

How to load PATSTAT data in your database

As explained in its [product page](#), PATSTAT data consists of biyearly PATSTAT Global and PATSTAT EP Register packages that can be purchased separately and downloaded from the [EPO Bulk Data Distribution Service](#).

Contrary to [PATSTAT Online](#) that allows you to run searches using SQL queries in an EPO graphical user interface, the use of PATSTAT data requires initially to load this data in your own database, this activity consisting of the following steps for a given PATSTAT package (spring or autumn edition of PATSTAT Global or PATSTAT EP Register):

1. Unzip PATSTAT package

Unzip all zip files so that all CSV files are extracted. There is one or multiple CSV files per table with the following characteristics:

- Windows-style line ending `\r\n` (CRLF)
- UTF-8 encoding
- Header with comma-separated list of table columns (attributes)
- Comma “,” used as separator between values
- Double quotes “” used as delimiters for values of type string, e.g. for names (“Smith, John”), titles, abstracts, and not used for other types, e.g. for dates and integers (2019-12-31, 123)
- Point “.” used as decimal separator, e.g. 0.125

2. Create PATSTAT database

The PATSTAT package also includes a number of SQL scripts in the folder SQL that you may need to adapt to your technical environment. Creating a PATSTAT database usually consists of the following:

- Create an empty database: see script in \SQL\CreateDatabase
- Create empty tables: see scripts in \SQL\CreateTable
- Load CSV data in tables: no scripts provided for this operation
- Create indexes: see scripts in \SQL\CreateIndex

Once the above operations are successfully completed, you may want to test your database using the SQL script located in \SQL\Test. This script allows you to count table rows so that you could compare your results with the EPO results included in the TXT file in the same folder.



Notes:

- For PATSTAT Global, the loading of patent abstracts in table TLS203_APPLN_ABSTR may take quite some time and may require quite a lot of storage capacity
- When using MySQL, you may select the character set `utf8mb4` because MySQL's character set `utf8` is restricted to 3 bytes
- A significant amount of disk space may be required to store loading logs and indexes

A PATSTAT package also has a folder DOC including:

- this document
- the edition release notes
- the data catalogue (for PATSTAT Global or for EP Register, depending on the package you purchased)

Many other PATSTAT-related documents are available on the [EPO website](#).

Contact us for

⇒ [questions on PATSTAT access, subscription or prices](#)

⇒ [other PATSTAT-related questions](#)