

## Les bases statistiques SAE 2019 – Documentation

### Sommaire :

A- Accès aux données et contenu des dossiers .....	1
B- Contenu des corrections .....	3
C- Ruptures de séries.....	4
D- Tables fournies .....	4
a- Données d'activité et de capacités.....	4
b- Données sur le personnel médical .....	6
c- Données sur le personnel non médical .....	7
d- Nouvelles bases mises à disposition .....	7

### A- Accès aux données et contenu des dossiers

Pour consulter les bases statistiques SAE 2019, il faut dé-zipper le dossier contenant les tables. Pour cela, il faut cliquer sur *SAE 2019 – Bases statistiques.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 2019 Bases statistiques » contient 4 sous-répertoires :

- un sous-répertoire « Bases statistiques » ;
- un sous-répertoire « Séries longues » ;
- un sous-répertoire « Fichier de transport – bases SAS » ;
- un sous-répertoire « Documentation ».

**Le sous-répertoire « Bases statistiques »** contient deux dossiers :

- un dossier « Base SAS » avec les différentes tables redressées au format SAS ;
- un dossier « Base CSV » avec les différentes tables redressées au format CSV.

Chacun de ces dossiers contient l'intégralité des bases (dans deux formats différents). Lorsqu'elles sont suffixées par la lettre « a », en fin de nom de fichier, cela indique que la base statistique est identique à la base administrative, contrairement au cas où elles sont suffixées par la lettre « r », qui indique que la base statistique est redressée.

**Le sous-répertoire « Séries longues »** contient les tables de séries longues, proposées là encore dans deux formats différents, avec les deux dossiers suivants :

- un dossier « Base SAS » avec les tables de séries longues au format SAS ;
- un dossier « Base CSV » avec les tables de séries longues au format CSV.

Chacun de ces dossiers contient (dans deux formats différents) :

- trois tables contenant des données structurantes d'activité et de capacité (CAPACT19 pour l'année 2019, CAPACT0019 et CAPACT\_TOT0019 pour la période 2000-2019) ;

- deux tables contenant des informations sur les effectifs et ETP de personnels rémunérés pour la période 2001-2019 (PM\_01\_19 pour le personnel médical et PNM\_01\_19 pour le personnel non médical).

**Le sous-répertoire « Fichier de transport – bases SAS »** contient les *fichiers de transport* pour les bases statistiques de la SAE 2019 ainsi que pour les tables de séries longues.

Les tables et formats SAS sont en effet 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. Les *fichiers de transport* devraient vous permettre de pallier ce problème et d'utiliser les formats. Vous trouverez la procédure d'utilisation à cette adresse : <http://support.sas.com/kb/44/047.html>.

**Le sous-répertoire « Documentation »** contient :

- les formats SAS ;
- l'aide au remplissage de l'enquête (*SAE\_AR\_2019.pdf*) ;
- les nomenclatures de l'enquête (*Nomenclatures\_2019.xls*) ;
- le fichier *SAE2019\_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 ;
- le fichier *Liste Variables.xlsx* : il reprend les informations du fichier précédent (*SAE2019\_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 ;
- le fichier *Liste GHT\_2019.xlsx* : il établit pour chaque GHT la liste des entités juridiques qui le compose ;
- ce document.

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](mailto:drees-sae@sante.gouv.fr)

## B- Contenu des corrections

Les corrections de la SAE 2019 sont de deux ordres : harmonisation et redressement des données.

**L'harmonisation des données** concerne la mise en cohérence des données entre bordereaux (cohérence entre les bordereaux Q20/Q21 et Q23/Q24 particulièrement, dont les ETP salariés et les effectifs de libéraux doivent être égaux), ainsi que l'harmonisation des autorisations avec le référentiel ARHGOS (notamment pour les autorisations des maternités, des urgences, en soin de suite et réadaptation, etc.).

**Le redressement des données** : contrairement aux bases administratives, qui doivent retranscrire exactement les réponses des établissements géographiques à la SAE telles que validées par eux suite à la phase de validation des données collectées, les bases statistiques permettent de réaliser, si nécessaire, des corrections de valeurs aberrantes, manquantes ou des corrections d'incohérences qui n'auraient pu être traitées pendant la phase de validation de la collecte.

À compter de la SAE 2019, nous apportons également une correction de la non réponse partielle et totale. Les établissements qui n'ont pas validé leur questionnaire mais ont renseigné une partie des données (statut de l'enquête « en cours de saisie ») ainsi que ceux n'ayant pas saisi leur SAE sont ajoutés aux bases statistiques.

Pour les établissements « en cours de saisie », les données complétées par l'établissement sont récupérées et intégrées aux bases statistiques, et les quelques données manquantes (notamment les capacités et les données de personnel) sont ajoutées en utilisant les données N-1. Pour les établissements n'ayant pas saisi leur SAE ou l'ayant validée mais n'ayant pas rempli un bordereau qu'ils auraient dû remplir (parce qu'ils ont l'autorisation par exemple), nous récupérons les données PMSI pour l'activité et nous imputons les données manquantes en utilisant, le plus souvent, les données de l'année précédente (pour les bordereaux et variables les plus importants). Par la suite, la dénomination « non réponse totale par bordereau » correspond à la correction de cette non réponse totale ou partielle.

Après intégration de ces établissements, le taux de couverture de l'enquête est de 99,8 % (contre 99,6 % de taux de réponse en base administrative).

**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).

Dans la SAE, seuls les établissements sanitaires détenant au moins une autorisation de soins sont interrogés, ce qui exclut les « GCS de moyen exploitants », qui exploitent les autorisations de soins de leurs membres mais ne les détiennent pas.

Exceptionnellement, et au vu de sa spécificité, nous avons décidé d'ajouter au champ de la SAE 2019 le GCS de moyen exploitant ICANS (670020098). En effet, ce GCS réalise son activité dans des locaux en propre. De plus, il n'est pas possible de répartir l'activité entre les établissements membres du GCS

– CLCC Paul Strauss (670000033) et HU Strasbourg (670018787) –, les capacités ayant notamment été ajustées et regroupées sans possibilité de distinction entre les établissements.

### C- Ruptures de séries

La SAE a subi une refonte très importante en 2013, sur laquelle vous trouverez des informations sur le site de la DREES ou dans la documentation relative aux bases administratives. **En raison de cette refonte, les évolutions 2012-2013 sont difficiles à analyser.**

Pour faciliter l'exploitation des données, nous avons malgré tout mis à disposition des tables qui regroupent les années 2000 à 2019. Cela ne signifie pas que l'évolution annuelle est interprétable : sur un grand nombre de variables, les concepts ont changé. Ainsi, les données d'activité importées du PMSI ne se basent pas exactement sur les mêmes notions que l'activité qui était déclarée dans la SAE précédemment : les évolutions 2012-2013 restent difficiles à analyser.

**Attention** : à partir de 2011, les données incluent Mayotte.

### D- Tables fournies

#### a- Données d'activité et de capacités

##### Pour l'année 2019

À partir de la SAE 2019, les bordereaux PHARMA et BIO seront interrogés alternativement. Pour la SAE 2019, le bordereau PHARMA n'a pas été collecté. Il le sera dans la SAE 2020.

La table **CAPACT19** (disponible dans le sous-répertoire « Séries longues ») regroupe les données structurantes d'activité et de capacités, par discipline d'équipement, pour chaque établissement interrogé en 2019.

Les bases auxquelles des corrections ont été apportées, par rapport aux bases administratives, sont listées ci-dessous en précisant les modifications effectuées :

La table **ID\_2019R** : elle est au même format que la table ID\_2019 des bases administratives, mais elle contient des corrections sur la variable ETAT\_SAISIE ainsi que sur le champ.

La table **STB\_2019R** : elle est au même format que la table STB\_2019 des bases administratives, mais elle contient des corrections sur le champ.

La table **PN\_2019R** : elle est au même format que la table PN\_2019 des bases administratives, mais elle contient des corrections sur l'activité (prise en charge des personnes en situation de précarité) et le personnel concourant à l'activité des CMRR.

La table **FILTRE\_2019R** : elle est au même format que la table FILTRE\_2019 des bases administratives, mais elle contient des corrections sur les activités et équipements disponibles. Elle est également corrigée de la non-réponse totale au bordereau.

La table **MCO\_2019R** : elle est au même format que la table MCO\_2019 des bases administratives, mais elle contient des corrections sur les capacités, le nombre de journées

exploitables et sur l'activité. Elle est également corrigée de la non-réponse totale au bordereau.

La table **MCOGER\_2019R** : elle est au même format que la table MCOGER\_2019 des bases administratives, mais elle contient des corrections sur les capacités (lits et places) et sur l'activité.

La table **PSY\_2019R** : elle est au même format que la table PSY\_2019 des bases administratives, mais elle contient des corrections sur les capacités et l'activité (temps complet, partiel et ambulatoire).

La table **PSY2\_2019R** : elle est au même format que la table PSY2\_2019 des bases administratives, mais elle contient des corrections sur les SMPR et le nombre d'actes ambulatoire par mode légal de soins.

La table **SSR\_2019R** : elle est au même format que la table SSR\_2019 des bases administratives, mais elle contient des corrections sur les autorisations ainsi que sur le nombre de lits, places, journées et séjours. Elle est également corrigée de la non-réponse totale au bordereau.

La table **USLD\_2019R** : elle est au même format que la table USLD\_2019 des bases administratives, mais elle contient des corrections sur l'activité et les capacités suite à la correction de la non-réponse totale au bordereau.

La table **HAD\_2019R** : elle est au même format que la table HAD\_2019 des bases administratives, mais elle contient des corrections sur le nombre de patients pouvant être pris en charge simultanément par la structure HAD, de séjours et de journées ainsi que sur la file active, notamment du fait de la correction de la non-réponse totale au bordereau.

La table **BLOCS\_2019R** : elle est au même format que la table BLOCS\_2019 des bases administratives, mais elle contient des corrections sur le nombre de salles. Elle est également corrigée de la non-réponse totale au bordereau.

La table **BLOCS\_P\_2019R** : elle est au même format que la table BLOCS\_P\_2019 des bases administratives, mais elle contient des modifications sur les effectifs et ETP.

La table **IMAGES\_2019R** : elle est au même format que la table IMAGES\_2019 des bases administratives, mais elle contient des corrections sur le nombre d'équipement et l'activité, notamment du fait de la correction de la non-réponse totale au bordereau.

La table **BIO\_2019R** : elle est au même format que la table BIO\_2019 des bases administratives, mais elle contient des corrections sur les effectifs et ETP de personnel concourant à l'activité.

La table **TELEMED\_2019R** : elle est au même format que la table TELEMED\_2019 des bases administratives, mais elle est corrigée de la non-réponse totale au bordereau.

La table **PERINAT\_2019R** : elle est au même format que la table PERINAT\_2019 des bases administratives, mais elle contient notamment des corrections sur les types de maternités et les capacités. Elle est également corrigée de la non-réponse totale au bordereau.

La table **CHIRCAR\_2019R** : elle est au même format que la table CHIRCAR\_2019 des bases administratives, mais elle contient des corrections sur les autorisations et le champ.

La table **CHIRCAR\_P\_2019R** : elle est au même format que la table CHIRCAR\_P\_2019 des bases administratives, mais elle contient des modifications sur le champ.

La table **PALIA\_2019R** : elle est au même format que la table PALIA\_2019 des bases administratives, mais elle contient des corrections sur l'activité en USP et de l'EMSP et sur le personnel concourant à l'activité.

La table **DIALYSE\_2019R** : elle est au même format que la table DIALYSE\_2019 des bases administratives, mais elle contient des corrections sur les capacités et le nombre de séances notamment. Elle est également corrigée de la non-réponse totale au bordereau.

La table **CANCERO\_2019R** : elle est au même format que la table CANCERO\_2019 des bases administratives, mais elle contient des corrections sur l'activité et le champ. Elle est également corrigée de la non-réponse totale au bordereau.

La table **URGENCES\_2019R** : elle est au même format que la table URGENCES\_2019 des bases administratives, mais elle contient des corrections sur les autorisations.

La table **URGENCES2\_2019R** : elle est au même format que la table URGENCES2\_2019 des bases administratives, mais elle contient des corrections sur le nombre de passages aux urgences générales notamment.

La table **SMURSAMU\_2019R** : elle est au même format que la table SMURSAMU\_2019 des bases administratives, mais elle contient des corrections sur les autorisations SMUR et SAMU, sur l'activité de SMUR et de SAMU et sur les carences ambulancières. Elle est également corrigée de la non-réponse totale au bordereau.

La table **SMURSAMU\_P\_2019R** : elle est au même format que la table SMURSAMU\_P\_2019 des bases administratives, mais elle contient des corrections sur les effectifs et ETP de personnel concourant à l'activité (champ).

La table **REA\_2019R** : elle est au même format que la table REA\_2019 des bases administratives, mais elle contient des corrections sur les capacités et l'activité notamment. Elle est également corrigée de la non-réponse totale au bordereau.

#### Séries longues (disponibles dans le sous-répertoire « Séries longues »)

La table **CAPACT0019** contient, par discipline d'équipement, les données relatives à l'activité et aux capacités pour les années 2000 à 2019. Les données sont disponibles au niveau établissement géographique pour les établissements privés et sont regroupées au niveau entité juridique pour les établissements publics (pour être en cohérence avec le mode d'interrogation pré-refonte de 2013).

Les données couvrant 19 années, il existe potentiellement plusieurs ruptures de séries, dont celle (forte) correspondant à la refonte de l'enquête en 2013.

La table **CAPACT\_TOT0019** reprend les données disponibles dans la table CAPACT0019, mais cette fois agrégées pour toutes les disciplines d'équipement, au niveau établissement (géographique pour le privé et entité juridique pour le public).

#### b- Données sur le personnel médical rémunéré

##### Pour l'année 2019

La table **Q20\_2019R** (personnel médical rémunéré par spécialité exercée) : elle est au même format que la table Q20\_2019 des bases administratives, mais elle contient des corrections sur les ETP et les effectifs rémunérés, notamment suite à la correction de la non-réponse totale au bordereau et à la mise en cohérence avec le bordereau Q21.

La table **Q21\_2019R** (personnel médical rémunéré par groupe de disciplines d'équipement) : elle est au même format que la table Q21\_2019 des bases administratives, mais elle contient des corrections sur les ETP et les effectifs rémunérés, notamment suite à la correction de la non-réponse totale au bordereau et à la mise en cohérence avec le bordereau Q20.

La table **Q22\_2019R** (internes et faisant fonction d'internes) : elle est au même format que la table Q22\_2019 des bases administratives, mais elle contient des corrections sur les effectifs.

Séries longues (disponibles dans le sous-répertoire « Séries longues »)

La table **PM\_01\_18** contient, pour chaque établissement privé et entité juridique publique, le nombre d'ETP salariés et les effectifs des libéraux, pour chaque année de 2001 à 2019.

Les données couvrant 19 années, il existe potentiellement plusieurs ruptures de séries, dont celle (forte) correspondant à la refonte de l'enquête en 2013.

c- Données sur le personnel non médical rémunéré

Pour l'année 2019

La table **Q23\_2019R** (personnel non médical rémunéré) : elle est au même format que la table Q23\_2019 des bases administratives, mais elle contient des corrections sur les effectifs et ETP rémunérés, notamment suite à la correction de la non-réponse totale au bordereau.

La table **Q24\_2019R** (personnel non médical rémunéré par disciplines d'équipement) : elle est au même format que la table Q24\_2019 des bases administratives, mais elle contient des corrections sur les ETP de personnel non médical rémunéré, notamment suite la correction de la non-réponse totale au bordereau et à la mise en cohérence avec le bordereau Q23.

Séries longues (disponibles dans le sous-répertoire « Séries longues »)

La table **PNM\_01\_19** contient, pour chaque établissement privé et entité juridique publique, le nombre d'ETP salariés par catégorie de personnel, pour chaque année de 2001 à 2019.

Les données couvrant 19 années, il existe potentiellement plusieurs ruptures de séries, dont celle (forte) correspondant à la refonte de l'enquête en 2013.

Les données de la table **PNM\_01\_19** bénéficient toutefois d'un travail de rétopolation des données antérieures à 2011, qui permet de s'affranchir de plusieurs ruptures de série. En effet, le champ du personnel dans la SAE a subi plusieurs évolutions entre 2008 et 2011, si bien qu'il n'est pas possible de calculer les évolutions annuelles en utilisant directement les tables de chaque année. Par exemple, la SAE 2010 incluait les personnels des établissements sociaux, médico-sociaux et de formation, tandis qu'à partir de la SAE 2011, ceux-ci en sont exclus.

d- Nouvelles bases mises à disposition

Depuis la SAE 2016, deux nouvelles tables sont mises à disposition : les tables SYGEN\_AAAAR et GHT\_AAAAR.

La table **SYGEN\_2019R** reprend les informations contenues dans la partie « Activité générale et activités autorisées » du bordereau SYGEN. Elle est au même format que la table SYGEN\_2019 des bases administratives, mais elle est mise à jour des redressements apportés aux données des différents bordereaux du questionnaire.

Pour la SAE 2019, l'alternance des bordereaux BIO et PHARMA a été mise en place pour alléger la charge de collecte. Les données de ces deux bordereaux étant reprises dans la fiche de synthèse, les variables de ces deux bordereaux sont toujours présentes dans la table associée, mais seules celles du bordereau interrogé pour l'année considérée sont alimentées. Aussi, pour la SAE 2019, seules les

données de BIO sont alimentées avec les données collectées ; les variables de PHARMA sont présentes (pour faciliter les alimentations automatiques des systèmes d'information) mais sont vides.

La table **GHT\_2019R** contient des données agrégées sur le contour des groupements hospitaliers de territoires (GHT) au 31 décembre tels que le nombre de lits, de places, de journées et de séjours et des informations sur les équipements, les plateaux techniques et le personnel des établissements de santé. Elle est au même format que la table GHT\_2019 des bases administratives, mais elle est mise à jour des redressements apportés aux données des différents bordereaux du questionnaire.

Tout comme pour la base SYGEN, suite à l'alternance des bordereaux BIO et PHARMA dans la SAE, seules les données du bordereau interrogé pour l'année considérée sont alimentées bien que toutes les variables soient présentes dans la table, y compris celles du bordereau non interrogé (ces variables sont donc à vide).

Depuis la SAE 2018, une table supplémentaire est mise à disposition : la table HPR\_AAAAR.

La table **HPR\_2019R** contient les informations de la table SYGEN\_2019R sur le contour des hôpitaux de proximité uniquement, selon la liste définie par l'arrêté de mai 2019.