

Base administrative SAE 2018 – Documentation

Sommaire :

A- Accès aux données et contenu des dossiers	1
B- Exploitation de la base SAE 2018	3
a. Le champ de l'enquête et les tables d'identification	3
b. Spécificités d'exploitation de certaines tables	4
c. Fiche de synthèse « Activité générale et activités autorisées »	4
d. Base de données agrégées au niveau des GHT	4
C- Quelques évolutions entre la SAE 2017 et la SAE 2018	4
ANNEXE 1 – Champ et mode d'interrogation de l'enquête SAE	6
ANNEXE 2 – La refonte de l'enquête (SAE 2013)	8

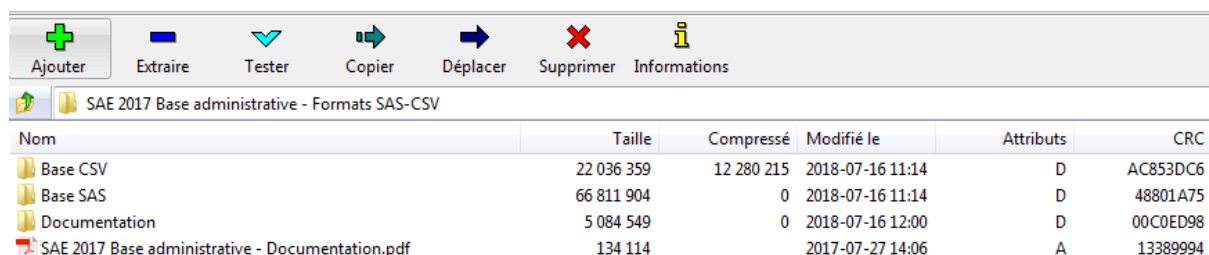
A- Accès aux données et contenu des dossiers

Pour consulter la base administrative SAE 2018, il faut dé-zipper le dossier contenant les tables. Pour cela, il faut cliquer sur *SAE 2018 – 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 2018 Base Administrative » contient trois 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 ».

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 de type OpenOffice ou LibreOffice et les logiciels de données SAS, R ou Python. Les fichiers SAS (.sas7bdat) peuvent être ouverts avec les logiciels SAS, R ou Python.



Nom	Taille	Compressé	Modifié le	Attributs	CRC
Base CSV	22 036 359	12 280 215	2018-07-16 11:14	D	AC853DC6
Base SAS	66 811 904	0	2018-07-16 11:14	D	48801A75
Documentation	5 084 549	0	2018-07-16 12:00	D	00C0ED98
SAE 2017 Base administrative - Documentation.pdf	134 114		2017-07-27 14:06	A	13389994

Le dossier « Documentation » contient :

- le fichier *SAE2018_Dictionnaire_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 :
 - l'onglet **Recap TABLES** recense l'ensemble des tables diffusées avec leur bordereau correspondant ;
 - l'onglet **Légende** est une aide à la lecture ;

- les onglets **ID1** à **Q24** indiquent les noms de chaque variable, case par case, pour chacun des bordereaux de l'enquête ;
 - l'onglet **SYGEN** indique le nom de chaque variable de la table, case par case. La table SYGEN correspond à la fiche de synthèse « Activité générale et activités autorisées » (synthèse des bordereaux) ;
 - 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 2018.xlsx* : il reprend les informations du fichier précédent (*SAE2018_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 celle-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 2018 » contenant plusieurs éléments d'information sur la collecte de la SAE 2018 :
 - l'aide au remplissage de l'enquête (*SAE_AR_2018.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_2018.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 2018 et fiche de synthèse.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 la page de la DREES :

<http://drees.solidarites-sante.gouv.fr/etudes-et-statistiques/open-data/etablissements-de-sante-sociaux-et-medico-sociaux/article/la-statistique-annuelle-des-etablissements-sae>.

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

B- Spécificité de l'exploitation de la base 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;
```

a. Le champ de l'enquête et les tables d'identification

Le champ de la SAE est constitué des établissements de santé qui ont exercé en 2018 une activité de soins autorisée.

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 2018 est de 99,6 %.

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_2018** 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_2018** contient les données d'identification pour les établissements non interrogés et non interrogés directement

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

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. Les données du CHU sont donc réparties sur 3 Finess au lieu d'un seul : les Finess 970100442 (CHU de Pointe-à-Pitre/Abîmes), 970112835 (Polyclinique de Guadeloupe - site CHU) et 970112850 (Clinique les eaux claires – site CHU).

b. Spécificités d'exploitation de certaines tables

Dans certaines tables, il existe des sous-modalités ; une ligne ne correspond donc pas aux données d'un établissement dans son ensemble mais à un croisement « établissement » x « sous-modalité ». C'est le cas notamment dans la table concernant la psychiatrie (distinction psychiatrie générale/infanto-juvénile/pénitentiaire), les tables de personnel (distinction selon le code « métier ») ou encore dans la table des urgences (distinction urgences générales/pédiatrique). Les différentes modalités sont rappelées dans l'onglet « Recap TABLES » du dictionnaire des variables.

Il y a donc différentes manières d'agrèger les données dans ces tables : somme sur tous les établissements par sous-modalité, somme de toutes les sous-modalités pour chaque établissement, etc.

Pour les agrégats des sous-modalités, il faut tout de même faire attention à certaines tables dans lesquelles il existe un sous-total, par exemple dans les tables PSY ou SSR. Dans ce cas, il faut alors choisir entre deux options : utiliser l'agrégat déjà calculé, ou recalculer le total en excluant les sous-totaux existants.

Par exemple, dans la table 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 les trois autres modalités, GEN/INF/PEN.

c. Fiche de synthèse « Activité générale et activités autorisées »

La fiche de synthèse « Activité générale et activités autorisées » est une synthèse des données structurantes des bordereaux. Elle est alimentée par les informations présentes dans les différents bordereaux de la SAE (MCO, psychiatrie, urgences, etc.).

Depuis la SAE 2016, afin d'éviter aux utilisateurs d'avoir à rechercher les informations de cette synthèse dans les tables des différents bordereaux, un fichier spécifique contenant toutes ces informations a été créé : il s'agit de la table SYGEN_2018.

d. Base de données agrégées au niveau des GHT

Depuis la SAE 2016, une base de données agrégées sur les contours des groupements hospitaliers de territoire (GHT) au 31 décembre a été créée (cette année : table GHT_2018). Cette table contient des variables pertinentes pour l'analyse au niveau GHT, telles que les lits, places, journées, séjours, informations sur les équipements, les plateaux techniques et le personnel des établissements de santé.

C- Quelques évolutions entre la SAE 2017 et la SAE 2018

Pour la SAE 2018, 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 :

- Afin d'alléger la charge de collecte, les questions sur le nombre de conseils de secteur de **Psychiatrie** et sur les ETP budgétés du bordereau **Q20** sont supprimées. Les modalités « brancardiers » et « assistants de service social » sont également supprimées de la liste des professionnels concourant à l'activité de SAMU du bordereau **SMURSAMU**.
- Pour la SAE 2018, c'est le bordereau **DOULEUR** qui est interrogé au lieu du bordereau **PALIA**, puisque ces bordereaux sont proposés en alternance une année sur deux.
- L'activité concernant les bébés restés auprès de leur mère est dorénavant recueillie dans le bordereau **PERINAT** afin de mieux répondre à la demande européenne de mesurer cette activité. Elle est pré-remplie à partir du PMSI. L'algorithme pour mieux repérer les bébés restés auprès de leur mère a également été amélioré. Pour rappel, l'activité générée par ces bébés n'est pas comptabilisée dans le bordereau MCO.
- Les questions devenues non informatives ont été supprimées du questionnaire pour l'alléger. C'est le cas des questions portant sur le contrat de bon usage du médicament dans le bordereau **PHARMA** et sur la participation de la médecine libérale à la régulation dans le cadre de la PDSA du bordereau **SMURSAMU**.
- Pour éviter une confusion lors du remplissage du bordereau **URGENCES**, les questions portant sur l'accueil des malades pour une spécialité donnée sont supprimées.
- Les questions sur les coopérations internationales ajoutées l'année dernière dans le bordereau Suivi des politiques nationales et missions de service public (**PN**) sont enlevées cette année, car la SAE n'est pas le bon vecteur pour collecter cette information.
- La variable sur les territoires de santé (TSANT) dans les tables **ID** et **IDB** a été supprimée.

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 FINES).

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 établissements 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'identification (HAD) ou de prestations inter-établissements (exemple : les actes dans le bordereau BLOCS). Dans le fichier **Documentation collecte SAE 2018\Questionnaire SAE 2018 et fiche de synthèse.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 