: Filing date of Supplementary Protection Certificate, opposition or similar

Domain: date

Default value: 9999-12-31

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L512EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.84 L513EP **PATSTAT Legal Status**

Name: Expiry date of IP right (except SPC)

Also Known As: n/a

Description: Expiry date of IP right. Note: Expiry dates for SPC are stored in L523EP.

Domain: date

Default value: 9999-12-31

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L513EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.85 L515EP **PATSTAT Legal Status**

Name: Inventor name

Also Known As: n/a

Description: New or changed inventor name, if inventor name is corrected

Domain: Up to 255 characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L515EP`

Source sub-field identifier: n/a

Comments: Almost all strings have been truncated at PATSTAT's data source, the INPADOC worldwide legal status database. Only a minority (about 25 instances in the 2013 October edition) of the values exceeds 50 characters (max 55 characters in the 2013 October edition).

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2014 – Comment added

6.86 L516EP **PATSTAT Legal Status**

Name: IPC

Also Known As: n/a

Description: New International Patent Classification symbol(s) in case of additions or changes

Domain: Up to 50 ASCII characters (not necessarily strictly conformant to [WIPO ST.8](#))

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag L516EP

Source sub-field identifier: n/a

Comments: This data is not as complete as the IPC data in table TLS209_APPLN_IPC and it is never re-classified.

This attribute might be removed in future PATSTAT editions.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.87 L517EP **PATSTAT Legal Status**

Name: Representative's name

Also Known As: n/a

Description: Name of representative (patent attorney) in case it has been changed or corrected

Domain: Up to 255 characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L517EP`

Source sub-field identifier: n/a

Comments: Almost all strings have been truncated at PATSTAT's data source, the INPADOC worldwide legal status database. Only a minority (265 instances in the 2013 October edition) of the values exceeds 50 characters.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2014 – Comment added

6.88 L518EP **PATSTAT Legal Status**

Name: Fee payment date

Also Known As: n/a

Description: Date of fee payment

Domain: date

Default value: 9999-12-31

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag L518EP

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.89 L519EP **PATSTAT Legal Status**

Name: Opponent name

Also Known As: n/a

Description: Name of the opponent (for events in opposition or re-examination procedures)

Domain: Up to 255 characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L519EP`

Source sub-field identifier: n/a

Comments: Many strings have been truncated at PATSTAT's data source, the INPADOC worldwide legal status database. In the 2013 October edition there were no values which exceed 50 characters.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2014 – Comment added

6.90 L520EP **PATSTAT Legal Status**

Name: Year of renewal fee payment

Also Known As: n/a

Description: Year of renewal fee payment, e. g. 7 for the seventh year.

Domain: Number 0 ... 25

Default value: 0

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag L520EP

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2014 – Domain changed to numeric

M. Kracker - 01-04-2015 – Domain changed to 0 ... 25

6.91 L522EP **PATSTAT Legal Status**

Name: Name of licensee or other requestor

Also Known As: n/a

Description: Licensee, for license events; or for other events a requester who is neither the owner nor the inventor nor the representative

Domain: Up to 255 characters

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L522EP`

Source sub-field identifier: n/a

Comments: As of 2014 Autumn Edition, all strings have been truncated at PATSTAT's data source, the INPADOC worldwide legal status database to 50 characters.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2014 – Comment added

6.92 L523EP **PATSTAT Legal Status**

Name: Expiry date of SPC

Also Known As: n/a

Description: Expiry date of a Supplementary Protection Certificate.

Note: L513EP is only used for the corresponding patent.

Domain: date

Default value: 9999-12-31

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L523EP`

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.93 L524EP **PATSTAT Legal Status**

Name: Countries concerned

Also Known As: n/a

Description: List of country codes

Domain: Up to 100 characters (A-Z), according to [WIPO ST.3](#)

The country codes of the countries concerned are concatenated without any separator.

Example: "ALLTLVMKROSI"

Default value: empty

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag L524EP

Source sub-field identifier: n/a

Comments: This attribute is rarely used.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.94 L525EP **PATSTAT Legal Status**

Name: Effective date

Also Known As: n/a

Description: Effective date of the legal event.

If set to the default value '9999-12-31': see also attribute PRS_GAZETTE_DATE

Domain: date

Default value: 9999-12-31

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: Take the value of tag `L525EP`

Source sub-field identifier: n/a

Comments: Most offices deliver for certain legal events the effective date of an event. If it is not delivered (= set to default value), then in many cases the PRS_GAZETTE_DATE attribute could be used instead.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.95 LAST_NAME

Name: Last name / Organization name

Also Known As: n/a

Description: Contains the last name (family name, surname) of a physical person or the name of a legal person

Domain: Up to 500 characters

Default value: empty string

Source database: USPTO data of published applications and published grants

Source field name:

<addressbook> <last-name>; if empty then <addressbook> <orgname>

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2016 – Domain extended to 500 characters

6.96 LEC_DESCR **PATSTAT Legal Status**

Name: Description of the legal event code.

Also Known As: n/a

Description: Short english text explaining the legal event code.

Domain: up to 250 ASCII characters

Default value: n/a

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.97 LEC_ID **PATSTAT Legal Status**

Name: Technical ID of a legal event code

Also Known As: n/a

Description: Surrogate key: Technical unique identifier without any business meaning.

Domain: Number 0 ... about 3 000;

0 indicates the dummy record / default record

Default value: 0

Source database:

- In table TLS221_INPADOC_PRS:

Derived from PATSTAT (tables TLS221_INPADOC_PRS, TLS201_APPLN and TLS802_LEGAL_EVENT_CODE)

- In table TLS802_LEGAL_EVENT_CODE:

Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.98 LEC_NAME **PATSTAT Legal Status**

Name: Legal event code of the IPR

Also Known As: n/a

Description: Part of the Alternate Key. To identify the type of a legal event, it has to be combined with the other 3 attributes of the Alternate Key.

Domain: up to 4 ASCII characters

Default value: n/a

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.99 LECG_DESCR **PATSTAT Legal Status**

Name: Description of the legal event code group.

Also Known As: n/a

Description: Short english text explaining the legal event code group.

Domain: up to 150 ASCII characters

Default value: n/a

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.100 LECG_ID **PATSTAT Legal Status**

Name: Technical ID of a legal event code group

Also Known As: n/a

Description: Surrogate key: Technical unique identifier without any business meaning.

Domain: Number 0 ... about 21

Default value: 0 (if no legal event code group has been assigned)

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.101 LECG_NAME **PATSTAT Legal Status**

Name: Legal event code group

Also Known As: n/a

Description: Short name of the legal event code group.

Domain: up to 6 ASCII characters

Default value: n/a

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.102 MIDDLE_NAME

Name: Middle name of a physical person

Also Known As: n/a

Description: Contains the middle name of a physical person

Domain: Up to 500 characters

Default value: empty string

Source database: USPTO data of published applications and published grants

Source field name:

<addressbook> <middle-name>

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2016 – Domain extended to 500 characters

6.103 NACE2_CODE

Name: 2-4 digit code of the Statistical Classification of Economic Activities in the European Community (Nomenclature statistique des activités économiques dans la CE)

Also Known As: n/a

Description: The 2-4 digits NACE2 code, like '17', '18.1', '20.51' or '20.60'.

Domain: Up to 5 ASCII characters;

It must not be defined as numerical field because trailing zeros are significant (e. g. NACE2 codes "20.6" and "20.60" are not the same).

Default value: n/a

Source database: See Eurostat's paper described in section 5.27 "TLS902_IPC_NACE2: Mapping between IPC and industrial sectors".

Source field name: n/a

Source sub-field identifier: n/a

Comments: This is a classification according to *industries*. A classification according to *technology* is the TECHN_FIELD_NR which can be found in the tables TLS901_TECHN_FIELD_IPC and TLS209_APPLN_IPC.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2015 - First version

6.104 NACE2_DESCR

Name: Description of the NACE2 code

Also Known As: n/a

Description: Short description of the NACE2 code

Domain: Up to 150 ASCII characters

Default value: n/a

Source database: See Eurostat's paper cited in section 5.27 "TLS902_IPC_NACE2: Mapping between IPC and industrial sectors".

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2015 - First version

6.105 NACE2_WEIGHT

Name: Indicator whether a NACE2 code will be assigned to a certain IPC

Also Known As: n/a

Description: Weight (number between 1 and 0) indicating whether there is a mapping between a particular IPC and a NACE2 code.

Domain: Number 0 or 1

Default value: 1

Source database: See Eurostat's paper cited in section 5.27 "TLS902_IPC_NACE2: Mapping between IPC and industrial sectors".

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2015 - First version

M. Kracker - 01-04-2016 – Changed domain and description

6.106 NAME_FREEFORM

Name: Full name in a single string

Also Known As: n/a

Description: Contains the full name in case the name is not available in structured form, where first, middle and last name are in different data fields.

Domain: Up to 500 characters

Default value: empty string

Source database: DOCDB and EP Register

For DOCDB data:

Source field name

```
<inventors>
  <inventor sequence="1" data-format="docdb">
    <inventor-name>
      <name>MACDONALD ALEX BRUCE</name>
    </inventor-name>
    <residence>
      <country>US</country>
    </residence>
  </inventor>
  <inventor sequence="2" data-format="docdb">
    <inventor-name>
      <name>AN LING LING</name>
    </inventor-name>
    <residence>
      <country>US</country>
    </residence>
  </inventor>
  <inventor sequence="1" data-format="docdba">
    <inventor-name>
      <name>WHITTUM- HUDSON, JUDITH A</name>
    </inventor-name>
  </inventor>
  <inventor sequence="2" data-format="docdba">
    <inventor-name>
      <name>MACDONALD, ALEX BRUCE, </name>
    </inventor-name>
  </inventor>
  <inventor sequence="3" data-format="docdba">
    <inventor-name>
      <name>AN, LING LING</name>
    </inventor-name>
  </inventor>
</inventors>

<applicants>
  <applicant sequence="1" data-format="docdba">
    <applicant-name>
      <name>THE JOHNS HOPKINS UNIVERSITY MACDONALD, ALEX BRUCE</name>
    </applicant-name>
  </applicant>
  <applicant sequence="2" data-format="docdba">
    <applicant-name>
      <name>AN, LING LING UNIVERSITY OF MASSACHUSETTS, A
PUBLIC INSTITUTION OF HIGHER EDUCATION OF THE COMMONWEALTH
OF MASSACHUSETTS, </name>
```

```

        </applicant-name>
    </applicant>
</applicants>

```

Source sub-field identifier

It occurs that DOCDB contains the names in DOCDB standardized format, but not in unstandardized format. So the following rule applies:

If **unstandardized applicants / inventors exist**, take format "docdba".

Otherwise take standardized format "docdb", if there is one:

- inventor-name ... data-format="docdba"
If no such entry exists, use inventor-name ... data-format="docdb"
- applicant-name ... data-format="docdba"
If no such entry exists, use applicant-name ... data-format="docdb"

For EP Register data:

Source field name

```

<parties>
  <applicants change-gazette-num="2000/29">
    <applicant app-type="applicant" designation="all" sequence="1">
      <addressbook>
        <name>Seidel, Helmut</name>
        <address>
          <address-1>Fliederstrasse 19</address-1>
          <address-2>65396 Walluf</address-2>
          <country>DE</country>
        </address>
      </addressbook>
      <nationality>
        <country/>
      </nationality>
      <residence>
        <country/>
      </residence>
    </applicant>
  </applicants>
  <inventors change-gazette-num="2000/29">
    <inventor sequence="01">
      <addressbook>
        <name>Franta, Georg</name>
        <address>
          <address-1>Ulrich-Rapp-Strasse 18</address-1>
          <address-2>87634 Obergünzburg</address-2>
          <country>DE</country>
        </address>
      </addressbook>
    </inventor>
    <inventor sequence="02">
      <addressbook>
        <name>Dojan, Viktor</name>
        <address>
          <address-1>Ludwig-Strecker-Strasse 5</address-1>
          <address-2>55129 Mainz</address-2>
          <country>DE</country>
        </address>
      </addressbook>
    </inventor>
  </inventors>

```

</inventors>

Source sub-field identifier: n/a

Comments:

DOCDB data in `data-format = "docdba"` are stored in PATSTAT "as received" by the EPO from other offices after converting to upper case and removing diacritics.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2014 – DOCDB data: Source sub-field identifier changed.

6.107 NAT_AUTH_CC **PATSTAT Legal Status**

Name: National authority country code.

Also Known As: n/a

Description: National authority after entry in the national phase of an international / regional application. Also in the special case of Hong Kong (HK) and Russian Federation (RU) for applications which have been originally filed in CN, GB or EP (in case of HK) or in SU (in case of RU).

It is part of the Alternate Key. To identify the type of a legal event, it has to be combined with the other 3 attributes of the Alternate Key.

Domain: 2 ASCII characters (according to WIPO standard ST.3) or empty

Default value: empty

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.108 NAT_LEC_NAME **PATSTAT Legal Status**

Name: National legal status code.

Also Known As: n/a

Description: National legal status code after entry in the national phase of an international / regional application. Also in the special case of Hong Kong (HK) and Russian Federation (RU) for applications which have been originally filed in CN, GB or EP (in case of HK) or in SU (in case of RU).

It is part of the Alternate Key. To identify the type of a legal event, it has to be combined with the other 3 attributes of the Alternate Key.

Domain: up to 4 ASCII characters

Default value: empty

Source database: Based on Excel table "Categorisation of recently used legal status codes", accessible on the EPO homepage

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.109 NAT_CLASS_SYMBOL

Name: National classification symbol

Also Known As: n/a

Description: Classification symbol according to a National classification scheme

Domain: Up to 15 characters

Default value: n/a

Source database: DOCDB

Source field name:

```
<classification-national>
  <text>002002500</text>
  <text>X002410000</text>
</classification-national>
```

If a string of symbols contains a comma, then split the string at the comma and create multiple entries.

Source sub-field identifier

n/a

Comments

These symbols are stored in PATSTAT against the APPLN_ID.

National classification is found in DOCDB mainly for AT, BR, CA, CH, DE, DK, GB and MX. US and JP national classification symbols are in tables TLS222_APPLN_JP_CLASS and TLS223_APPLN_DOCUS.

These national classification symbols are stored exactly as received by the EPO. No corrections are made.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 01-07-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 19-02-2010 - Modified comments

D. Lingua - 11-10-2011 - US and JP symbols have been moved to separate tables

6.110 NAT_PHASE

Name: Indicator whether the application is in the national phase

Also Known As: n/a

Description: Indicates that an application is in the national phase.

Domain: 1 ASCII character

Y	Yes
N	No
space	not known (In case of uncertain interpretations; used very little or not at all)

Default value: N

Source database: PATSTAT;

Source field name: Derived from table TLS201_APPLN

Y	if the application has APPLN_KIND <> W and APPLN_AUTH is a national office;
N	otherwise

Source sub-field identifier: n/a

Comments

For explanation and disclaimer see attribute INT_PHASE in section 6.57.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2016 - First version

6.111 NB_APPLICANTS

Name: Number of applicants of an application according to the most recent publication

Also Known As: n/a

Description: Number of applicants of an application according to the most recent publication

Domain: Number 0 ... about 250

Default value: n/a

Source database: PATSTAT

Source field name: Derived from table TLS207_PERS_APPLN

Source sub-field identifier: n/a

Comments: Only the latest known set of applicants is considered (e. g. from the latest publication)

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.112 NB_CITING_DOCDB_FAM

Name: Number of forward citations on family level

Also Known As: n/a

Description: Number of distinct DOCDB simple families citing at least one of the publications or applications of the DOCDB simple family of the current application (search report citations from TLS212_CITATION)

Domain: Number 0 .. about 3.000

Default value: n/a

Source database: PATSTAT

Source field name: Derived from table TLS228_DOCDB_FAM_CITN

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-04-2015 - Name of attribute changed for clarification (was: NB_CITATION)

6.113 NB_INVENTORS

Name: Number of inventors of an application according to the most recent publication

Also Known As: n/a

Description: Number of inventors of an application according to the most recent publication

Domain: Number 0 ... about 250

Default value: n/a

Source database: PATSTAT

Source field name: Derived from table TLS207_PERS_APPLN

Source sub-field identifier: n/a

Comments: Only the latest known set of inventors is considered (e. g. from the latest publication)

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.114 NOT_WITH_IPC

Name: IPC main group not co-occurring with IPC sub class

Also Known As: n/a

Description: Empty or first 8 characters of an IPC symbol according to [WIPO ST.8](#).

Domain: Up to 8 ASCII characters; Example: 'A61K 6'

Default value: empty

Source database: See Eurostat's paper described in section 5.27 "TLS902_IPC_NACE2: Mapping between IPC and industrial sectors".

Source field name: n/a

Source sub-field identifier: n/a

Comments: IPC main group which must not co-occur with the IPC in attribute IPC.
In the most cases this field is empty

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2015 - First version

6.115 NPL_BIBLIO

Name: Non-Patent Literature bibliography

Also Known As: n/a

Description: Bibliographic data of the Non-Patent Literature

Domain: Up to 3.000 characters

Default value: n/a

Source database: DOCDB

Source field name:

Each NPL citation leads to a record in the PATSTAT Non-Patent Literature table. Ignore any text which is empty (spaces), and in these cases take care with the calculation of the NPL_CITN_SEQ_NR and the CITN_ID.

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit num="1">
      <document-id>
        <country>DE</country>
        <doc-number>19723659</doc-number>
        <kind>A1</kind>
      </document-id>
    </patcit>
    <category>X</category>
  </citation>
  <citation srep-phase="SEA" sequence="2">
    <patcit num="2">
      <document-id>
        <country>WO</country>
        <doc-number>0137477</doc-number>
        <kind>A1</kind>
      </document-id>
    </patcit>
    <category>X</category>
  </citation>
  <citation srep-phase="SEA" sequence="3">
    <nplcit num="1">
      <text>LI GONG ET AL: "MULTICAST SECURITY AND ITS EXTENSION TO
A MOBILE ENVIRONMENT" WIRELESS NETWORKS, ACM, US, Bd. 1, Nr. 3, 1. Oktober 1995
(1995-10-01), Seiten 281-295, XP000538241 ISSN: 1022-0038</text>
    </nplcit>
    <corresponding-docs>
      <refno>000538241</refno>
    </corresponding-docs>
    <category>Y</category>
  </citation>
  <citation srep-phase="SEA" sequence="4">
    <nplcit num="2">
      <text>"Universal Mobile Telecommunications System (UMTS);3G
Security; Security Architecture (3g TS 33.102 version 3.1. Release 1999)" ETSI TS
133 102 V3.3.1, XX, XX, 1. Januar 2000 (2000-01-01), Seiten 1-60,
XP002167547</text>
    </nplcit>
    <corresponding-docs>
      <refno>002167547</refno>
    </corresponding-docs>
    <category>Y</category>
  </citation>
  <citation srep-phase="SEA" sequence="5">
    <nplcit num="3">
```



```

        <text>TANAKA S ET AL: "A key distribution and rekeying
framework with totally ordered multicast protocols" INTERNATIONAL CONFERENCE ON
INFORMATION NETWORKING (ICOIN'01), Februar 2001 (2001-02), Seiten 831-838,
XP010534359</text>
    </nplcit>
    <corresponding-docs>
        <refno>010534359</refno>
    </corresponding-docs>
    <category>A</category>
</citation>
</references-cited>

```

Source sub-field identifier

n/a

Comments

- For NPL citations including a reference to a patent document (e. g. Derwent abstracts, Patent Abstracts of Japan), see description of table TLS212_CITATION and the source field name in attribute description of CITED_PAT_PUBLN_ID.
- Corresponding documents:
Corresponding documents are documents which have the same or substantially the same technical content, so they are stored together NPL_BIBLIO attribute, separated only by the string “-&” .

When an examiner is citing a document during the search or examination phase, and the document which is being cited is NOT an original document but a database or an abstract made available by third party, for this document the examiner is mentioning the information provided by third party (name of database, acc. number etc.) as first part and the source or corresponding document (main article: journal name, volume, number, pages etc.) information as second part. And these two parts of information are separated by an ‘&’ sign.

Other typical examples of corresponding documents are:

- One document of the patent family is published in a non-EPO language before the priority date of the application, whereas a different member of the patent family is published in an EPO language after the priority date of the application to be examined. In this case, the earliest publication will be cited and the later publication will be mentioned as &-document.
- Abstracts of documents where the original document is not easily available or in a non-EPO language. If only the abstract of a document is cited, the original document will be cited as &-document.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 01-07-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-04-2015 – Added “Corresponding documents” to comment

6.116 NPL_CITN_SEQ_NR

Name: Sequence number of the NPL citation

Also Known As: n/a

Description: Number for an NPL (non-patent literature) citation in the series of NPL citations for one publication/origin combination

Domain: Number 0 ... about 1000

Default value: 0

Source database: Computed from PATSTAT. It is a sequential number for each NPL citation. The numbering starts with 1 for each origin of citations (CITN_ORIGIN). The NPL_CITN_SEQ_NR will be set to 0 when the citation is not a NPL citation, but a patent citation.

Source field name: n/a

Source sub-field identifier: n/a

Comments

The NPL_CITN_SEQ_NR attribute does **not** indicate the order of appearance of a NPL citations.

The sequence numbers start at 1 for each origin of the citations.

See also attributes PAT_CITN_SEQ_NR and CITN_ID.

References in the EP Search Report to the original WO search report publication are also included as NPL citations, because most likely the EP search report did not repeat the references which were cited in the original PCT search report. E.g. see references of EP1468879A1, which contains a NPL citation with the text "See also references of WO 03064220A1".

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 09-05-2005 - First version CITN_EXTRACT-NPL

R. Heijna - 21-11-2005 - CITN_EXTRACT-NPL deleted

R. Heijna - 21-11-2005 - First version NPL_CITN_SEQ_NR

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-12-2015 - Changed processing instructions and comments

6.117 NPL_PUBLN_ID

Name: Non-Patent Literature publication identification

Also Known As: n/a

Description: Surrogate key for Non-Patent Literature publications

Domain: Number 0 ... 999 999 999

There are 2 number ranges (see below)

Range 1: NPL with XP number (EPO internal accession number, including the dummy NPL_PUBLN_ID = 0):

0 ... 949 999 999

Range 2: NPL without XP number:

950 000 000 ... 999 999 999

Default value: 0

Source database: DOCDB, PATSTAT

Source field name:

1) NPL citations with XP number

```
<refno>000538241</refno>
```

If the `nplcit` reference in DOCDB XML contains a so-called "XP number" in the bibliographic data, the same number without its prefix "XP" is included in the field `<refno>`.

```
<citation cited-phase="SEA" srep-office="EP" sequence="3">
  <nplcit num="1" npl-type="a" extracted-xp="000538241">
    <text>LI GONG ET AL: "MULTICAST SECURITY AND ITS EXTENSION TO A
MOBILE ENVIRONMENT" WIRELESS NETWORKS, ACM, US, vol. 1, no. 3, 1 October 1995
(1995-10-01), pages 281-295, XP000538241 ISSN: 1022-0038</text>
  </nplcit>
  <category>X</category>
</citation>
```

In case of NPL referring to a patent document, the XP number is still given, but additionally there will be also a reference to the document-ID of the patent publication. This case should be treated the same as before, that is the XP number in `<refno>` is used to create the surrogate key, but the reference to the patent publication is kept as mentioned in the descriptions of table `TLS212_CITATION` and of attribute `CITED_PAT_PUBLN_ID` (see case c) in the table of the Business Rules in 5.11 "TLS212_CITATION: Citation".

Usage example showing a WPI abstract of a patent:

```
<citation cited-phase="SEA" srep-office="EP" sequence="3">
  <nplcit num="1" npl-type="a" extracted-xp="002594548">
    <text>DATABASE WPI Week 200214 Thomson Scientific, London, GB; AN
2002-099849 XP002594548 -& JP 2001 237065 A (TOPPAN PRINTING CO LTD) 31
August 2001 (2001-08-31)</text>
    <source-doc>
      <document-id doc-id="289417544">
        <country>JP</country>
        <doc-number>2001237065</doc-number>
        <kind>A</kind>
      </document-id>
    </source-doc>
  </nplcit>
  <category>Y</category>
</citation>
```

2) NPL citations without XP number

If there is no XP number given in DOCDB XML, a NPL_PUBLN_ID surrogate key has to be created. In this case the surrogate key will get a number starting from 950 000 001.

Comments

These numbers are not allocated sequentially. Only the "replenished" surrogate keys starting from 950 000 001 are allocated sequentially.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 01-07-2005 - First version

R. Heijna - 21-11-2005 - Name and definition adapted

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 13-04-2012 - PATSTAT uses <refno> as surrogate key

M. Kracker - 01-04-2014 – Clarification in source field name and comment

6.118 NPL_TYPE

Name: Type of the Non-Patent Literature

Also Known As: n/a

Description: Indicates the type of the Non-Patent Literature

Domain: 1 ASCII character:

For articles:

- b Book citation
- c Chemical abstracts citation
- i Biological abstract citation
- j Patent Abstracts of Japan citation
- s Serial / Journal / Periodical citation

For online citations:

- d Derwent citation
- e Database citation
- w World Wide Web / Internet search citation

For poor NPL citations (no rich NPL structure):

- a Abstract citation of no specific kind

For the dummy entry:

space dummy

Default value: "a"

Note: all NPL citations with poor data content (in contrast to "rich" NPL structure) will have NPL_TYPE = "a" .

Source database: [DOCDB](#)

Source field name:

```
<publn_nplcit>
  <nplcit num="1" npl-type="s" extracted-xp="055067747">
    <text>JINEK M. ET AL: "A PROGRAMMABLE DUAL-RNA-GUIDED DNA
ENDONUCLEASE IN ADAPTIVE BACTERIAL IMMUNITY (SUPPLEMENTARY MATERIAL)", SCIENCE,
vol. 337, no. 6096, 17 August 2012 (2012-08-17), XP055067747</text>
    <article>
      <author>
        <name>JINEK M. ET AL</name>
      </author>
      <atl>A PROGRAMMABLE DUAL-RNA-GUIDED DNA ENDONUCLEASE IN
ADAPTIVE BACTERIAL IMMUNITY (SUPPLEMENTARY MATERIAL)</atl>
      <serial>
        <sertitle>SCIENCE</sertitle>
        <pubdate>
          <sdate>20120817</sdate>
        </pubdate>
        <vid>337</vid>
        <ino>6096</ino>
      </serial>
    </article>
  </nplcit>
</publn_nplcit>
```

Source sub-field identifier: n/a

Comments: In case of data inconsistencies (diverging NPL types for the same NPL) in the source database, priority is given to the type which occurs most frequently in rich citations.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2016 - First version

6.119 NUTS

Name: NUTS region, region

Also Known As: Nomenclature of Territorial Units for Statistics

Description: This attribute contains the NUTS code as defined by Eurostat

Domain: 5 ASCII characters or empty

5 characters, when NUTS_LEVEL = 3: indicates a NUTS 3 level,

Example: "DE211" for Ingolstadt

empty: no region has been allocated

Default value: empty

Source database: made available by ECOOM (K.U. LEUVEN)

Source field name: n/a

Source sub-field identifier: n/a

Comments:

The regionalisation procedure is performed on all person addresses of EP applications available in the PATSTAT 2016 Spring Edition, and adopts the new NUTS 2013 codes for 33 countries (EU28 plus Iceland, Liechtenstein, Norway, Switzerland and Turkey).

See table TLS904_NUTS for a details about the NUTS codes.

The 2016 Autumn version contains some special codes which are not true NUTS codes and therefore not in the TLS904_NUTS reference table:

- Country codes + XXX:
Examples: UKXXX, BEXXX
Meaning: Unassigned
The address was never treated and therefore remains unassigned.
- Country codes + ZZZ:
Examples: UKZZZ, BEZZZ
Meaning: Unassignable
The address has been treated, but it was impossible to allocate. Reasons and examples are:
 - Deceased: the address field only contains the word "deceased".
 - Foreign: there is an inconsistency between the country code and the country in the address field.
 - Spelling errors and address errors: mistakes in postcodes, street-names, city-names and counties.
 - Missing NUTS code for postcodes and/or city-names in the source NUTS list.
- NUTS1 + ZZ or NUTS2 + Z:
Examples: UKIZZ (London), UKNOZ (Northern Ireland)
Meaning: NUTS codes on a lower aggregation than NUTS3 + "Z" or "ZZ"
Some regions or addresses cannot be allocated at the lowest aggregation level of NUTS3. We allocate these on the next lowest possible level of aggregation (NUTS2 or NUTS1) and we complete the code with "Z" or "ZZ" respectively.

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available some time later. Consequently, additions and

changes introduced by the current PATSTAT edition are not included in the current edition, but will be included in the next edition of PATSTAT.

The methodology for regionalisation is described in

<http://ec.europa.eu/eurostat/documents/3859598/5916785/KS-RA-11-008-EN.PDF>>

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2016 - First version

6.120 NUTS_LEVEL

Name: Level of NUTS region code

Also Known As: n/a

Description: Indicates the level of the regionalisation code in attribute NUTS (according to NUTS - Nomenclature of Territorial Units for Statistics)

Domain: 1 digit number with values 0, 1, 2, 3 or 9;

The value 9 indicates that no regionalisation was performed, and consequently attribute NUTS is empty.

Default value: 9

Source database: made available by ECOOM (K.U. LEUVEN)

Source field name: n/a

Source sub-field identifier: n/a

Comments:

See attribute NUTS in table TLS206_PERSON / TLS906_PERSON.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2016 - First version

6.121 NUTS3

Name: Level 3 of NUTS region code

Also Known As: n/a

Description: The code of the region according to NUTS level 3 (Nomenclature of Territorial Units for Statistics) , version 2013

Domain: 5 ASCII characters;.

Default value: n/a

Source database:

Data provided by Univ. Leuven (17.10.2017).

Source field name: n/a

Source sub-field identifier: n/a

Comments:

n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2016 - First version

6.122 NUTS3_NAME

Name: Name of the NUTS level 3 region code

Also Known As: n/a

Description: The name of the region according to NUTS level 3 (Nomenclature of Territorial Units for Statistics), version 2013

Domain: up to 250 characters

Default value: n/a

Source database:

Data provided by Univ. Leuven (17.10.2017).

Source field name: n/a

Source sub-field identifier: n/a

Comments:

n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2016 - First version

6.123 OECD_MEMBER

Name: Member of the Organisation for Economic Co-operation and Development

Also Known As: n/a

Description: Indicates whether this country is a member state of the OECD

Domain: 1 ASCII character: Y or space

Y If a country is member of the OECD.
space otherwise

Default value: n/a

Source database: <http://www.oecd.org/about/membersandpartners/>

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.124 PARENT_APPLN_ID

Name: Application identification of the earlier application

Also Known As: n/a

Description: Surrogate key of the application which was the basis for the continuation application

Domain: Number 1 ... 999 999 999

Only earlier applications for which a continuation is filed with the same authority (domestic). E.g. the country in the priority-claim is the same as the country in the application-reference. Clearly self-claimers are to be ignored. The **continuation** is published with an INID-code in the 60-series ([WIPO ST.9](#)) (plus inner priority, INID (23) as used by DE). The case to be taken into account is case # 6 from section 4.6 "Relation Types".

The PARENT_APPLN_ID is taken from the APPLN_ID allocated in PATSTAT for the earlier application. All of the applications must have been collected from the DOCDB before this logic can be used.

Default value: n/a

Source database: DOCDB, PATSTAT

Source field name

```
<priority-claim sequence="2" data-format="docdb" status="A">
  <country>US</country>
  <doc-number>9885602</doc-number>
  <kind>A</kind>
  <date>20020314</date>
  <priority-linkage-type>3</priority-linkage-type>
  <priority-active-indicator>N</priority-active-indicator>
</priority-claim>
<priority-claim sequence="2" data-format="epodoc">
  <doc-number>US20020098856</doc-number>
</priority-claim>
<priority-claim sequence="1" data-format="original">
  <doc-number>9885602</doc-number>
</priority-claim>
```

With `<country>US</country>` `<doc-number>9885602</doc-number>` `<kind>A</kind>` in DOCDB the corresponding application in PATSTAT is determined (via APPLN_AUTH, APPLN_NR and APPLN_KIND) and the value of APPLN_ID for this corresponding application is the PARENT_APPLN_ID

If there is no corresponding application in PATSTAT it should be created, see section 4.4 "Application replenishment".

Source sub-field identifier

n/a

Comments

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 03-05-2005 - First version

R. Heijna - 20-07-2005 - Source field definition improved
R. Heijna - 07-07-2005 - Value zero for the physical model
J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.125 PAT_CITN_SEQ_NR

Name: Sequence number of the patent citation

Also Known As: n/a

Description: Number for a patent citation in the series of patent citations for one publication/"origin of citation" combination

Domain: Number 0 ... about 1000

Default value: 0

Source database: Computed from PATSTAT. It is a sequential number for each patent citation, regardless whether the patent citation is referring to a patent publication (CITED_PAT_PUBLN_ID > 0) or patent application (CITED_APPLN_ID > 0). The numbering starts with 1 for each origin of citations (CITN_ORIGIN).

The PAT_CITN_SEQ_NR will be set to 0 when the citation is not a patent citation, but a NPL (non-patent literature) citation.

Source field name: n/a

Source sub-field identifier: n/a

Comments

The PAT_CITN_SEQ_NR attribute does **not** indicate the order of appearance of patent citations.

The sequence numbers start at 1 for each origin of the citations.

The sequence number identifies all patent citations, i.e. it does not distinguish between citations of publications (CITED_PAT_PUBLN_ID) and citations of applications (CITED_APPLN_ID).

See also attributes NPL_CITN_SEQ_NR and CITN_ID.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

R. Heijna - 01-07-2005 - Specifically for patent citations

R. Heijna - 15-07-2005 - For PL as well as NPL citations

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 08-05-2013 - Clarification in comment

M. Kracker - 01-12-2015 – Changed processing instructions and comments

6.126 PAT_PUBLN_ID

Name: Patent publication identification

Also Known As: n/a

Description: Surrogate key for patent publications

Domain: Number 0 ... 999 999 999

Range 1 - from 1 to 900 000 000 is used for publications of standard applications (= non-artificial applications)

Range 2 - from 900 000 001 to 999 999 999 for artificial publication references created in PATSTAT for those cited publications which do not themselves have a publication reference in DOCDB (see section 4.5 "Publication replenishment").

Default value: 0

Source database: DOCDB (for Range 1), PATSTAT (for Range 2)

Source field name

For Range 1:

```
<exchange-document country="EP" doc-number="0681755" kind="B1" doc-id="300943156" date-  
publ="19960904" family-id="21747543" is-representative="NO" date-of-last-  
exchange="20150206" date-added-docdb="19960831" originating-office="EP">
```

For Range 2: A unique number is generated for each unique combinations of the alternate key (PUBLN_AUTH, PUBLN_NR, PUBLN_KIND, PUBLN_DATE).

Source sub-field identifier: n/a

Comments

Note: For reasons of database consistency, there must be a dummy publication with a PAT_PUBLN_ID value of 0.

Within range 1 (1 to 900 000 000) this key will remain stable, i.e. it will not change between PATSTAT editions. For details see section 4.3.2 "Stable IDs".

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M.Kracker - 0104-2014 – Changed comment; PATPUBLN_ID need not be sequential

M.Kracker - 0104-2015 – IDs of non-artificial publications are now taken from DOCDB

6.127 PERSON_ADDRESS

Name: Person Address

Also Known As: Correspondence address

Description: All address elements of the person apart from the country. Example: street, city, postal code.

Domain: Up to 1.000 characters

Default value: empty

Source database: see attribute PERSON_NAME

For DOCDB data:

Source field name and Source sub-field identifier:

see attribute ADDRESS_FREEFORM

For EP Register data:

Source field name

The attributes ADDRESS_1, ADDRESS_2, ADDRESS_3, ADDRESS_4 and ADDRESS_5 of table TLS226_PERSON_ORIG are concatenated with a comma.

For USPTO data of published applications and published grants:

Source field name:

The attributes STREET, CITY, ZIP_CODE and STATE of table TLS226_PERSON_ORIG are concatenated.

Alternatively – if available- the attributes ADDRESS_1, ADDRESS_2 and ADDRESS_3 are concatenated.

Regardless of the source, each source data element of this attribute is cleaned:

- Leading and trailing spaces are removed
- whitespace characters (tabs, line feed, carriage return, ...) are replaced by a space
- multiple spaces are reduced to a single space

Comments

Address data in DOCDB is only available for a few authorities with scattered coverage: EP, IT, WO, CA, FI, AT and for older GB and IE documents. Therefore better quality address for EP and US patents is taken from other sources: The EPO address data is sourced from the EP Register. The USPTO address data is sourced from the USPTO publication files on USPTO's website.

In PATSTAT Online due to data privacy reasons, the PERSON_ADDRESS has been emptied for all persons who might be a natural person (e. g. all inventors, or where the SECTOR attribute contains "INDIVIDUAL" or "UNKNOWN" or is empty.)

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 01-12-2004 - First version

R. Heijna - 31-05-2005 - Applicants and Inventors integrated

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 01-10-2013 - For EP changed source to EP Register; Changed description of source

M. Kracker - 05-10-2014 – Comment updated

M. Kracker - 01-05-2015 – Comment updated

M. Kracker - 01-10-2015 – Section “Source field name” for EP Register data is amended

M. Kracker - 01-04-2016 – USPTO source may use ADDRESS_1 – ADDRESS_3

6.128 PERSON_CTRY_CODE

Name: Person country code

Also Known As:

Description: Country part of the correspondence address of the person or business

Domain: 2 characters (A-Z), according to [WIPO ST.3](#) or spaces

Exception: in case of bad data it may be any characters (e. g. "UK" is not ST.3 compliant, but should be "GB")

Default value: spaces

Source database: DOCDB

Source field name: see attribute PERSON_NAME

DOCDB data:

Source field name

```
<inventors>
  <inventor sequence="1" data-format="docdb">
    <inventor-name>
      <name>STACY N SMITH</name>
    </inventor-name>
    <residence>
      <country>US</country>
    </residence>
  </inventor>
  <inventor sequence="1" data-format="docdba">
    <inventor-name>
      <name>STACY N. SMITH</name>
    </inventor-name>
    <address>
      <text>305 Cottonwood Lane, NC 27540 Holly Springs,
UNITED STATES OF AMERICA (USA)</text>
    </address>
  </inventor>
  <inventor sequence="1" data-format="original">
    <inventor-name>
      <name>Stacy N. Smith</name>
    </inventor-name>
  </inventor>
</inventors>

<applicants>
  <applicant sequence="1" data-format="docdb">
    <applicant-name>
      <name>ERICSSON INC</name>
    </applicant-name>
    <residence>
      <country>US</country>
    </residence>
  </applicant>
  <applicant sequence="1" data-format="docdba">
    <applicant-name>
      <name>ERICSSON INC.</name>
    </applicant-name>
    <address>
      <text>7001 Development Drive, 27709-3969 Research
Triangle Park, UNITED STATES OF AMERICA (USA)</text>
    </address>
  </applicant>
</applicants>
```

```

    <applicant sequence="1" data-format="original">
      <applicant-name>
        <name>Ericsson Inc.</name>
      </applicant-name>
    </applicant>
  </applicants>

```

Source sub-field identifier

data-format="docdb"

For EP Register data:

Source field name

```

<parties>
  <applicants change-gazette-num="2000/29">
    <applicant app-type="applicant" designation="all" sequence="1">
      <addressbook>
        <name>Seidel, Helmut</name>
        <address>
          <address-1>Fliederstrasse 19</address-1>
          <address-2>65396 Walluf</address-2>
          <country>DE</country>
        </address>
      </addressbook>
      <nationality>
        <country/>
      </nationality>
      <residence>
        <country/>
      </residence>
    </applicant>
  </applicants>
  <inventors change-gazette-num="2000/29">
    <inventor sequence="01">
      <addressbook>
        <name>Franta, Georg</name>
        <address>
          <address-1>Ulrich-Rapp-Strasse 18</address-1>
          <address-2>87634 Obergünzburg</address-2>
          <country>DE</country>
        </address>
      </addressbook>
    </inventor>
    <inventor sequence="02">
      <addressbook>
        <name>Dojan, Viktor</name>
        <address>
          <address-1>Ludwig-Strecker-Strasse 5</address-1>
          <address-2>55129 Mainz</address-2>
          <country>DE</country>
        </address>
      </addressbook>
    </inventor>
  </inventors>

```

Source sub-field identifier: n/a

For USPTO data of published applications and published grants:

Source field name:

```

<applicant sequence="00" app-type="applicant-inventor" designation="us-only">
  <addressbook>

```

```
<last-name>Eckhoff</last-name>
<first-name>Philip A.</first-name>
<address>
  <city>Bellevue</city>
  <state>WA</state>
  <country>US</country>
</address>
</addressbook>
<nationality>
  <country>omitted</country>
</nationality>
<residence>
  <country>US</country>
</residence>
</applicant>
```

Comments

The country code as well as the full name of the country are usually indicated as "the country". Note: For only 50% of the inventors the country codes are known.

Note that the EPO does not receive the Country Code value with the Japanese data which is loaded into DOCDB; for this reason there are no PERSON_CTRY_CODES in PATSTAT for Japanese documents.

This code is copied from the 'standard' DOCDB table and added to the 'bypass' data, matching on the application id of authority, number and kind code and inventor sequence number or applicant sequence number.

Country code does not necessarily indicate the "Nationality" of inventor or applicant.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 20-12-2004 - First version

R. Heijna - 31-05-2005 - Applicants and Inventors integrated

D. Lingua - 13-02-2008 - Comment and other fields modified

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 15-05-2013 - Added exception to Domain; For EP changes source to EP Register; Changed description of source

6.129 PERSON_ID

Name: Person identification

Also Known As: n/a

Description: Surrogate key based on the elements in the alternate primary key of table TLS206_PERSON

Domain: Number 1 ... 999 999 999

Default value: n/a

Source database: PATSTAT

Source field name

PERSON_NAME and PERSON_ADDRESS and PERSON_CTRY_CODE in PATSTAT.

Allocate a surrogate key PERSON_ID for each combination of these fields. Upper case and lower case are considered equal. E.g. "James Bond" is considered to be the same person name as "JAMES BOND".

Source sub-field identifier: n/a

Comments

Sequential number unique for each unique combination of the elements in the candidate primary key.

Persons are the legal or physical persons that have a relation with the patent granting procedure. Currently included are applicants and inventors.

This key will normally remain stable, i.e. it will not change between PATSTAT editions. However, in exceptional cases some values of PERSON_ID might change. For details see section 4.3.2 "Stable IDs".

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 15-04-2005 - First version

R. Heijna - 31-05-2005 - Applicant integrated with Inventor

M. Kracker - 01-10-2013 - Clarification in description

M. Kracker - 15-10-2015 - Clarification in comment

6.130 PERSON_NAME

Name: Person name

Also Known As: n/a

Description: Name of the Applicant or Inventor

Domain: Up to 500 characters

Default value: empty

Source database:

1) EPO Register for EP patent applications

2) OECD patents database for US data post 1976-01-01 up to and including November 15th 2005 for Published Grants.

3) PATSTAT weekly file extracts from USPTO website for Published Grants from November 22nd 2005 until today; Published Applications from September 29th 2005 to today inclusive.

4) Inventor & Applicant names for USPTO Published Applications from March 1st 2001 to September 22nd 2005 from DOCDB, data-format="docdba".

5) all other names from DOCDB, data-format="docdba".

For DOCDB data:

Source field name and Source sub-field identifier: see attribute NAME_FREEFORM

For EP Register data:

Source field name and Source sub-field identifier: see attribute NAME_FREEFORM

For USPTO data of published applications and published grants:

Source field name:

The attributes LAST_NAME, FIRST_NAME and MIDDLE_NAME of table TLS226_PERSON_ORIG are concatenated. The results, depending on the availability of the data, are like

- Kennedy, John F
- Kennedy, John
- Kennedy

Regardless of the source, each source data element of this attribute is cleaned:

- Leading and trailing spaces are removed
- whitespace characters (tabs, line feed, carriage return, ...) are replaced by a space
- multiple spaces are reduced to a single space

Comments

See also Business Rules in section 5.6 "TLS206_PERSON / TLS906_PERSON: Person".

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 01-12-2004 - First version

R. Heijna - 15-04-2005 - Size, source, comments updated

R. Heijna - 31-05-2005 - Applicants and Inventors integrated
J. Rollinson - 18-04-2006 - Comments extended, source field and domain updated
D. Lingua - 15-10-2008 - Comments extended
J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML
M. Kracker - 01-10-2013 - Source fields for EP Register added; Comments updated
M. Kracker - 15-10-2014 - Comments updated
M. Kracker - 01-04-2016 – Domain extended to 500 characters

6.131 PERSON_ORIG_ID

Name: Key for the unmodified person data record

Also Known As: n/a

Description: Number which uniquely identifies a row in the TLS226_PERSON_ORIG table

Domain: Number 1 ... 999 999 999

Default value: n/a

Source database: PATSTAT

Source field name: n/a

Source sub-field identifier: n/a

Comments: This key will remain stable, i.e. it will not change between PATSTAT editions. However, in exceptional cases some values of PERSON_ORIG_ID might change. For details see section 4.3.2 "Stable IDs".

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.132 PRIOR_APPLN_ID

Name: Application identification of claimed Paris Convention priority

Also Known As: n/a

Description: Surrogate key of an application of which the priority is claimed under the Paris convention

Domain: Number, 1 ... 999 999 999

Only "pure" priorities i.e. those according to the Paris convention and published with an INID-code in the 30-series ([WIPO ST.9](#)). The case to be taken into account is case # 2 from section 4.6 "Relation Types".

Default value: n/a

Source database: DOCDB, PATSTAT

Source field name

```
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>DE</country>
      <doc-number>10331291</doc-number>
      <kind>A</kind>
      <date>20030710</date>
    </document-id>
    <priority-active-indicator>Y</priority-active-indicator>
  </priority-claim>
```

With

```
<country>DE</country>
<doc-number>10331291</doc-number>
<kind>A</kind>
```

in DOCDB the corresponding priority application in PATSTAT is determined (via APPLN_AUTH, APPLN_NR and APPLN_KIND) and the value of APPLN_ID of this priority application will be assigned to PRIOR_APPLN_ID.

If there is no corresponding application in PATSTAT it should be created, see section 4.4 "Application replenishment".

Source sub-field identifier

n/a

Comments

n/a

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 22-04-2005 - First version

R. Heijna - 20-07-2005 - Source field definition improved

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.133 PRIOR_APPLN_SEQ_NR

Name: Sequence number of claimed priority

Also Known As: n/a

Description: Number indicating the place in the list of priorities claimed in the application.

Domain: Number, 1... about 500

Default value: n/a

Source database: DOCDB

Source field name:

```
<...>
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>DE</country>
      <doc-number>10331291</doc-number>
      <kind>A</kind>
      <date>20030710</date>
    </document-id>
    <priority-active-indicator>Y</priority-active-indicator>
  </priority-claim>
</priority-claims>
```

If an application is claiming itself as a priority, then this priority is not stored in PATSTAT. So if a priority-claim element is the same as the application-reference, the application is claiming itself as a priority. These are normally the last priority in the priority-claims list of DOCDB.

This means that the sequence numbers of any subsequent priorities claimed by this application must be reduced by 1. See the rules for PRIOR_APPLN_ID to see which priorities are to be ignored.

Source sub-field identifier

n/a

Comments

The sequence number is assigned based on the sequence in which the priorities have been provided by the supplier.

For US data - where sequence is extremely important with continuations/divisions/continuations in part - the sequence numbers is on filing-date descending. Earliest filing date last.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 22-12-2004 - First version

D. Lingua - 05-06-2009 - Added comments

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.134 PRS_CODE **PATSTAT Legal Status**

Name: PRS Code

Also Known As: n/a

Description: The code which classifies the legal status event in conjunction with the country code of the application.

Domain: Up to 4 ASCII characters

Default value: n/a

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name:

Take the value of tag L008EP "PRS Code" from PRS

```
<IPREVENT RECORD=DATA DATE=20080424 CY=AT STATUS=N>
  <L001EP>AT</L001EP>
  <L002EP>F</L002EP>
  <L003EP>6195</L003EP>
  <L004EP>A </L004EP>
  <L005EP>PI</L005EP>
  <L007EP>20080415</L007EP>
  <L008EP>ELJ </L008EP>
  <L013EP>N</L013EP>
  <L014EP>19950117</L014EP>
  <L015EP>AT1995      61A</L015EP>
  <L016EP>AT      406287B</L016EP>
  <L017EP> AT      406287B</L017EP>
  <L019EP>20080421</L019EP>
</IPREVENT>
```

Source sub-field identifier: n/a

Comments

The worldwide legal status database (formerly known as INPADOC PRS Patent Register Service) uses thousands of codes to classify legal events in the lives of industrial property rights. Always check which national patent law is used with the PRS code.

For example (see elements <L001EP> and <L008EP> in the data example above): "AT ELJ" means "Ceased due to non-payment of the annual renewal fee in Austria".

A full list of PRS codes is given in the documents "Legal status codes in original language" and "Legal status codes in English" on <http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/weekly.html> .

Explanation of *regional* legal event codes:

There are 4 regional codes: REG, PGFP, PG25, and PGRI. These codes are used for applications where some event took place in the *national phase* of these applications.

3 of these 4 regional event codes indicate a *specific* legal event of an EP patent. The national office where this event takes place is indicated in attribute L501EP.

- PGFP Post grant: Annual fees paid to the national office
- PG25 Lapsed in a contracting state announced via post grant information from national office to EPO
- PGRI Post grant: Patent reinstated in contracting state

One specific regional code, which is REG, is of a more general nature. It indicates that *some* legal event took place in a national phase. The REG code is not limited to EP applications, but is applied to these applications and their national phases:

- National phase of an EP application
- National phase of an PCT application
- RU (Russian) phase of an earlier SU (Soviet Union) application
- HK (Hong Kong) phase of an GB (United Kingdom) application
- HK (Hong Kong) phase of an CN (Chinese) application

The national office and the code of the legal event which took place in the national phase are stored in the attributes L501EP and L502EP.

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 27-07-2010 - First version

M. Kracker - 01-10-2013 - Update of description and comment

M. Kracker - 01 - 11-2013 – Comment added

6.135 PRS_EVENT_SEQ_NR **PATSTAT Legal Status**

Name: Sequence number of the legal status event

Also Known As: n/a

Description: Number indicating the sequence of the PRS event in the series of events for one patent application

Domain: Number 1 ... about 2.000; max value in April 2013 was 1.055

Default value: n/a

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name: n/a

Source sub-field identifier: n/a

Comments

For a given patent application, each legal event PRS code is assigned a sequence number so as to make each row identifiable by the combination APPLN_ID and PRS_EVENT_SEQ_NR.

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 27-07-2010 - First version

6.136 PRS_GAZETTE_DATE **PATSTAT Legal Status**

Name: Publication date of the legal event

Also Known As: n/a

Description: Date on which the PRS event has been made available to the public

Domain: Date

Default value: n/a

Source database: INPADOC Worldwide Legal Status database - PRS

Source field name:

Take the value of tag L007EP "Gazette date" from PRS

```
<IPREVENT RECORD=DATA DATE=20080424 CY=AT STATUS=N>
  <L001EP>AT</L001EP>
  <L002EP>F</L002EP>
  <L003EP>6195</L003EP>
  <L004EP>A </L004EP>
  <L005EP>PI</L005EP>
  <L007EP>20080415</L007EP>
  <L008EP>ELJ </L008EP>
  <L013EP>N</L013EP>
  <L014EP>19950117</L014EP>
  <L015EP>AT1995          61A</L015EP>
  <L016EP>AT      406287B</L016EP>
  <L017EP> AT      406287B</L017EP>
  <L019EP>20080421</L019EP>
</IPREVENT>
```

Source sub-field identifier: n/a

Comments

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 27-07-2010 - First version

6.137 PSN_ID

Name: ID for the PATSTAT standardized name

Also Known As: n/a

Description: For each unique occurrence of the PSN_NAME an ID will be assigned. This will facilitate faster grouping.

Multiple rows may have the same PDN_ID, if multiple person names in table TLS206_PERSON have been harmonized into one PATSTAT standardized name.

Domain: Number 1 ... 999 999 999

Default value: n/a

Source database: made available by ECOOM (K.U. LEUVEN)

Source field name: n/a

Source sub-field identifier: n/a

Comments:

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available 3-4 months afterwards. Consequently, additions and changes introduced by the current PATSTAT edition are not harmonized in the current edition, but will be harmonized in the next edition of PATSTAT.

See also comment of attribute PSN_NAME.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2014 – Comment updated

M. Kracker - 01-04-2015 – Comment updated

M. Kracker - 01-04-2016 – Attribute renamed (was: HRM_L2_ID); Comment updated

6.138 PSN_LEVEL

Name: Harmonization level of PATSTAT standardized name

Also Known As: n/a

Description: This attribute indicates for each name in PSN_NAME the level of harmonisation which has been applied

Domain: Number 0 ... 2

- 0: No harmonization has taken place
(PSN_NAME is the same as attribute PERSON_NAME)
- 1: Automated harmonization only has been applied
- 2: Automated harmonization plus manual refinement have been applied.

Default value: n/a

Source database: made available by ECOOM (K.U. LEUVEN)

Source field name: n/a

Source sub-field identifier: n/a

Comments:

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available 3-4 months afterwards. Consequently, additions and changes introduced by the current PATSTAT edition are not harmonized in the current edition, but will be harmonized in the next edition of PATSTAT.

See also comment of attribute PSN_NAME.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2014 – Comment updated

M. Kracker - 01-04-2016 – Attribute renamed (was: HRM_LEVEL); Comment updated

6.139 PSN_NAME

Name: PATSTAT standardized name

Also Known As: n/a

Description: PATSTAT standardized name.

The attribute is populated for all persons. Names of persons which have not been harmonized are just copied from the attribute PERSON_NAME.

Domain: Up to 500 characters

Default value: n/a

Source database: made available by ECOOM (K.U. LEUVEN)

Source field name: n/a

Source sub-field identifier: n/a

Comments:

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available 3-4 months afterwards. Consequently, additions and changes introduced by the current PATSTAT edition are not harmonized in the current edition, but will be harmonized in the next edition of PATSTAT.

The PATSTAT standardized names are the results of an approach to standardize the original name (<http://www.ecoom.be/en/EEE-PPAT>). It is done in an automated way with additional manual refinements.

Background papers on the production of the PSN_NAME:

- i. Du Plessis, M., Van Looy, B., Song, X & Magerman, T. (2009) Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation. EUROSTAT Working Paper and Studies, Luxembourg.
- ii. Magerman T, Grouwels J., Song X. & Van Looy B. (2009). Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization. EUROSTAT Working Paper and Studies, Luxembourg.
- iii. Peeters B., Song X., Callaert J., Grouwels J., Van Looy B. (2009). Harmonizing harmonized patentee names: an exploratory assessment of top patentees. EUROSTAT Working paper and Studies, Luxembourg.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2014 – Updated description and comment

M. Kracker - 01-04-2016 – Attribute renamed (was: HRM_L2);
Domain extended to 500 characters; Comment updated

6.140 PSN_SECTOR

Name: Sector of the applicant

Also Known As: n/a

Description: This is a by-product of the PATSTAT standardized name harmonization effort: Applicants may have been assigned to one or more sectors, like company, government or non-profit organization, university or hospital. If the sector of an applicant cannot be determined, then the sector is UNKNOWN. If a person (e.g. a person who is only an inventor, but not an applicant) is not assigned a sector, then this field is empty.

So this column may contain zero, one or more of these keywords:

INDIVIDUAL
COMPANY
UNKNOWN
GOV
NON-PROFIT
UNIVERSITY
HOSPITAL

This list of keywords may change.

Domain: Up to 50 ASCII characters

Default value: empty

Source database: made available by ECOOM (K.U. LEUVEN)

Source field name: n/a

Source sub-field identifier: n/a

Comments:

The processing of the PATSTAT standardized name starts as soon as PATSTAT data is released. Typically the result will be available 3-4 months afterwards. Consequently, additions and changes introduced by the current PATSTAT edition are not harmonized in the current edition, but will be harmonized in the next edition of PATSTAT.

See also comment of attribute PSN_NAME.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 15-10-2014 – Comment updated

M. Kracker - 01-04-2016 – Attribute renamed (was: SECTOR); Comment updated

6.141 PUBLN_AUTH

Name: Publication Authority

Also Known As: Publishing office

Description: Patent Authority that issued the publication of the application

Domain: 2 characters (A-Z) according to [WIPO ST.3](#) or spaces

Exception: in case of bad data it may be any characters

Default value: spaces

Source database: DOCDB

Source field name:

1) Standard publication reference (PAT_PUBLN_ID between 0 and 900 000 000):

```
<publication-reference data-format="docdb">
  <document-id lang="en">
    <country>US</country>
    <doc-number>2007011914</doc-number>
    <kind>A1</kind>
    <date>20070118</date>
  </document-id>
</publication-reference>
```

2) Artificial publications from cited references (PAT_PUBLN_ID between 900 00 001 and 999 999 999):

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>US</country>
        <doc-number>3380531</doc-number>
        <kind>A</kind>
      </document-id>
    </patcit>
    <category>A</category>
  </citation>
```

Source sub-field identifier

data-format="docdb"

Comments

Take all the publication-references in DOCDB into PATSTAT.

For all artificial publications which were themselves artificially created for those cited publications, where the cited publications are not registered in DOCDB as publications: use the authority (country) of the cited publication: <country>US</country>

Check if the cited publication has a publication-reference in DOCDB and if not, then create an artificial publication and an artificial application.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

M. Kracker - 15-05-2013 - Added exception to Domain

6.142 PUBLN_CLAIMS

Name: Indicator of the number of claims in the given publication

Also Known As: n/a

Description: This indicator provides the number of claims that has been attributed to the specific publication. Currently only available for certain EP and US publications.

Domain: Number 0 ... about 1.000;

0 means

- that a publication contains no claims (e. g. EP publications of kind A3, A8, B8 and non-republished EuroPCT applications)
or
- that the number of claims is not known (e. g. US-publications published on or before 1974; publications not from EP or US)

Default value: 0

Source database: Special delivery files for EP and US data only

Source field name:

There are two separate sources for US and EP data. Specific formats are used in each case.

1) US data: relates to granted patents only (A documents until 2000, B1 or B2 documents afterwards) which were published on or after 1975-01-01

The backfile published by the USPTO has this format:

- Columns 1-7: US Patent document
I.e. issued patent to which the other information in the record applies (character field)
- Columns 9-12: Number of Claims
If information for this field is missing the field is marked with a period (".") (integer field)
- Columns 14-17: Number of Drawing Figures
If information for this field is missing, the field is marked with a period (".") (integer field)
- Columns 19-22: Number of Submitted Drawing Sheets
If information for this field is missing, the field is marked with a period (".") (integer field)

Usage example:

7585234	24	4	3
7585235	8	18	7
7585236	42	23	13

Only the information in columns 1-7 and 9-12 is used, the remaining is ignored.

The information in column 1-7 is mapped to publication data in PATSTAT, where

PUBLN_AUTH is US

PUBLN_NR is the content of column 1-7

PUBLN_KIND is "A" until 2000, "B1" or "B2" from 2001

The content of column 9-12 is used to populate element PUBLN_CLAIMS.

2) EP data: relates to both published applications (kind code "A") from 1978 and granted patents (kind code "B") from 1980.

Data was extracted from the EPO publications XML, in this format:

```
EP publication number;kind code;publication date;number of claims  
1123811;A2;20010816;17  
1124248;A2;20010816;20  
1123812;A2;20010816;34
```

The information needs to be mapped to publication data in PATSTAT as follows:

PUBLN_AUTH is EP

PUBLN_NR is the content of "EP publication number";

PUBLN_KIND is the content of "kind code";

The content of "number of claims" is used to populate element PUBLN_CLAIMS.

Source sub-field identifier: n/a

Comments

In a minority of cases for EP B (European granted patents) publications, multiple sets of claims are published, each set applying to a specific group of designated states. For the sake of simplification, only the highest number of claims has been considered.

Warning: The number of claims will be "0" for all EP A documents originating from a PCT published in English, French or German (so called "Euro-PCTs"). For all these Euro-PCT documents, as the EPO does not republish the application (by recognising the PCT publication as being sufficient), the claim count for the EP document will be equal to "0" as there is no real EP A publication. For those Euro-PCT documents whose original PCT language is not English, French or German, there is a new publication in one EPO official language and thus the claim count is available.

Modification history

Author of update - Date of update - Explanation of update

D. Lingua - 04-08-2011 - First version

D. Lingua - 13-04-2012 - Update on EP B documents (1980 to 2005) and warning

D. Lingua - 25-09-2012 - Update on US claim coverage

M. Kracker - 01-10-2013 - Clarification for value 0

M. Kracker - 15-10-2015 - Clarification for coverage (US starting from 1975-01-01; EP)

6.143 PUBLN_DATE

Name: Publication date

Also Known As: n/a

Description: Date on which the publication was made available to the public

Domain: Date

Default value: 9999-12-31

Source database: DOCDB

Source field name

```
<publication-reference data-format="docdb">
  <document-id lang="en">
    <country>US</country>
    <doc-number>2007011914</doc-number>
    <kind>A1</kind>
    <date>20070118</date>
  </document-id>
</publication-reference>
```

With `country`, `doc-number` and `kind` in `document-id` in `patcit` in `citation` in `references-cited` in `DOCDB` the corresponding publication in `PAT_PUBLN` in `PATSTAT` is determined (via `PUBLN_AUTH`, `PUBLN_NR` and `PUBLN_KIND`). The value of `PUBLN_DATE` for this publication is the value of `date` in `document-id`. If it is an invalid date or empty, then use 9999-12-31.

Source sub-field identifier

`data-format="docdb"`

Comments

n/a

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.144 PUBLN_FIRST_GRANT

Name: Identifier of the granting publication of an application

Also Known As: n/a

Description: Indication that the publication can be considered as the first publication of grant of a given application

Domain: Number, 0 ... 1:

0 - this publication step is **not** the first publication of a grant.

1 - this publication step is the first publication of a grant.

Default value: n/a

Source database: DOCDB

Source field name: in tag <date-of-public-availability> when categories <printed-with-grant> or <not-printed-with-grant> are given

Case 1) <printed-with-grant>

```
<exchange-document country="ES" doc-number="2340887" kind="T3" date-
publ="20100610" family-id="38220640" is-representative="YES" date-of-last-
exchange="20100610" date-added-docdb="20100601" originating-office="EP"
status="C">
```

```
  <bibliographic-data>
    <publication-reference data-format="docdb">
      <document-id lang="es">
        <country>ES</country>
        <doc-number>2340887</doc-number>
        <kind>T3</kind>
        <date>20100610</date>
      </document-id>
    </publication-reference>
    <publication-reference data-format="epodoc">
      <document-id lang="es">
        <doc-number>ES2340887T</doc-number>
      </document-id>
    </publication-reference>
    ...
    <dates-of-public-availability>
      <printed-with-grant>
        <document-id>
          <date>20100610</date>
        </document-id>
      </printed-with-grant>
    </dates-of-public-availability>
  </bibliographic-data>
```

Case 2) <not-printed-with-grant>

```
<exchange-document country="HK" doc-number="21891" kind="A" date-publ="19910328"
family-id="26321336" is-representative="YES" date-of-last-exchange="20100610"
date-of-previous-exchange="20001030" date-added-docdb="20000401" originating-
office="EP" status="A">
```

```
  <bibliographic-data>
    <publication-reference data-format="docdb">
      <document-id>
        <country>HK</country>
        <doc-number>21891</doc-number>
        <kind>A</kind>
        <date>19910328</date>
      </document-id>
    </publication-reference>
  </bibliographic-data>
```



```

        </document-id>
    </publication-reference>
    <publication-reference data-format="epodoc">
        <document-id>
            <doc-number>HK21891</doc-number>
        </document-id>
    </publication-reference>
...
    <dates-of-public-availability>
        <not-printed-with-grant>
            <document-id>
                <date>19910328</date>
            </document-id>
        </not-printed-with-grant>
    </dates-of-public-availability>
</bibliographic-data>

```

Source sub-field identifier: n/a

Comments

This indicator provides a somewhat simplistic view to identify the date of first publication of grant. It is based on the DOCDB XML element `<date-of-public-availability>` and will have a value "0" if this tag does not contain any of the two categories listed above. It will have the value "1" if the tag contains one of the two categories listed above.

A value "0" is also given in case the element `<date-of-public-availability>` is not present.

In case there are multiple publications of a grant, the first publication only is given the "1" indicator (first publication of grant).

Annex IV to the [DOCDB ST36 Layout Description](http://www.wipo.int/export/sites/www/standards/en/pdf/03-30-01.pdf) lists a concordance of this element with the WIPO ST.30 (<http://www.wipo.int/export/sites/www/standards/en/pdf/03-30-01.pdf>); codes "450" and "470" are the key to identify the grant date.

Although the EPO has taken great care in analysing the grant information, this process is the result of interpretations and assumptions for which no responsibility whatsoever can be accepted.

If an application has a publication with PUBLN_FIRST_GRANT = 1 then it can be concluded that the application has been granted. In all other cases, the application has not (yet) been granted.

The publication date can be considered as being (close to) the granting date. However, exception exists, like the Austrian Utility models which are granted 2 month before being published as U1 publication.

Modification history

Author of update Date of update - Explanation of update

D. Lingua - 23-02-2009 - First version

D. Lingua - 14-06-2010 - Changed source to DOCDB XML

M. Kracker - 01-04-2014 – Comment extended

6.145 PUBLN_KIND

Name: Kind of Publication

Also Known As: n/a

Description: Publication kind attributed by the Patent Authority issuing the publication

Domain: Up to 2 ASCII characters, as laid down in the "Kind Code concordance list" for databases within the EPO in column "DOCDB" available on <http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/regular.html> . See also "Comments" below.

Default value: n/a

Source database: DOCDB

Source field name

1) Standard publication reference (PAT_PUBLN_ID between 0 and 900 000 000):

```
<publication-reference data-format="docdb">
  <document-id lang="en">
    <country>US</country>
    <doc-number>2007011914</doc-number>
    <kind>A1</kind>
    <date>20070118</date>
  </document-id>
</publication-reference>
```

2) Artificial publications from cited references (PAT_PUBLN_ID between 900 00 001 and 999 999 999):

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>US</country>
        <doc-number>3380531</doc-number>
        <kind>A</kind>
      </document-id>
    </patcit>
    <category>A</category>
  </citation>
```

Source sub-field identifier

data-format="docdb"

Comments

Take all the publication-references in DOCDB into PATSTAT.

For all artificial publications which were themselves artificially created for those cited publications, where the cited publications are not registered in DOCDB as publications: use the authority (country) of the cited publication and the publication kind code as cited (by the applicant). Because of this, a substantial number of PUBLN_KIND codes will not occur in the DOCDB "Kind Code concordance list". Example: the data base contains more than 2 000 US publications with PUBLN_KIND code B. We assume this might be B1, B2, B3, ... but we have kept the kind code "B" as originally cited.

Check if the cited publication has a publication-reference in DOCDB and if not, then create an artificial publication and an artificial application.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

D. Lingua - 11-10-2011 - Updated figures to October 2011 edition

6.146 PUBLN_LG

Name: Publication language

Also Known As: n/a

Description: Language of the publication

Domain: 2 ASCII characters, according to ISO language codes (ISO 639-1) or spaces

Default value: spaces

Source database: DOCDB

Source field name:

```
<publication-reference data-format="docdb">
  <document-id lang="en">
    <country>US</country>
    <doc-number>2007011914</doc-number>
    <kind>A1</kind>
    <date>20070118</date>
  </document-id>
</publication-reference>
```

If country = 'DE', then PUBLN_LG = 'DE'.

Source sub-field identifier

data-format="docdb"

Comments

Present in about 10% of cases only (NB not always necessary, e.g. DE publications are always in German)

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.147 PUBLN_NR

Name: Publication number

Also Known As: n/a

Description: Number given by the Patent Authority issuing the publication

Domain: Up to 15 ASCII characters (since April 2013 without leading spaces)

Default value: n/a

Source database: DOCDB

Source field name

1) standard publication reference:

```
<publication-reference data-format="docdb">
  <document-id lang="en">
    <country>US</country>
    <doc-number>2007011914</doc-number>
    <kind>A1</kind>
    <date>20070118</date>
  </document-id>
</publication-reference>
```

2) Publications from cited references:

```
<references-cited>
  <citation srep-phase="SEA" sequence="1">
    <patcit>
      <document-id>
        <country>US</country>
        <doc-number>3380531</doc-number>
        <kind>A</kind>
      </document-id>
    </patcit>
    <category>A</category>
  </citation>
```

Source sub-field identifier

n/a

Comments

The term "Publication number" is sometimes used for the combination with authority as well, like "EP1123456"

Take all the publication-references in DOCDB into PATSTAT.

For all artificial publications which were themselves artificially created for those cited publications, where the cited publications are not registered in DOCDB as publications, use the authority (country) of the cited publication: <doc-number>3380531</doc-number>.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 04-05-2005 - First version

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.148 PUBLN_NR_ORIGINAL

Name: Publication number in original format

Also Known As: Original publication number

Description: Publication number in original format as provided by the supplier. It is assumed that the number is as printed on the publication.

The availability and the format of the original publication number depends on the publishing authority.

Domain: Up to 100 characters

Default value: empty

Source database: DOCDB

Source field name

1) Source for the standard (= non-artificial) publications:

```
<exch:publication-reference data-format="original">
  <document-id>
    <doc-number>07691110</doc-number>
  </document-id>
</exch:publication-reference>
```

If DOCDB does not provide an original publication number, then PUBLN_NR_ORIGINAL will contain an empty string.

2) For all artificial publications the attribute PUBLN_NR_ORIGINAL will contain an empty string.

Source sub-field identifier

data-format="original"

Source codes

n/a

Comments

This attribute is useful to combine publication data of PATSTAT with another publication data set which also contains the original publication number.

On average, only about 20% of all publications do have an original publication number.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2016 - First version

6.149 REG_PHASE

Name: Indicator whether the application *is* or *has been* in the regional phase

Also Known As: n/a

Description: Indicates that an application *is* or *has been* in the regional phase.

Domain: 1 ASCII character

Y	Yes
N	No
space	not known (In case of uncertain interpretations; used very little or not at all)

Default value: N

Source database: PATSTAT;

Source field name: Derived from tables TLS201_APPLN , TLS211_PAT_PUBLN and the "Kind code concordance List" (Excel format, see <http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/regular.html>)

Y	if the APPLN_KIND <> W and (APPLN_AUTH is a regional office or (APPLN_AUTH is a member of an regional office and the PUBLN_KIND code indicates that the patent publication is the result of a regional phase));
N	otherwise

Source sub-field identifier: n/a

Comments

For explanation and disclaimer see attribute INT_PHASE in section 6.57.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2016 - First version

6.150 RESIDENCE_CTRY_CODE

Name: Country code of a person's residence

Also Known As: n/a

Description: The country of the residence - in contrast to the country of the correspondence addresses which is conventionally used.

Domain: 2 characters (A-Z), according to [WIPO ST.3](#) or spaces

Exception: in case of bad data it may be any characters (e. g. "UK" is not ST.3 compliant, but should be "GB")

Default value: spaces

Source database

USPTO data of published applications and published grants

Source field name:

<residence> <country>

Source sub-field identifier: n/a

Comments: n/a

This data is not available for applicants, only for inventors. Note that patent offices do not check the validity of the residence information.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.151 ROLE

Name: The assignee's role according to the USPTO.

Also Known As: n/a

Description: Classification of an assignee as assigned by the USPTO.

Domain: 2 characters or empty,

The meaning of the values is as follows:

Note: A "1" in the first position identifies a partial owner.

01 or 11	Unassigned
02 or 12	United States company or corporation
03 or 13	Foreign company or corporation
04 or 14	United States individual
05 or 15	Foreign individual
06 or 16	U.S. Federal government
07 or 17	Foreign government
08 or 18	U.S. county government
09 or 19	U.S. state government

All other values are data errors and therefore not defined.

Default value: empty

Source database

USPTO data of published applications and published grants

Source field name:

<assignee> <addressbook> <role>

Source sub-field identifier: n/a

Comments: n/a

This data is not available for inventors, only for applicants.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.152 SOURCE

Name: Name of the data source

Also Known As: n/a

Description: Identifies the source of the data in this record

Domain: 5 ASCII characters

DOCDB → DOCDB, EPO's Bibliographic Database

EPREG → EP Register

USPTO → USPTO's Published Applications and Published Grants

Default value: n/a

Source database: n/a - Generated; value depends on data source

Source field name: n/a

Source sub-field identifier: n/a

Comments: The data source may be refined with the attribute SOURCE_VERSION.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.153 SOURCE_VERSION

Name: Version of the data source

Also Known As: n/a

Description: Refines the attribute SOURCE

Domain: Up to 10 ASCII characters

If SOURCE = "DOCDB" → SOURCE_VERSION is empty

If SOURCE = "EPREG" → SOURCE_VERSION is empty

If SOURCE = "USPTO" → SOURCE_VERSION is "BACKFILE", "4.2", "4.3", "4.4", ...

Default value: n/a

Source database: n/a - Generated; value depends on data source

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.154 ST3_NAME

Name: Country Name

Also Known As: cc

Description: Short English name of a state, other entity or intergovernmental organisation, as defined in WIPO standard [WIPO ST.3](#)

Domain: up to 100 characters

Default value: n/a

Source database: WIPO standard ST.3

Source field name: n/a

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.155 STATE

Name: US state as part of the address

Also Known As: n/a

Description: Contains the US state as part of the address

Domain: Up to 2 ASCII characters or empty

Default value: empty string

Source database

USPTO data of published applications and published grants

Source field name:

<addressbook> <address> <state>

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

6.156 STATE_INDICATOR

Name: State indicator

Also Known As: n/a

Description: Indicates that a country code indicates a state (and not an organisation)

Domain: 1 ASCII character: Y or space

Y if this country code indicates a state (and not an organisation)
space otherwise

Default value: n/a

Source database: WIPO standard ST.3

Source field name: n/a

Source sub-field identifier: n/a

Comments: This attribute is useful if data has to be displayed on a map

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.157 STREET

Name: Street part of the address

Also Known As: n/a

Description: Contains the street part of the address

Domain: Up to 500 characters

Default value: empty string

Source database: USPTO data of published applications and published grants

Source field name:

<addressbook> <address> <street>

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

M. Kracker - 01-10-2015 – Removed source “EP Register data”;
cf. attributes ADDRESS_1, ..., ADDRESS_5

6.158 TECH_REL_APPLN_ID

Name: Application identification of the technically related application

Also Known As: n/a

Description: Surrogate key based on the elements in the candidate primary key chosen

Domain: Number 1 ... 999 999 999

Applications for which a technical relation had been found and for which no other relation is in existence. The case to be taken into account is case # 5 from section 4.6 "Relation Types", using the <priority-linkage-type> value T.

Source database: DOCDB, PATSTAT

Source field name:

```
<application-reference is-representative="YES" data-format="docdb">
  <document-id>
    <country>US</country>
    <doc-number>44896706</doc-number>
    <kind>A</kind>
    <date>20060607</date>
  </document-id>
</application-reference>
.....
<language-of-publication>en</language-of-publication>
<priority-claims>
  <priority-claim sequence="1" data-format="docdb">
    <document-id>
      <country>US</country>
      <doc-number>44896706</doc-number>
      <kind>A</kind>
      <date>20060607</date>
    </document-id>
    <priority-active-indicator>Y</priority-active-indicator>
  </priority-claim>
  <priority-claim sequence="2" data-format="docdb">
    <document-id>
      <country>US</country>
      <doc-number>32859306</doc-number>
      <kind>A</kind>
      <date>20060110</date>
    </document-id>
    <priority-linkage-type>T</priority-linkage-type>
    <priority-active-indicator>N</priority-active-indicator>
  </priority-claim>
```

The corresponding application in PATSTAT is determined (via APPLN_AUTH, APPLN_NR and APPLN_KIND) and the value of APPLN_ID for this application is the TECH_REL_APPLN_ID

If there is no corresponding application in PATSTAT it should be created, see section 4.4 "Application replenishment".

Source sub-field identifier: n/a

Comments

This field defines the relationship between an application and a prior application (priority). If the priority-linkage-type = T, then the priority is a technically related priority.

Modification history

Author of update - Date of update - Explanation of update

R. Heijna - 20-07-2005 - Source field definition improved

J. Rollinson - 17-06-2009 - Changed source to DOCDB Exchange XML

6.159 TECHN_FIELD

Name: Name of a technology field

Also Known As: n/a

Description: English name of the technology field.

Domain: Up to 50 characters

Default value: n/a

Source database:

http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/xls/ipc_technology.xls

Source field name: See FIELD_EN of the above mentioned Excel file

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.160 TECHN_FIELD_NR

Name: Number of a technology field

Also Known As: n/a

Description: Uniquely identifies a technology field. The number has little business meaning.

Domain: Number 1 ... 35

Default value: n/a

Source database:

http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/xls/ipc_technology.xls

Source field name: See FIELD_NUMBER of the above mentioned Excel file.

Source sub-field identifier: n/a

Comments: This is a classification according to *technology*. A classification according to *industries* is the NACE code which can be found in the tables TLS902_IPC_NACE2 and TLS229_APPLN_NACE2.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

M. Kracker - 01-04-2015 –comment amended

6.161 TECHN_SECTOR

Name: Name of a technology sector

Also Known As: n/a

Description: The 35 technology fields are grouped in 5 technology sectors. This attribute contains the English name of the technology sector.

Domain: Up to 50 characters

Default value: n/a

Source database:

http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/xls/ipc_technology.xls

Source field name: See SECTOR_EN of the above mentioned Excel file

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2014 - First version

6.162 UNLESS_WITH_IPC

Name: IPC main group limiting the effect of attribute NOT_WITH_IPC

Also Known As: n/a

Description: Empty or first 8 characters of an IPC symbol according to WIPO ST.3.

Domain: Up to 8 ASCII characters; Example: 'A61K 8'

Default value: empty

Source database: See Eurostat's paper described in section 5.27 "TLS902_IPC_NACE2: Mapping between IPC and industrial sectors".

Source field name: n/a

Source sub-field identifier: n/a

Comments: IPC main group which nullifies the effect of the attribute NOT_WITH_IPC column if it co-occurs with the symbol in the attribute IPC.

In the most cases this field is empty.

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2015 - First version

6.163 WEIGHT

Name: Weight of this association between the application and a classification according to an industry or a technical field.

Also Known As: n/a

Description: The higher the number, the stronger the relationship between an application and an industry / a technical field. The total of all weights of an application always equals 1.

Domain: Real number between 0 and 1

Default value: n/a

Source database: Computed from PATSTAT using reference table TLS902_IPC_NACE2 resp. TLS901_TECHN_FIELD_IPC.

Source field name: n/a

Source sub-field identifier: n/a

Comments:

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-04-2015 - First version

M. Kracker - 01-10-2015 – Extended to be also applicable to Technical Fields

6.164 ZIP_CODE

Name: Zip code of the address

Also Known As: postal code, postcode

Description: Contains the zip code part of the address

Domain: Up to 30 characters

Default value: empty string

Source database

USPTO data of published applications and published grants

Source field name:

<addressbook> <address> <postcode>

Source sub-field identifier: n/a

Comments: n/a

Modification history

Author of update - Date of update - Explanation of update

M. Kracker - 01-10-2013 - First version

7 HISTORY OF MAJOR CHANGES TO TABLES AND ATTRIBUTES

October 2005	original
March 2006	Table tls213_npl_citn has been deleted. Table TLS212_pat_citn is now called tls212_citation (our citations data model needed reworking in December)
April 2006	TLS214_NPL_TEXT has been renamed to TLS214_NPL_PUBLN to comply with the Data Catalog. The indexes on TLS207_pers_appln changed from unique-primary on person_id & appln_id to non-unique on person_id and non-unique-clustering on appln_id
September 2006	No changes.
April 2007	No changes.
October 2007	New column IPC_CLASS_LEVEL in table tls209_appln_IPC to store the advanced or core indicator; new table TLS217_appln_I_CLS to store the applications classified in the Y01N In Computer Only EPO classification scheme for Nanotechnology.
April 2008	New table TLS218_docdb_fam - links applications which have exactly the same Paris Convention priorities in table TLS204_appln_prior; new web application to download parts of the data; TLS216 column renamed to parent_appln_id (used to be parent_appl_id)
September 2008	New table TLS219_inpadoc_fam links applications directly or indirectly - this corresponds to the extended INPADOC patent family in Espacenet or OPS web service.
April 2009	<p>In table tls211_pat_publn, there is a new element publn_first_grant. This is a very tricky area - if you feel that you can improve on our rules, please discuss it with us via patstat@epo.org .</p> <p>The FAQs are now available on the internet forum; access to the forum is available via patstat@epo.org .</p> <p>In the area of citations, we no longer show the citations of patents in Non-Patent Literature (NPL) patent abstracts as separate citations. This means that before April 2009 if your programs were counting the citations for a patent , then your counts in April 2009 will be lower by the number of patents which were cited within patent abstracts.</p> <p>We no longer copy US inventor names into US assignee fields. The names and addresses from US Grants take precedence over those from US Published Applications.</p> <p>Referential integrity has been implemented for table tls207_pers_appln , to avoid applications having duplicate persons. Where possible (at the moment only for US data) , the person_name column in table tls206_person has been implemented as a concatenation of the last name, the first name and the middle names - separated by a comma.</p> <p>The separate files of person data, the TLS206_ASCII files, are intended only for users who wish to write special programs to process the name and address data. The last name, the first name and the middle names are stored in separate fields. Where the incoming data does not show the separation between the last first and middle names, then all are stored in the last-name field. For US data, the 'role' of the assignee is given.</p> <p>The web application which allows users to download a subset of the data is still active, but only for the edition of September 2008 . The usage of the service has been monitored, and it has been decided not to extend this service beyond April 2009.</p>
September 2009	Data coverage - PATSTAT has now increased data coverage for the Latin American

	<p>countries: Guatemala (GT), Chile (CL), Ecuador (EC), Nicaragua (NI), Dominican Republic (DO), Panama (PA), Costa Rica (CR), Cuba (CU), Peru (PE) and El Salvador (SV). This will result in better family information as well.</p> <p>The web application which allowed users to download a subset of the data has been switched off.</p>
April 2010	<p>New table: Table TLS217_APPLN_I_CLS has been renamed TLS217_APPLN_ECLA and contains all ECLA codes and all ICO codes (including nanotech). This table covers ECLA (EPO Classifications scheme), ICO (InComputerOnly EPO Classification scheme,), IDT (Indeling der Techniek), ECNO (ECLA symbols allocated by National Offices, not by EPO). The NANO-Technology symbols (Y01N) as provided in previous PATSTAT editions are part of the ICO scheme. This table contains extra columns that indicate the authority that assigned the code, the classification scheme and the symbol itself. You will need to keep this in mind for your scripts and queries.</p> <p>PATSTAT production process: the production of PATSTAT has been outsourced and the main data source is now the XML version of DOCDB. As a result of this change, the data quality has improved and a systematic user acceptance testing has been put in place. During this process we were able to eliminate about ten thousand duplicate "D2" applications. This new production process will guarantee a better synchronisation between DOCDB, PATSTAT and other patent information products.</p> <p>IPC related searches for documents published after 2006 allow now to find all the DOCDB simple family members consistently (in the previous editions you might have found only one or more but not all members of the family). This is due to the fact that, as (different) IPC classes can be present in DOCDB for all publication levels of an application, in PATSTAT these are now aggregated and de-duplicated at application level.</p>
September 2010	<p>PATSTAT production process: various adaptations have been introduced, the main ones being a) introduction of new citation sources in TLS212 (element 31 CITN_ORIGIN) and b) change of source to DOCDB XML for element PUBLN_FIRST_GRANT.</p> <p>The table TLS211_PAT_PUBLN contains the column PUBLN_FIRST_GRANT. If this has the value '1', then that publication is the 'first grant'. In April 2010, the method for calculating this was based on the publication kind code representing a grant in each country, and then selecting the earliest publication. In September 2010 we use the 'public-availability' tag in the DOCDB XML product from the EPO.</p> <p>New table: Table TLS221_INPADOC_PRS containing INPADOC worldwide legal status data has been created and integrated into the PATSTAT database structure. However it has been produced on a test basis only, it will be available as of April 2011 edition but will have to be acquired separately.</p>
April 2011	<p>Table TLS201_APPLN: New permanent unique application identifier introduced in APPLN_ID. With the April 2011 edition, the DOCDB "doc-id" unique and stable identifier has been used to populate APPLN_ID instead of creating a PATSTAT-edition-specific surrogate key (<i>but not for the artificial applications in PATSTAT</i>). DOCDB attribute "doc-id" contains a stable and unique identifier that will allow for linking up a number of EPO raw data products through the application in a reliable way. This attribute will remain the same across PATSTAT editions and will always refer to the same combination of application authority, application number and application kind.</p> <p>Table TLS209_APPLN_IPC: IPC Core Level symbols are no longer maintained in WIPO ST8. Until September 2011, PATSTAT collected both the Advanced and Core sets of symbols for each application. The IPC Core symbols are now obsolete and have been eliminated from DOCDB, unless a publication had a Core symbol but no Advanced symbol. PATSTAT now shows the Advanced symbols, however Core symbols are still shown in those cases where no Advanced symbol is available but a Core symbol is still present in DOCDB.</p>

October 2011	<p>Creation of two additional tables TLS222 and TLS223 for JP and US national classifications, to reflect the additional national classification symbols now available in DOCDB XML.</p> <p>Addition of elements CITN_GENER_AUTH and CITED_APPLN_ID in citation table TLS212 to provide the International Search Authority (ISA) for PCT published application and to additionally provide details on cited applications (cited by the applicant).</p> <p>Addition of the number of claims for EP and US publications in table TLS211: the values are provided in the element PUBLN_CLAIMS.</p> <p>Improvements to existing data include adding 9 million abstracts in English language for Japanese published applications in table TLS203</p> <p>To avoid confusion among table TLS210 and the newly created tables TLS222 and TLS223, all JP and US national classification symbols present in TLS210 have been removed.</p>
April 2012	<p>Table TLS214 NPL: the surrogate key has been replaced with XP number <refno> from DOCDB, this 9 digit number is used now as surrogate key.</p> <p>Element 57 PUBLN_CLAIMS: for the number of claims relating to EP B publications (granted patents) we have added the missing values for the years 1980-2005 in addition to the number of claims already available in October 2011 (2006 to date). For the US B publications (granted patents) we are now able to provide a more timely coverage (up to end 2011 in this edition).</p>
October 2012	<p>In table TLS201_APPLN, an additional rule has been implemented to further identify national applications originating from the PCT. This has raised the number of applications with an INTERNAT_APPLN_ID >0 from 4,850,479 to 5,319,404.</p> <p>In Table TLS212_CITATION, the domain for element 31 CITN_ORIGIN has changed: a new value PRS (for "PRE-Search" citations), to be attributed value 9 in PATSTAT, has been added.</p> <p>Remark: the ECLA classification scheme will be replaced by the new CPC - Cooperative Patent Classification on 01-01-2013. The next PATSTAT edition April 2013 will contain only CPC.</p>
April 2013	<ul style="list-style-type: none"> • Table TLS224_APPLN_CPC has replaced TLS217_APPLN_ECLA . • Leading blanks in the attributes APPLN_NR and PUBLN_NR have been removed. • The APPLN_ID ranges for artificial applications and the PUBLN_ID ranges for artificial publications are now fixed and do not depend any more on the number of applications / publications. • Artificial applications which do not have a publication (= their APPLN_ID is > 930 000 000) are not assigned to any INPADOC family
October 2013	<ul style="list-style-type: none"> • EP Register is used (again) instead of EP Bulletin as a data source for names and address of applicants and inventors of EP patents. • New table TLS226_PERSON_ORIG which replaces the previous file TLS206_PERSON_ASCII. It contains the unmodified name and address data for all persons in PATSTAT • New table TLS227_PERS_PUBLN which links applicants and inventors to publications. • Attributes PERSON_ID and PERSON_ORIG_ID are from now on stable • Freeform name strings from the USPTO data source are constructed more consistently with other freeform names • De-duplication rules for table TLS206_PERSONS are changed • New attribute TLS201_APPLN.APPLN_EPODOC_NR to easily connect to Espacenet etc. • Additional values for TLS212_CITATION.CITN_ORIGIN and reordering of columns • Unused columns of TLS221_INPADOC_PRS removed: L514EP, L521EP, L526EP, L527EP • Change in computation of TLS201_APPLN.IPR_TYPE
2014 Spring Edition	<ul style="list-style-type: none"> • New tables TLS801_COUNTRY and TLS802_LEGAL_EVENT_CODE

	<ul style="list-style-type: none"> • Table TLS901_TECHN_FIELD_IPC replaces table TECHN_FIELD_IPC • New attribute LEC_ID in table TLS221_INPADOC_PRS • New default values for dummy applications (APPLN_ID = 0) and dummy publications (PAT_PUBLN_ID = 0) • Attribute L519EP has extended string length • Attribute L520EP is now numeric • Attribute TECHN_FIELD in table TLS209_APPLN_IPC renamed to TECHN_FIELD_NR • Value "P" (provisional application) defined for attribute APPLN_KIND
2014 Autumn Edition	<ul style="list-style-type: none"> • Double quotes (") are replaced by single quotes (') • PUBLN_ID is now a stable attribute • IPC classification symbols which are classified on subclass level only are now included (IPC_CLASS_LEVEL = 'S') • Deduplication rules when merging IPCs from different publications into their application has been slightly adapted. Same rules apply also for CPC classification symbols.
2015 Spring Edition	<ul style="list-style-type: none"> • Several tables and attributes which were available only in PATSTAT Online are now also in PATSTAT Raw Data: details see below. • New table TLS906_PERSON has been added. It includes all data from TLS206_PERSON, plus all data about harmonized names previously only available for PATSTAT Online. It is advised that users replace the original table TLS206_PERSON by this new extended table TLS906_PERSON. • Table TLS208_DOC_STD_NMS has been removed and its content integrated into TLS206_PERSON. • Table INDUSTRY_IPC has been replaced by table TLS902_IPC_NACE, which is now also available in PATSTAT Raw Data. The new table represents the new IPC – NACE concordance table published by Eurostat in 2014. • New table TLS229_APPLN_NACE2 has been added. • Table DOCDB_FAMILY_CITATION has been renamed to TLS228_DOCDB_FAM_CITN and made available to PATSTAT Raw Data. Its attributes have been re-ordered and partly renamed. • New column ISO_ALPHA3 has been added to table TLS801_COUNTRY. • In table TLS201_APPLN the name of attribute NB_CITATIONS has been changed to NB_CITING_DOCDB_FAM. • The order of attributes has changed in table TLS901_TECHN_FIELD_IPC. • The attribute DOC_STD_NAME_ID is no longer a stable ID. • Duplicates in persons have been removed. • Some computed attributes of PATSTAT Online are regarded as deprecated and may be removed in future editions: In table TLS201_APPLN: - APPLN_FILING_YEAR_MONTH - APPLN_FILING_YEAR - PRIOR_EARLIEST_YEAR_MONTH - PRIOR_EARLIEST_YEAR - PUBLN_EARLIEST_YEAR_MONTH - PUBLN_EARLIEST_YEAR - DOCDB_FAMILY_ID In table TLS209_APPLN_IPC: - IPC_SUBCLASS_SYMBOL In table TLS224_APPLN_CPC: - CPC_MAINGROUP_SYMBOL
2015 Autumn Edition	<ul style="list-style-type: none"> • Removed differences between the data models of PATSTAT Raw Data and PATSTAT Online, by removing these pre-computed and redundant attributes: - APPLN:FILING_YEAR_MONTH from TLS201_APPLN - PRIOR:EARLIEST_YEAR_MONTH from TLS201_APPLN - PUBLN:EARLIEST_YEAR_MONTH from TLS201_APPLN - PUBLN:EARLIEST_REF from TLS201_APPLN - IPC_SUBCLASS_SYMBOL from TLS209_APPLN_IPC

	<ul style="list-style-type: none"> - TECHN_FIELD_NR from TLS209_APPLN_IPC - CPC_MAINGROUP_SYMBOL from TLS224_APPLN_CPC. • Re-ordered attributes in table TLS201_APPLN and added attribute EARLIEST_FILING_ID • Renamed these attributes of TLS201_APPLN: <ul style="list-style-type: none"> - PRIOR_EARLIEST_DATE to EARLIEST_FILING_DATE - PRIOR_EARLIEST_YEAR to EARLIEST_FILING_YEAR - PUBLN_EARLIEST_DATE to EARLIEST_PUBLN_DATE - PUBLN_EARLIEST_YEAR to EARLIEST_PUBLN_YEAR • Moved attribute APPLN_TITLE_LG from TLS201_APPLN to TLS202_APPLN_TITLE • Moved attribute APPLN_ABSTRACT_LG from TLS201_APPLN to TLS203_APPLN_ABSTRACT • Table TLS212_CITATION: In attribute CITN_ORIGIN code "115" has been renamed to "TPO" (Third Party Observation) • Tables TLS218_DOCDB_FAM and TLS219_INPADOC_FAM have been integrated into table TLS201_APPLN • TLS226_PERSON_ORIG: 5 new attributes for 5 address lines for addresses of EP applications. • New table TLS230_APPLN_TECHN_FIELD
2015 Autumn Edition - Amended	<ul style="list-style-type: none"> • Extended the domains for attributes DOCDB_STD_NAME, CITN_ID, PAT_CITN_SEQ_NR and NPL_CITN_SEQ_NR • New processing rules for attributes of table TLS212_CITATION, to adapt for changes in DOCDB
2016 Spring	<ul style="list-style-type: none"> • Table TLS201_APPLN: New attributes APPLN_NR_ORIGINAL, INT_PHASE, REG_PHASE and NAT_PHASE. • Tables 206_PERSON and 906_PERSON: Attributes for EEE-PPAT names have been renamed from "HRM_..." to "PSN_..." (<u>PATSTAT Standardised Name</u>) and are now also available in TLS206_PERSON. Attribute HRM_L1 has been removed. The length of all person attributes has been unified to 500 characters. • Table TLS211_PAT_PUBLN: New attribute PUBLN_NR_ORIGINAL • Table TLS212_CITATION: Attribute NPL_PUBLN_ID has been renamed to CITED_NPL_PUBLN_ID • Table TLS214_NPL_PUBLN: New attribute NPL_TYPE. NPL references which contain in their biblio text something like "none" or "See also references .." are removed. • TLS226_PERSON_ORIG: Attributes ADDRESS_1 - ADDRESS_3 are now also used for the USPTO data source. • TLS229_APPLN_NACE2: Due to an update by Eurostat the logic has slightly changed.
2016 Autumn	<ul style="list-style-type: none"> • Line breaks are replaced by " \n" in text attributes. • All artificial applications now also belong to a DOCDB family and an INPADOC family. They will have no other family member. See attributes DOCDB_FAMILY_ID and INPADOC_FAMILY_ID. • Table TLS205_TECH_REL now also contain inverse relations, because technical relations are by definition symmetric. • Regionalisation information has been added: <ul style="list-style-type: none"> - new attributes NUTS and NUTS_LEVEL in table TLS906_PERSON - new reference table TLS904_NUTS

8 KNOWN DEFICIENCIES

Data coverage issues are not described in this section. Data coverage information can be found in <http://www.epo.org/searching-for-patents/helpful-resources/raw-data/data/tables/weekly.html> .

- **TLS206_PERSON / TLS906_PERSON:**

- **DOCDB standardized names:**

- Some DOCDB standardised names are wrongly assigned to persons of US patents, because the sequence of persons in the USPTO data source and that in DOCDB sometimes do not match correctly.

- There is no know fix. When working with US patent applicants or inventors, you should avoid using the DOCDB standardised name. Instead, you might consider other harmonized names available in tables TLS206_PERSON and TLS906_PERSON.