The REG101_APPLN Table

Welcome to the first table of PATSTAT Register, namely table **REG101_APPLN**. This table contains:

- EP applications with their identifiers and some additional data. In case of Euro-PCT, i.e. an international application which has entered the EP regional phase, then the number of the international application number is given. If no international application number is given, then it is an EP direct filing.
- International applications, which have not (yet) entered the EP regional phase.

In [14]:
```
from epo.tipdata.patstat import PatstatClient
from epo.tipdata.patstat.database.models import REG101_APPLN
from sqlalchemy import func
import pandas as pd

# Initialise the PATSTAT client
patstat = PatstatClient(env='PROD')

# Access ORM
db = patstat.orm()
```

ID (Primary Key) ¶

A technical identifier for an application, without business meaning. Its values will not change from one PATSTAT edition to the next.

In [2]:
```
i = db.query(
    REG101_APPLN.id
).limit(1000)

df = patstat.df(i)
df
```

Out[2]:

	id
0	6719360
1	9005130
2	97306940
3	97932748
4	10714541
...	...
995	19947900
996	91309804
997	90916485
998	13862355
999	20184626

1000 rows × 1 columns

APPLN_ID

Application identifier. `APPLN_ID` > 0 indicates that the application is either an EP direct application or a PCT international application which has entered the EP regional phase. The corresponding EP application in the PATSTAT Global database is linkable via the attribute `APPLN_ID` which occurs in both databases. `APPLN_ID` = 0 indicates that this application has not entered the EP regional phase. Therefore the PATSTAT Global database does not contain a corresponding EP application.

```
In [3]: i = db.query(
            REG101_APPLN.id,
            REG101_APPLN.appln_id
        ).limit(1000)

        df = patstat.df(i)
        df
```

```
Out[3]:
               id    appln_id
     0    6719360    16331990
     1    9005130    57180350
     2   97306940    17246447
     3   97932748    17276609
     4   10714541   315486707
   ...        ...         ...
   995   19947900           0
   996   91309804    16897624
   997   90916485           0
   998   13862355   419206699
   999   20184626   534194298

1000 rows × 2 columns
```

APPLN_AUTH

Office where the application has been filed. In table `REG101_APPLN` this attribute is always 'EP', because the scope of this database is European patents.

In [4]:
```
# Use the count function in the query and rename the column via t
he label command
group_q = db.query(
    func.count(REG101_APPLN.appln_id).label('total_application
s'),
    REG101_APPLN.appln_auth
).group_by(
    REG101_APPLN.appln_auth  # Here we use the group_by function
on the 'appln_auth' field
).order_by(
    func.count(REG101_APPLN.appln_id)
)

# Convert it in a dataframe
grouped_res = patstat.df(group_q)
grouped_res
```

Out[4]:

	total_applications	appln_auth
0	6933144	EP

The only application authority is indeed 'EP', since PATSTAT Register concerns EPO applications and patents only.

APPLN_NR

Application number as issued by the application authority. Note that the attribute must not be a numerical attribute but a text string attribute because leading zeros are significant.

In [2]:
```
appln_nr = db.query(
    REG101_APPLN.appln_nr,
    REG101_APPLN.appln_id
).limit(1000)

appln_nr_df = patstat.df(appln_nr)
appln_nr_df
```

Out[2]:

	appln_nr	appln_id
0	12804625	379455231
1	20181931	533200863
2	12707607	352155859
3	21929883	0
4	10183917	323912451
...	...	...
995	88118974	16718189
996	23879304	0
997	98962991	0
998	12775175	0
999	04712268	16135089

1000 rows × 2 columns

Except for the value 0, the `appln_id` attribute is associated to only one `appln_nr` .

In [6]:
```
tot_nr = db.query(
    func.count(REG101_APPLN.appln_nr).label('tot_appln_nr'),
    REG101_APPLN.appln_id
).group_by(
    REG101_APPLN.appln_id
).order_by(
    func.count(REG101_APPLN.appln_nr).label('tot_appln_nr')
)

tot_nr_df = patstat.df(tot_nr)
tot_nr_df
```

Out[6]:

	tot_appln_nr	appln_id
0	1	523850827
1	1	496077405
2	1	15889179
3	1	553184033
4	1	405107888
...	...	...
4396500	1	579631981
4396501	1	16963292
4396502	1	17067662
4396503	1	488158865
4396504	2536640	0

4396505 rows × 2 columns

APPLN_FILING_DATE

Date on which the application was received at the Patent Authority.

Notice that there several applications with filing application year equal to 9999. These correspond to missing dates.

In [9]:
```
missing_dates = db.query(
    REG101_APPLN.appln_filing_date,
    REG101_APPLN.appln_id
).order_by(
    REG101_APPLN.appln_filing_date.desc()
).limit(50000)

missing_dates_df = patstat.df(missing_dates)
missing_dates_df
```

Out[9]:

	appln_filing_date	appln_id
0	9999-12-31	446034451
1	9999-12-31	0
2	9999-12-31	0
3	9999-12-31	0
4	9999-12-31	0
...	...	...
49995	2023-10-18	0
49996	2023-10-18	0
49997	2023-10-18	0
49998	2023-10-18	0
49999	2023-10-18	600319405

50000 rows × 2 columns

Missing dates may correspond to withdrawn applications. Whichever the reason is, these cases represent a tiny portion of the database.

In [7]:
```
num_missing_dates = db.query(
    REG101_APPLN.appln_filing_date,
    func.count(REG101_APPLN.appln_id).label('num_appln')
).filter(
    REG101_APPLN.appln_filing_date == '9999-12-31'
).group_by(
    REG101_APPLN.appln_filing_date
).order_by(
    REG101_APPLN.appln_filing_date.desc()
)

num_missing_dates_df = patstat.df(num_missing_dates)
num_missing_dates_df
```

Out[7]:

	appln_filing_date	num_appln
0	9999-12-31	22

FILING_LG

The language in which the application was filed.

Let's see which are the most frequent languages in the database.

In [8]:
```
lg = db.query(
    REG101_APPLN.filing_lg,
    func.count(REG101_APPLN.appln_id).label('tot_appln')
).group_by(
    REG101_APPLN.filing_lg
).order_by(
    func.count(REG101_APPLN.appln_id).desc()
)

lg_df = patstat.df(lg)
lg_df
```

Out[8]:

	filing_lg	tot_appln
0	en	3920040
1	de	929429
2	ja	816829
3	zh	541945
4	fr	296158

5	ko	193328
6	it	80381
7	es	38136
8	ru	22311
9	nl	21126
10	sv	17668
11		15553
12	fi	10482
13	pt	7252
14	tr	5557
15	no	3991
16	da	3941
17	pl	3574
18	cs	1731
19	hu	1061
20	sl	856
21	el	503
22	sk	373
23	ar	245
24	hr	130
25	lt	117
26	th	106
27	et	104
28	ro	88
29	bg	68
30	lv	37
31	uk	4
32	se	3
33	sh	3
34	ch	2
35	be	1
36	xx	1
37	ee	1
38	sa	1

39	za	1
40	sc	1
41	lb	1
42	kn	1
43	ka	1
44	sr	1
45	an	1
46	mh	1

STATUS

Status of the application. The status of granted patents (e. g. still valid or not) is not included here. The domain consists of numbers from 1 to 18. This attribute permits to link this table to table REG403_APPLN_STATUS. Therefore, its meaning and utility will be explained later on in this series of notebooks.

In [9]:
```
status = db.query(
    REG101_APPLN.status
).limit(1000)

status_df = patstat.df(status)
status_df
```

Out[9]:

	status
0	10
1	14
2	7
3	10
4	10
...	...
995	10
996	10
997	10
998	9
999	10

1000 rows × 1 columns

Raw numbers like this are not meaningful. We can see the actual text corresponding to the status joining table REG101 with table REG403 as said above.

In [15]:
```
from epo.tipdata.patstat.database.models import REG403_APPLN_STAT
US

text_status = db.query(
    REG101_APPLN.status,
    REG403_APPLN_STATUS.status_text
).distinct(
    REG101_APPLN.status
).join(
    REG101_APPLN, REG403_APPLN_STATUS.status == REG101_APPLN.stat
us
).order_by(
    REG101_APPLN.status
)

text_status_df = patstat.df(text_status)
text_status_df
```

Out[15]:

	status	status_text
0	1	Patent revoked by proprietor
1	2	The patent has been limited
2	3	Patent maintained as amended
3	4	Patent revoked
4	5	Opposition rejected
5	6	Opposition procedure closed
6	7	No opposition filed within time limit
7	8	The patent has been granted
8	9	The application has been withdrawn
9	10	The application is deemed to be withdrawn
10	11	The application has been refused
11	12	Grant of patent is intended
12	13	Proceedings closed following consolidation wit...
13	14	Examination is in progress
14	15	Request for examination was made
15	16	The application has been published
16	17	The international publication has been made

INTERNAT_APPLN_ID

Application ID of the international application. This attribute allows for each PCT application in PATSTAT EP Register to easily retrieve the corresponding PCT application in PATSTAT Global. You just have to join the non-zero values of attribute `REG101_APPLN.INTERNAT_APPLN_ID` with the attribute `TLS201_APPLN.APPLN_ID` .

In [10]:
```
from epo.tipdata.patstat.database.models import TLS201_APPLN

internat = db.query(
    REG101_APPLN.internat_appln_id,
    TLS201_APPLN.appln_id
).filter(
    REG101_APPLN.internat_appln_id != 0
).join(
    REG101_APPLN, TLS201_APPLN.appln_id == REG101_APPLN.internat_appln_id
)

internat_df = patstat.df(internat)
internat_df
```

Out[10]:

	internat_appln_id	appln_id
0	379450165	379450165
1	590299660	590299660
2	445451735	445451735
3	54533126	54533126
4	42921430	42921430
...	...	...
4870667	568713537	568713537
4870668	523325058	523325058
4870669	443833080	443833080
4870670	315873477	315873477
4870671	54531355	54531355

4870672 rows × 2 columns

INTERNAT_APPLN_NR

International application number. It consists of up to 15 characters.

In [11]:
```
internat_appln_nr = db.query(
    REG101_APPLN.internat_appln_id,
    REG101_APPLN.internat_appln_nr
).limit(1000)

internat_appln_nr_df = patstat.df(internat_appln_nr)
internat_appln_nr_df
```

Out[11]:

	internat_appln_id	internat_appln_nr
0	523246300	WO2019EP83225
1	0	
2	587829525	WO2023EP53809
3	25475238	WO2001JP09629
4	57900284	WO2008US73146
...	...	...
995	56433255	WO2007US84376
996	17541504	WO2007EP59050
997	0	
998	561964070	WO2020CN93371
999	0	

1000 rows × 2 columns

BIO_DEPOSIT

An indicator telling whether one or more deposits of biological material have been made or not. The domain consists of one ASCII character: Y or N.

Let's count how many Y and N there are in the database.

In [13]:
```
bio = db.query(
    REG101_APPLN.bio_deposit,
    func.count(REG101_APPLN.appln_id).label('number_of_applicatio
ns')
).group_by(
    REG101_APPLN.bio_deposit
).order_by(
    func.count(REG101_APPLN.appln_id).desc()
)

bio_df = patstat.df(bio)
bio_df
```

Out[13]:

	bio_deposit	number_of_applications
0	N	6911225
1	Y	21919

In [ ]:
```
```