

Base administrative SAE 2023 – Documentation

Sommaire

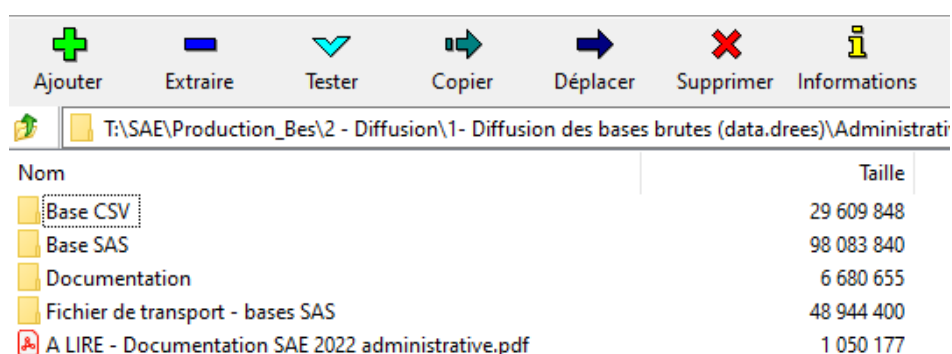
A-	Accès aux données et contenu des dossiers	1
a.	Les bases SAS/CSV et les fichiers de transports	1
b.	La documentation des bases administratives	2
B-	Quelques évolutions entre la SAE 2022 et la SAE 2023	4
C-	Le champ de l'enquête, les tables d'identification et la localisation des établissements	5
a.	Le champ de l'enquête	5
b.	L'identification des établissements de santé	5
c.	Localisation des données collectées	6
D-	Informations sur les tables composant la base administrative	6
a.	Structure des tables présentes dans la base	6
b.	Limites de l'utilisation des données d'activités importées depuis le PMSI	7
	ANNEXE 1 – Champ et mode d'interrogation de l'enquête SAE	8
	ANNEXE 2 – La refonte de l'enquête (SAE 2013)	10

A- Accès aux données et contenu des dossiers

Pour consulter la base administrative SAE 2023, il faut dé-zipper le dossier contenant les tables. Pour cela, il faut cliquer sur *SAE 2023 – Base administrative.zip*, et enregistrer le fichier dans un répertoire. Il faut ensuite double-cliquer sur ce dossier, et cliquer sur « Extraire ».

Une fois dé-zippé, le répertoire « SAE 2023 Base Administrative » contient quatre sous-dossiers :

- un dossier « Base SAS » avec les différentes tables au format SAS ;
- un dossier « Base CSV » avec les différentes tables au format CSV ;
- un dossier « Documentation » ;
- un dossier « Fichier de transport – bases SAS ».



The screenshot shows a Windows File Explorer window with the following details:

- Toolbar: Ajouter, Extraire, Tester, Copier, Déplacer, Supprimer, Informations.
- Address bar: T:\SAE\Production_Bes\2 - Diffusion\1- Diffusion des bases brutes (data.drees)\Administrati
- Table of contents:

Nom	Taille
Base CSV	29 609 848
Base SAS	98 083 840
Documentation	6 680 655
Fichier de transport - bases SAS	48 944 400
A LIRE - Documentation SAE 2022 administrative.pdf	1 050 177

a. Les bases SAS/CSV et les fichiers de transports

Chacun des dossiers « Base SAS » et « Base CSV » contient l'intégralité des bases. Les fichiers .csv peuvent être ouverts avec Excel, mais aussi avec les tableurs des logiciels libres (OpenOffice ou

LibreOffice par exemple), le logiciel de données SAS ou les langages de programmation R ou Python. Les fichiers SAS (.sas7bdat) peuvent être ouverts avec le logiciel SAS ou à l'aide des langages de programmation R ou Python.

Fichiers de transports

Les tables et formats SAS sont créés en 32bits à la DREES. Dans le cas où votre système d'exploitation (et SAS) est en 64bits, cela empêche l'utilisation des formats. Pour pallier ce problème, vous trouverez dans le répertoire « SAE 2023 Base Administrative » (dossier « Fichier de transport – bases SAS ») les *fichiers de transports* pour les bases administratives de la SAE 2023 qui devraient vous permettre d'utiliser les formats. Vous trouverez la procédure d'utilisation à cette adresse : <http://support.sas.com/kb/44/047.html>.

Affichage des formats sous SAS

Lors de l'utilisation des bases SAS, vous pouvez choisir d'afficher les tables avec ou sans formats.

Pour utiliser les formats de la SAE, il faut copier ces lignes de code au début de votre programme SAS :

```
libname fSAE "chemin du dossier où se trouve formats.sas7bcat";  
options fmtsearch= (fSAE);
```

Si lors d'une instruction vous ne voulez pas utiliser les formats, il faut ajouter à votre procédure l'instruction suivante : `format _all_;`

Par exemple, pour créer une version de table sans formats :

```
data table_sans_formats;  
set table_avec_formats;  
format _all_;  
run;
```

Si vous ne voulez pas du tout utiliser les formats, il faut ajouter cette ligne de code : `options nofmterr;`

b. La documentation des bases administratives

Le dossier « Documentation » contient :

- le fichier *SAE2023_Dictionnaire_des_variables_par_bordereaux.xlsx* : il établit, pour chaque bordereau de la SAE, une correspondance entre le visuel du questionnaire et les variables des fichiers CSV ou SAS :
 - o l'onglet **Recap TABLES** recense l'ensemble des tables diffusées avec leur bordereau correspondant ;
 - o l'onglet **Légende** est une aide à la lecture ;
 - o les onglets **ID1** à **Q24** indiquent, case par case, le nom de chaque variable pour chacun des bordereaux de l'enquête ;
 - o l'onglet **SYGEN** reprend, case par case, le nom de chacune des variables de la table. La table SYGEN correspond à la fiche de synthèse « Activité générale et activités autorisées » (synthèse des bordereaux). Cet onglet sert aussi de correspondance pour la base HPR (Hôpitaux de proximité) ;
 - o l'onglet **GHT** donne la liste des variables présentes dans la base agrégée au niveau groupement hospitalier de territoire (GHT).

- le fichier *Liste Variables SAE 2023.xlsx* : il reprend les informations du fichier précédent (*SAE2021_Dictionnaire_variables_par_bordereaux.xlsx*) mais sous forme de tableau. Il contient, pour chaque table SAS, la liste des variables ainsi que leur libellé (et éventuellement la modalité associée), et la case du questionnaire à laquelle elle correspond. Pour certaines variables, il n'y a pas de correspondance entre celles-ci et une case du questionnaire : il s'agit de variables créées a posteriori. Elles sont repérées par l'indication « (Var SAS) » dans la colonne Bordereau ;
- un dossier « Documentation collecte SAE 2023 » contenant plusieurs éléments d'information sur la collecte de la SAE 2023 :
 - l'aide au remplissage de l'enquête (*SAE_AR_2023.pdf*). Ce document est fourni aux établissements enquêtés au moment de la collecte. Il contient des informations sur les concepts ainsi que des consignes pour la réponse à l'enquête ;
 - les nomenclatures de l'enquête (*Nomenclatures_2023.xls*), incluant les nomenclatures spécifiques à la diffusion. Ce document est également fourni aux établissements enquêtés au moment de la collecte ;
 - l'image du questionnaire de collecte (*Questionnaire SAE 2023.xlsx*).

Pour information, depuis la SAE 2016, les données ne sont plus diffusées au format ACCESS, l'objectif étant de s'orienter vers des formats ouverts (CSV), tout en conservant les formats SAS.

Vous pourrez trouver d'autres informations concernant la SAE sur :

- le site internet de la DREES : <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/00-la-statistique-annuelle-des-etablissements-sae>
- la FAQ de la SAE : [Accueil | Aides et FAQ enquête SAE \(sante.gouv.fr\)](#)

Pour toute question, vous pouvez nous écrire à l'adresse : DREES-INFOS@sante.gouv.fr

B- Quelques évolutions entre la SAE 2022 et la SAE 2023

Pour la SAE 2023, le questionnaire de l'enquête a peu évolué pour assurer une stabilité. Cependant, quelques changements ont été apportés, dont les principaux sont recensés ci-dessous :

Ajout et suppression de bordereaux

- Pour la SAE 2023, c'est le bordereau **PALIA** qui est proposé au lieu du bordereau **DOULEUR**, puisque ces bordereaux sont proposés en alternance une année sur deux.
- Depuis la SAE 2019, les deux bordereaux **PHARMA** et **BIO** sont interrogés alternativement afin de poursuivre l'allègement de la charge de collecte. Pour la SAE 2023, le bordereau **BIO** est collecté.

C- Le champ de l'enquête, les tables d'identification et la localisation des établissements

a. Le champ de l'enquête

Le champ de la SAE est constitué des établissements de santé qui ont exercé en 2023 une activité de soins autorisée. Tous les établissements ne sont pas interrogés directement : dans certaines entités juridiques (EJ), un établissement géographique est désigné pour remplir le questionnaire SAE pour lui-même et pour d'autres établissements. C'est le cas des établissements de santé mentale et des unités de soins longue durée (USLD). Pour les établissements de santé mentale, pour une EJ donnée, un établissement par département est désigné pour répondre pour l'ensemble des structures de psychiatrie de son EJ situées dans son département. Pour les USLD, ceux ayant la même adresse qu'un autre établissement géographique interrogé dans la SAE ne reçoivent pas de questionnaires et leur réponse est comprise dans le questionnaire de l'autre établissement.

En tout, 3 975 établissements géographiques sont interrogés directement dans la SAE 2023.

Le recueil de la SAE s'effectue sur un site de collecte en ligne. Cette enquête permet de récupérer des informations sur les capacités, l'activité, les plateaux techniques et le personnel des établissements de santé. Le taux de réponse de la SAE 2023 (pourcentage d'établissements enquêtés ayant transmis et validé leurs données) est de 99,5 %.

Pour plus d'information sur le champ et le mode d'interrogation de l'enquête SAE, vous pouvez vous reporter à l'annexe 1.

b. L'identification des établissements de santé

Dans le questionnaire et sur le site de collecte, deux bordereaux contiennent des données d'identification sur les établissements de santé : les bordereaux ID1 et ID2. Ces informations sont en partie importées du répertoire Finess.

Dans les bases CSV et SAS fournies, deux tables, réorganisées par rapport aux bordereaux du questionnaire, contiennent les données d'identification des établissements :

- La table **ID_2023** contient les données d'identification de l'établissement (raison sociale, statut juridique, catégorie d'établissement, etc.) issues des bordereaux ID1 et ID2, et des données propres à la collecte de l'enquête (vague d'interrogation, statut de validation) pour les établissements directement interrogés par la SAE.
- La table **IDB_2023** contient les données d'identification pour les établissements non interrogés et non interrogés directement.

Attention : un incendie au CHU de Pointe-à-Pitre/Abîmes en Guadeloupe fin 2017 a endommagé une partie des bâtiments, entraînant le déménagement temporaire des services impactés vers d'autres établissements. Afin de pouvoir séparer l'activité du CHU de ces autres établissements, des Finess « provisoires » ont été créés. Dans les SAE 2018 et 2019, les données du CHU sont réparties sur 3 Finess cette année : deux Finess « provisoires » (970112835 (Polyclinique de Guadeloupe - site CHU) et 970112850 (Clinique les eaux claires – site CHU)) et le Finess 970100442 (CHU de Pointe-à-Pitre/Abîmes). Depuis la SAE 2020, les données remontent de nouveau uniquement sur le Finess géographique du CHU de Pointe-à-Pitre/Abîmes.

De plus, dans le questionnaire, deux autres bordereaux permettent d'identifier les structures d'interrogation au niveau EJ et établissements :

- La table **ST_2023** contient tous les liens entre les entités juridiques recevant les courriers de l'enquête et leurs établissements géographiques ayant un questionnaire associé
- La table **STB_2023** contient tous les établissements en psychiatrie ou en USLD qui sont inclus dans la réponse de l'établissement de la même EJ ayant reçu le questionnaire.

c. Localisation des données collectées

Dans le cas particulier des établissements privés ne faisant que de la psychiatrie et tous situés dans le même département, c'est l'entité juridique (EJ) de ces établissements qui est interrogée et qui répond pour l'ensemble des établissements de psychiatrie qui lui sont rattachés. Dans certains cas, cette EJ n'est pas située dans le même département ou dans la même région que les établissements qui lui sont rattachés.

Au final, les données collectées sont localisées sur le lieu d'implantation de l'établissement répondant, y compris lorsque la réponse est collectée au niveau de l'EJ.

Depuis 2016, deux nouvelles variables (REG_DIFF et DEP_DIFF) sont disponibles dans la table ID diffusées sur Data.Drees, permettant de rendre compte du lieu réel de l'activité dans les cas particuliers où établissements et EJ ne sont pas localisés au même endroit.

D- Informations sur les tables composant la base administrative

a. Structure des tables présentes dans la base

Les tables de la SAE sont créées de telle sorte qu'une ligne correspond aux données d'un établissement dans son ensemble.

Cependant, dans le cas de bordereaux contenant des tableaux de grande taille, afin de faciliter la lecture des tables et d'éviter d'avoir trop de variables, une variable de la table correspondra à plusieurs cases du questionnaire : on crée alors une variable « modalité » qui distingue chaque case concernée. Ainsi, une ligne ne correspondra pas aux données d'un établissement dans son ensemble mais à un croisement « établissement » x « modalité ».

Exemples :

- dans la table concernant la psychiatrie (bordereau PSY), on distingue la psychiatrie générale, la psychiatrie infanto-juvénile et la psychiatrie pénitentiaire à l'aide de la variable « DIS » ;
- dans les tables de personnel rémunéré par les établissements, la distinction porte sur le code « métier » (bordereaux Q20 et Q23) avec la variable « PERSO » ou sur la discipline d'équipement (bordereaux Q21 et Q24) avec la variable « DISCI » ;
- dans la table sur les urgences (bordereau URGENCES), on distingue les urgences générales et pédiatriques à l'aide de la variable « URG ».

Les différentes modalités sont rappelées dans l'onglet « Recap TABLES » du dictionnaire des variables.

Pour ces tables, il existe donc différentes manières d'agrégier les données : somme sur tous les établissements par sous-modalité, somme de toutes les sous-modalités pour chaque établissement, etc.

Néanmoins, il faut tout de même faire attention à certaines tables dans lesquelles il existe une sous-modalités « Total » ou des sous-totaux (par exemple dans les tables PSY ou SSR). Dans ce cas, il faut choisir entre deux options : utiliser l'agrégat déjà calculé ou recalculer le total en excluant les totaux et sous-totaux existants.

Exemple : dans la table concernant la psychiatrie (bordereau PSY), la variable « DIS » (discipline de psychiatrie) peut prendre quatre modalités : *GEN* (psychiatrie générale), *INF* (psychiatrie infanto-juvénile), *PEN* (psychiatrie en milieu pénitentiaire) et *TOT* (total psychiatrie).

La modalité *TOT* est directement calculée sur le site de collecte et correspond à la somme *GEN+INF+PEN*. Pour éviter les doubles-comptes dans les agrégats, il faut donc choisir entre utiliser uniquement la modalité *TOT* ou utiliser uniquement les trois autres modalités, *GEN/INF/PEN*.

De plus, les données de personnel présents au sein des bordereaux d'activité (personnel concourant à l'activité ou PCA) sont diffusées dans des tables spécifiques : elles sont suffixées par la lettre « _p », excepté pour les bordereaux PSY, REA et CANCERO.

b. Limites de l'utilisation des données d'activités importées depuis le PMSI

Pour certains bordereaux, les cases portant sur l'activité des établissements sont pré-remplies à partir d'une source auxiliaire, le PMSI¹ (voir annexe 2) produit par l'ATIH². Dans le cadre du programme « Dites-le nous une fois » mené au sein du service statistique publique, ces données sont diffusées uniquement à but de contrôle, pour assurer la cohérence avec les données sur les capacités et sur le personnel.

Les données d'activité de la SAE pré-remplies avec le PMSI ne doivent donc pas être utilisées pour suivre l'activité des établissements de santé. La DREES ne les mobilise jamais pour ses propres publications.

La liste des cases pré-remplies est disponible dans l'aide au remplissage de l'enquête SAE (*SAE_AR_2023.pdf*).

¹ Programme de Médicalisation des Systèmes d'Information.

² Agence Technique de l'Information sur l'Hospitalisation.

ANNEXE 1 – Champ et mode d’interrogation de l’enquête SAE

A) Champ

La SAE concerne tous les établissements de santé publics et privés, disposant d’une autorisation pour une activité de soins, installés en France (métropole et DROM), y compris les structures qui ne font qu’un seul type d’hospitalisation (exemple : hospitalisation à temps partiel, hospitalisation à domicile). Sont également inclus les services pénitentiaires des établissements de santé et les secteurs militaires des établissements de santé. Entrent dans le champ tous les établissements ayant une activité de soins au cours de l’année, y compris ceux créés ou fermés au cours de l’année.

B) Entité interrogée

La nature de l’entité interrogée (établissement géographique – ET – ou entité juridique – EJ) dépend du statut et de la catégorie de l’établissement. Sont distingués 10 groupes d’entités interrogées directement et 8 groupes d’entités qui ne sont pas interrogées directement.

Un même établissement géographique ne peut pas figurer dans plusieurs groupes.

1. SECTEUR PUBLIC (statut juridique < 40)

Principe général :

Les établissements géographiques sont interrogés (un questionnaire par numéro FINESS).

Les établissements se répartissent en 5 groupes :

- **Groupe 03** : Établissements des trois AP (EJ 750712184 : APHP, 690781810 : HCL et 130786049 : APHM)
- **Groupe 13** : Établissements du Service de santé des armées (EJ 750821092)
- **Groupe 10** : Établissements sanitaires publics non rattachés à une EJ publique sanitaire
- **Groupe 14** : Établissements de santé mentale publics (voir ci-dessous)
- **Groupe 04** : Autres établissements publics (hors établissements de santé mentale)

Cas particuliers :

Les établissements de santé mentale sont des entités très hétérogènes. Par conséquent, nous n’interrogeons pas directement tous ces établissements. Au sein d’une entité juridique publique, un établissement géographique par département a été désigné pour recevoir un questionnaire. Cet établissement, classé en **groupe 14**, répond pour lui-même et pour tous les autres établissements de santé mentale se trouvant dans le même département au sein de la même entité juridique. Ces établissements non interrogés directement sont alors classés en **groupe 24**.

Par ailleurs, au sein d’une même entité juridique, lorsque toutes les structures psychiatriques sont ambulatoires (CMP, CATTP etc...) ou lorsque le mode d’organisation le justifie, l’ensemble des établissements psychiatriques sont classés en groupe 24 (il n’y a donc plus d’établissement en groupe 14) et rattachés à un établissement de groupe 04.

Lors de la SAE 2013, les USLD publiques se situant dans les mêmes locaux qu’un autre établissement interrogé ne remplissaient pas de questionnaire SAE propre. À partir de la SAE 2014, il a été décidé de créer un groupe pour ces ET non interrogés : il s’agit du **groupe 34** « USLD situés à la même adresse qu’un ET public » (rattachés à un ET de groupe 04).

La liste des établissements interrogés (groupe 03, 13, 10, 14, 04) se trouve dans le bordereau ID alors que celle des groupes 24 et 34 (non interrogés directement) est dans le bordereau IDB.

2. SECTEUR PRIVE (code statut juridique >= 40)

Principe général :

Les établissements géographiques sont interrogés.

Les établissements privés interrogés directement, hors cas particuliers décrits ci-dessous, constituent le **groupe 09**.

Cas particuliers :

Radiothérapie

C'est l'établissement géographique dans lequel est implanté l'équipement qui répond directement pour son activité de radiothérapie.

S'il n'a qu'une activité de radiothérapie ou si c'est un centre de lutte contre le cancer (CLCC – catégorie 131), il fait partie du **groupe 18**. S'il a une autre activité décrite dans la SAE, il fait alors partie du **groupe 09**.

Établissements privés de santé mentale

Pour les établissements privés de santé mentale, il existe trois cas :

- L'entité juridique n'a qu'un seul établissement géographique de santé mentale. L'établissement géographique est alors classé en **groupe 15** et sera interrogé directement.
- L'entité juridique a plusieurs établissements géographiques de santé mentale, tous situés dans le même département (hors structures ambulatoires, CMP et CATT). **L'entité juridique** est alors classée en **groupe 05** et remplira un questionnaire unique pour l'ensemble de ses établissements de santé mentale. Les établissements géographiques qui lui sont rattachés (et donc non interrogés directement) sont classés en **groupe 25**.
- L'entité juridique a plusieurs établissements géographiques de santé mentale, situés dans des départements différents. Dans ce cas, un établissement géographique est choisi par département. Ces établissements seront classés en **groupe 15** et répondront pour eux-mêmes et les autres établissements géographiques de santé mentale situés dans le même département qu'eux (et appartenant à la même EJ). Ces établissements géographiques non interrogés directement sont classés en **groupe 35**.

Dialyse

Les établissements interrogés sont les établissements de dialyse ayant une autorisation (catégories 138, 139, 140, 141, 146, ou 422). Chaque centre de dialyse remplit un questionnaire sous son propre numéro FINISS. Ils constituent le **groupe 07**.

USLD

À partir de la SAE 2014, il a été décidé de généraliser au secteur privé le mode d'interrogation des USLD du secteur public. En conséquence, les USLD privées situées à la même adresse qu'un autre établissement interrogé sont classées en **groupe 39** « ET USLD situés à la même adresse qu'un ET privé » (rattachés à un ET de groupe 09), et ne sont pas interrogées directement.

La liste des entités interrogés (groupe 09, 05, 07, 15, 18) se trouve dans le bordereau ID alors que celle des groupes 25, 35 et 39 (non interrogés directement) est dans le bordereau IDB.

ANNEXE 2 – La refonte de l'enquête (SAE 2013)

La SAE a été rénovée en profondeur pour la collecte des données 2013 intervenue en 2014. Trois principes ont présidé à cette refonte :

- simplification du niveau d'interrogation (désormais au niveau établissement géographique pour le secteur public comme pour le secteur privé, exception faite de la psychiatrie) ;
- actualisation du questionnaire (avec mise en place d'un questionnaire filtre permettant de soumettre à chaque établissement uniquement les bordereaux qui le concernent) ;
- mise en cohérence avec d'autres sources de données sur les établissements de santé, dont le programme de médicalisation des systèmes d'information (PMSI).

En effet, l'objectif de la refonte est aussi de s'appuyer sur d'autres recueils réglementaires (notamment le PMSI) pour alléger la charge de collecte des établissements. Ainsi, la SAE 2013 a bénéficié du pré-remplissage par des données d'activités issues du PMSI MCO et HAD (en collaboration avec l'ATIH) de certaines variables jusque-là déclarées par les établissements. À partir de la SAE 2016, les données d'activité du bordereau SSR ont également été pré-remplies avec les données issues du PMSI SSR.

L'ampleur de la refonte induit un profond changement du format des données obtenues : les données concernent désormais les entités géographiques et les données au niveau entité juridique doivent donc être calculées, les noms et formes des bordereaux ont changé, les variables également. En pratique, les principaux impacts sur l'exploitation des bases sont :

1. Les changements dans le questionnaire

Les questions posées dans la SAE 2013 étaient en grande partie nouvelles. Dans le dossier de documentation de la SAE 2013 se trouvent les fichiers **Liste Variables.xlsx** (liste exhaustive des variables ainsi que leur label) et **Correspondance principales variables SAE 2012.pdf** (qui permet de retrouver, pour certaines variables de la SAE 2012, les données qui sont sur le même thème dans la SAE rénovée).

Cette refonte a introduit pour de nombreuses variables une rupture de série en 2013, qui complique l'analyse des évolutions entre les années antérieures et postérieures à 2013 : celles-ci doivent donc être interprétées avec précaution.

2. Les changements dans le mode d'interrogation

Le champ de l'enquête est resté identique entre 2012 et 2013, mais le niveau d'interrogation a changé. Cette modification simplifie grandement l'exploitation des bases de la SAE. En effet, il n'y a plus de double-compte dans la SAE rénovée : pour exploiter les bases jusqu'à la SAE 2012 incluse, il fallait supprimer certains groupes d'interrogation pour ne pas compter les données en double.

Pour les SAE 2013 et suivantes, cela n'est plus nécessaire : les résultats nationaux peuvent être obtenus par simple agrégation des données de chacun des établissements interrogés.

3. Les imports PMSI

Dans les SAE 2013 et suivantes, une partie des variables d'activité sont directement importées du PMSI. Parmi ces variables importées, certaines sont non-modifiables par l'établissement (par exemple, le nombre de séjours de médecine dans le bordereau MCO) et certaines sont modifiables par l'établissement, en raison d'un problème de couverture du champ (dialyse, radiothérapie), d'une mauvaise répartition par autorisation (SSR) ou de prestations inter-établissements (exemple : les actes dans le bordereau BLOCS). Dans le fichier **Documentation collecte SAE 2023\Questionnaire SAE 2023.xlsx**, les cases contenant des variables PMSI non modifiables sont codées comme suit :

Pré-remplissage de données PMSI –
modifiable par l'établissement



Pré-remplissage de données PMSI -
non modifiable

