



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

*Liberté
Égalité
Fraternité*

Brevets de l'Office européen des brevets depuis 2010 et où au moins une partie au brevet est localisée en France

Ministère de l'Enseignement supérieur et de la Recherche

2 août 2023

Table des matières

1	Origine des données	2
1.1	La base de données PATSTAT de l'Office européen des brevets	2
1.2	La base de données brevets de l'INPI	2
1.3	Classification coopérative des brevets	2
2	Les traitements des données	3
2.1	Schéma de processus de traitement	3
2.2	Familles de brevets	3
2.3	Parties aux brevets	3
3	Jeux de données disponibles	4
3.1	Jeu de données sur les familles de brevets	4
3.2	Jeu de données sur les technologies des familles de brevets	6
3.3	Jeu de données sur les demandes de brevets	7
3.4	Jeu de données sur les déposants des brevets	9

1 Origine des données

Les données sont issues de la :

- base de données PATSAT de l'Office européen des brevets,
- base de données des brevets de l'INPI et
- Classification coopérative des brevets

1.1 La base de données PATSTAT de l'Office européen des brevets

Les données utilisées sont extraites des tables :

- 201 pour les demandes,
- 202 pour les titres des demandes,
- 203 pour les résumés des demandes,
- 204 pour les priorités des demandes,
- 206 et 207 pour les personnes parties à la demande
- 209, 224 et 225 pour les classifications des technologies
- 211 pour les publications des brevets

Seules les demandes réalisées depuis 2010 et où une des parties est identifiée comme étant localisée en France sont retenues.

1.2 La base de données brevets de l'INPI

Elle est utilisée pour obtenir les numéros de publication, les noms des déposants et les SIREN pour les organisations françaises.

1.3 Classification coopérative des brevets

La base PATSTAT contient les codes de classification des technologies des brevets. Pour enrichir cette information, il est fait usage du fichier CPC Scheme XML du site de la Classification coopérative des brevets – coopération entre les offices européen et américain des brevets. Cette classification étend celle de l'Organisation mondiale de la propriété intellectuelle et est donc plus complète.

2 Les traitements des données

Les données issues des extractions des bases PATSAT et INPI, ainsi que du fichier CPC Scheme XML, du Bulletin officiel des annonces commerciales (BODACC), de la base Sirene de l'INSEE et ScanR sont soumis à un processus de mise en qualité et d'enrichissement.

2.1 Schéma de processus de traitement

Le code est disponible sur GitHub : <https://github.com/dataesr/harvest-patstat>.

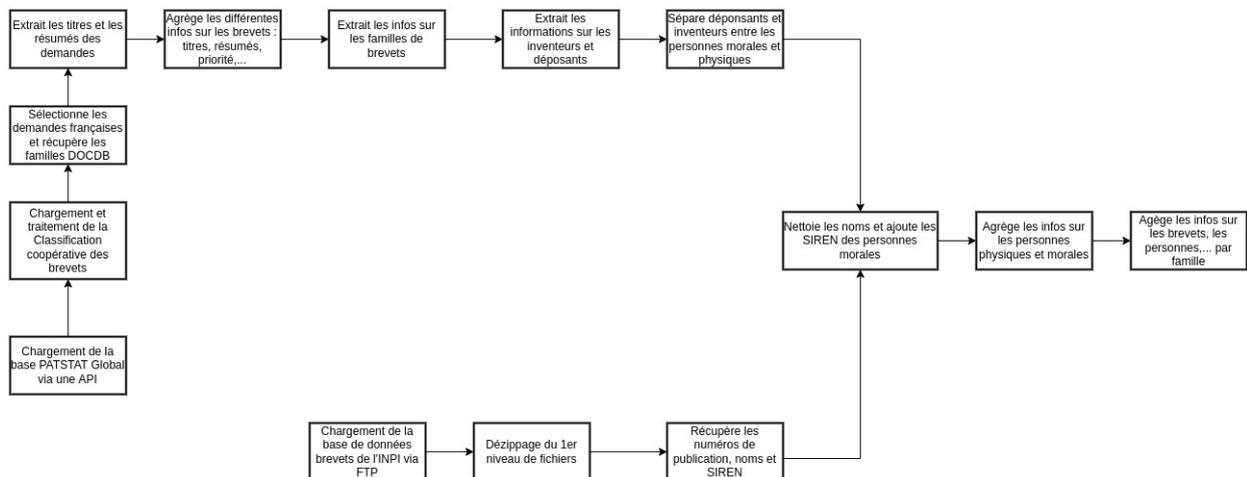


FIGURE 2.1 – Schéma du processus de traitement

2.2 Familles de brevets

Les brevets qui partagent les mêmes priorités font partie d'une même famille. Les informations sont organisées autour de ces familles. Le traitement a donc pour objectif d'agréger les données concernant les technologies, parties, publications, . . . par famille. Les dates, titres, résumés, . . . sont ceux de la première demande ou de la première délivrance au sein de la famille.

2.3 Parties aux brevets

Pour pouvoir enrichir les données, les personnes physiques sont séparées des personnes morales à l'aide :

- de la variable `invnt_seq_nr` où toutes les valeurs supérieures à zéro correspondent à des inventeurs, donc à des personnes physiques et
- d'un algorithme de classification de noms des parties à la demande.

Pour les personnes morales, les données de PATSTAT sont croisées avec celle de l'INPI, du BODACC, de ScanR et de l'INSEE pour récupérer les SIREN.

3 Jeux de données disponibles

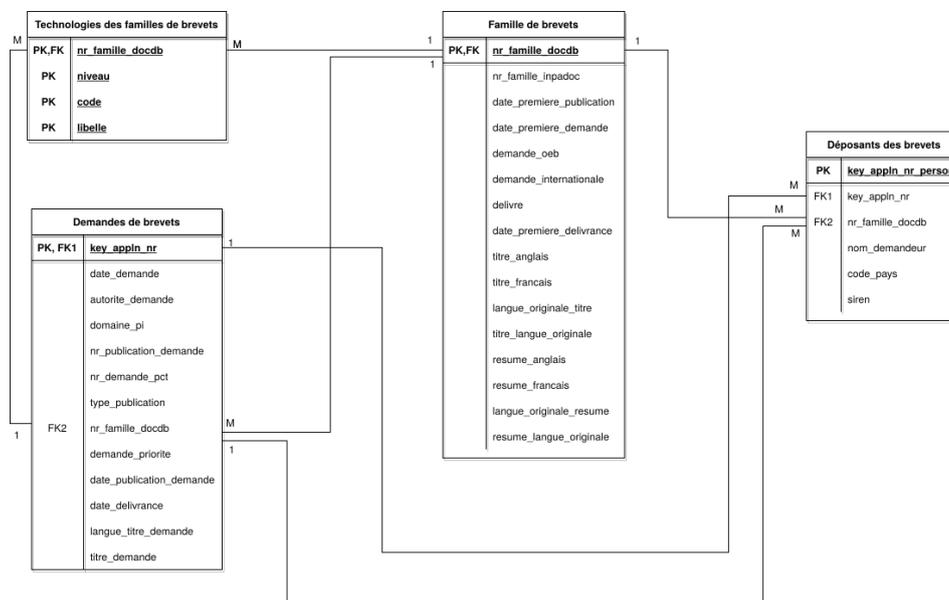


FIGURE 3.1 – Schéma des jeux de données disponibles

3.1 Jeu de données sur les familles de brevets

Ce jeu de données est disponible à l’adresse <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-familles-brevets/>. Il donne des informations sur :

- les numéros de familles DOCDB et INPADOC,
- les dates de la première demande, de la première publication et de la première délivrance des brevets de la famille,
- l’existence d’une demande au niveau européen ou international (PCT),
- la délivrance ou non de la famille,
- les titres et résumés de la famille en anglais, français et langue originale.

La clef primaire de ce jeu de données est `nr_famille_docdb`.

Nom de la variable	Descriptif de la variable	Type
nr_famille_docdb	Numéro de la famille DOCDB; au sein d'une famille DOCDB, les différentes demandes partagent exactement les mêmes priorités	texte
nr_famille_inpadoc	Numéro de famille INPADOC; la demande partage une priorité directe ou via une demande tierce	texte
date_premiere_publication	Date de première publication de la famille	date
date_premiere_demande	Date de première demande de la famille	date
demande_oeb	Existence d'une demande auprès de l'Office européen des brevets	texte
demande_internationale	Existence d'une demande internationale PCT	texte
delivre	Délivrance d'au moins une demande de brevet au sein de la famille	texte
date_premiere_delivrance	Date de première délivrance d'une des demandes de brevets de la famille	date
titre_anglais	Titre en anglais	texte
titre_francais	Titre en français	texte
langue_originale_titre	Code de la langue originale du titre	texte
titre_langue_originale	Titre dans la langue originale	texte
resume_anglais	Résumé en anglais	texte
resume_francais	Résumé en français	texte
langue_originale_resume	Code de la langue originale du résumé	texte
resume_langue_originale	Résumé dans la langue originale	texte

TABLE 3.1 – Familles de brevets

3.2 Jeu de données sur les technologies des familles de brevets

Ce jeu de données est disponible à l'adresse <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-technologies-brevets/>. Il donne des informations sur :

- le numéro de famille DOCDB et
- les niveaux, codes et libellés des familles de brevets

Les technologies des familles de brevets se basent sur la Classification coopérative des brevets. Seules les sections, classes et sous-classes sont présentées.

Nom de la variable	Descriptif de la variable	Type
nr_famille_docdb	Numéro de la famille DOCDB ; au sein d'une famille DOCDB, les différentes demandes partagent exactement les mêmes priorités	texte
niveau	Niveau de subdivision de la classification coopérative des brevets	texte
code	Code de la subdivision	texte
libelle	Libellé de la subdivision	texte

TABLE 3.2 – Technologies des familles de brevets

Chaque famille ayant plusieurs niveaux et codes des technologies, la clef correspond à l'ensemble des variables.

3.3 Jeu de données sur les demandes de brevets

Ce jeu de données est disponible à l'adresse <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/demandes-de-brevets>. Il donne des informations sur :

- l'identifiant constitué à partir de l'autorité qui a reçu la demande, du numéro de demande auprès de l'autorité, du type de la demande, de l'autorité qui a reçu la demande internationale
- les numéros de famille DOCDB
- la date de la demande et l'autorité auprès de qui la demande a été faite,
- le domaine de propriété industrielle couvert,
- le numéro de publication et le type de publication
- le numéro de demande relevant de la procédure du Traité de coopération sur les brevets (Patent Cooperation Treaty, PCT) et la revendication d'un droit de priorité,
- la date de publication de la demande
- la date de délivrance de la demande et
- la langue du titre de la demande et le titre qui lui est associée.

La clef primaire de ce jeu de données est `key_appln_nr`.

Nom de la variable	Descriptif de la variable	Type
key_appln_nr	Identifiant unique constitué de l'autorité qui a reçu la demande, du numéro de demande auprès de l'autorité, du type de la demande, de l'autorité qui a reçu la demande internationale	texte
date_demande	Date de la demande	date
autorite_demande	Autorité nationale, régionale ou internationale en charge du traitement de la demande	texte
domaine_pi	Domaine de propriété intellectuelle couvert par la demande (brevet, modèle ou certificat d'utilité)	texte
nr_publication_demande	Numéro de publication	texte
nr_demande_pct	Numéro d'identification de la demande de brevet de la procédure Patent Cooperation Treaty (PCT)	texte
type_publication	Type de publication tel qu'attribué par l'autorité de publication ; leur signification dépend de l'autorité	texte
nr_famille_docdb	Numéro de la famille DOCDB ; au sein d'une famille DOCDB, les différentes demandes partagent exactement les mêmes priorités	texte
demande_priorite	Existence d'une demande de priorité	texte
date_publication_demande	Date de publication de la demande	date
date_delivrance	Date de délivrance de la demande	date
langue_titre_demande	Code de la langue du titre de la demande	texte
titre_demande	Titre de la demande	texte

TABLE 3.3 – Demandes de brevets

3.4 Jeu de données sur les déposants des brevets

Ce jeu de données est disponible à l'adresse <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/deposants-des-brevets/>. Il donne des informations sur :

- l'identifiant constitué à partir de l'autorité qui a reçu la demande, du numéro de demande auprès de l'autorité, du type de la demande, de l'autorité qui a reçu la demande internationale
- les numéros de famille DOCDB
- le nom du demandeur, son SIREN et son code pays

Les SIREN sont récupérés de quatre sources différentes :

- la base de données brevets de l'INPI à l'aide des numéros de publication,
- la base de données du BODACC à l'aide des noms et adresses,
- la base de données ScanR à l'aide des noms ou
- la base de données Sirene de l'INSEE à l'aide des noms.

Nom de la variable	Descriptif de la variable	Type
key_appln_nr	Identifiant unique constitué de l'autorité qui a reçu la demande, du numéro de demande auprès de l'autorité, du type de la demande, de l'autorité qui a reçu la demande internationale	texte
nr_famille_docdb	Numéro de la famille DOCDB ; au sein d'une famille DOCDB, les différentes demandes partagent exactement les mêmes priorités	texte
nom_demandeur	Nom standardisé du déposant tel qu'enregistré dans la base de données de l'Office européen des brevets	texte
code_pays	Code pays du déposant	texte
siren	SIREN	texte
key_appln_nr_person	Identifiant unique constitué de l'autorité qui a reçu la demande, du numéro de demande auprès de l'autorité, du type de la demande, de l'autorité qui a reçu la demande internationale et identifiant de la personne	texte

TABLE 3.4 – Déposants des brevets

La clef primaire de ce jeu de données est key_appln_nr_person.

Table des figures

2.1	Schéma du processus de traitement	3
3.1	Schéma des jeux de données disponibles	4

Liste des tableaux

- 3.1 Familles de brevets 5
- 3.2 Technologies des familles de brevets 6
- 3.3 Demandes de brevets 8
- 3.4 Déposants des brevets 9