



## Accéder aux données diffusées par le CASD

Cette fiche, rédigée par l'équipe projet du Programme prioritaire de recherche (PPR) Autonomie piloté par le CNRS, a pour objectif de guider les équipes souhaitant accéder aux données confidentielles diffusées par le Centre d'accès sécurisé aux données (CASD). Elle présente la procédure de dépôt et de gestion des demandes d'accès ainsi que les modalités pratiques d'exploitation des données.

### Les données diffusées par le CASD

Les recherches sur l'autonomie des personnes (dans le cadre de leur vieillissement ou d'une situation de handicap) peuvent nécessiter la mobilisation de données individuelles détaillées (caractéristiques sociodémographiques, limitations fonctionnelles, état de santé, difficultés dans la vie quotidienne, ressources, localisation) qui ont souvent un caractère confidentiel.

C'est le CASD qui est principalement chargé d'organiser et de mettre en œuvre des services d'accès sécurisé pour ce type de données confidentielles produites par la statistique publique, des administrations ou d'autres producteurs (de données de santé, par exemple). Ainsi, il met à disposition les principales enquêtes de la DREES<sup>1</sup> sur le handicap et la perte d'autonomie : EHPA<sup>2</sup>, ES-Handicap<sup>3</sup>, RI-PCH<sup>4</sup>, VQS<sup>5</sup>, etc.

Sont également mis à disposition des appariements de données, réalisés directement par les producteurs des données ou par le CASD qui peut lui-même procéder à de nouveaux appariements sur demande des chercheurs, en accord avec les administrations productrices de données. Par exemple, sont diffusées les données de l'enquête CARE<sup>6</sup> appariées avec celles du SNDS<sup>7</sup>.

Toutes les données diffusées sont recensées dans le catalogue du CASD<sup>8</sup> qui en offre une description : année, producteur, dictionnaire des variables et de leurs modalités, liste de projets mobilisant les données, liste de publications déclarées par les utilisateurs, institutions autorisant l'accès, DOI<sup>9</sup>, etc. Une fois l'accès aux données autorisé au travers d'un processus *ad hoc*, le CASD met à disposition dans ses environnements sécurisés une documentation supplémentaire à celle présente sur le site, qui a été produite et transmise par le producteur des données. Cette documentation peut contenir des éléments techniques sur la construction des bases de données, les corrections apportées, le dictionnaire des codes et les questionnaires, etc.

1. Direction de la recherche, des études, de l'évaluation et des statistiques.

2. Enquête auprès des établissements d'hébergement pour personnes âgées. [Découvrir l'enquête EHPA](#).

3. Enquête auprès des établissements et services pour enfants et adultes handicapés. [Découvrir l'enquête ES-Handicap](#).

4. Remontées individuelles de la prestation de compensation du handicap. [Découvrir l'enquête RI-PCH](#).

5. Enquête « Vie quotidienne et santé ». [Découvrir l'enquête VQS](#).

6. Capacités, aides et ressources des seniors. [Découvrir l'enquête CARE](#).

7. Système national des données de santé.

8. [Découvrir le catalogue du CASD](#).

9. *Digital Object Identifier*, Identifiant numérique d'objet, utilisé pour le référencement des sources de données.

## BON À SAVOIR

Le CASD a pour objet principal d'organiser et de mettre en œuvre des services d'accès sécurisé pour les données confidentielles à des fins non-lucratives de recherche, d'étude, d'évaluation ou d'innovation. Il est membre du réseau IDAN<sup>10</sup> qui rassemble différents centres européens d'accès sécurisé aux données. Le but de ce réseau est de permettre également d'accéder aux données de l'ONS<sup>11</sup>, de l'UKDS<sup>12</sup>, du CBS<sup>13</sup>, du GESIS<sup>14</sup> et de l'IAB<sup>15</sup> selon des procédures définies avec chaque partenaire.

## Accéder aux données du CASD

Selon la nature et le calendrier de la demande, le délai pour accéder aux données peut varier. Anticiper les différentes étapes et préparer la demande en amont permettent de le réduire.

### Avant de déposer une demande

Avant de demander l'accès à des données mises à disposition par le CASD, il est recommandé :

- D'identifier la ou les sources de données pertinentes pour son projet et s'assurer qu'une version alternative (moins détaillée) ne suffirait pas. En effet, certaines données sont disponibles dans une version anonymisée en accès libre (par exemple sur la plateforme ouverte des données publiques françaises<sup>16</sup>) ou dans une version pseudonymisée avec des procédures d'accès moins strictes (par exemple sur Progedo<sup>17</sup> – on parle de FPR<sup>18</sup>). Bien que ces données soient moins précises, elles peuvent tout de même répondre aux besoins du projet.

- De bien évaluer et mettre en regard le temps nécessaire pour réaliser les démarches d'accès aux

10. *International Data Access Network*, Réseau international d'accès aux données. [Voir le site du réseau IDAN](#).

11. *Office for National Statistics*, Office des statistiques nationales (Royaume-Uni).

12. *United Kingdom Data Services*, Services des données du Royaume-Uni (Royaume-Uni).

13. *Centraal bureau voor de statistiek*, Bureau central des statistiques (Pays-Bas).

14. *Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen*, Institut Leibniz pour les sciences sociales (Allemagne).

15. *Institut für Beschäftigung und Employability*, Institut de recherche sur l'emploi (Allemagne).

16. [Accéder à la Plateforme ouverte des données publiques](#).

17. [Accéder à Progedo](#).

18. Fichiers de production et de recherche.

données (plusieurs mois) et le temps disponible pour cette étape dans le cadre du projet. Il s'agit par exemple de penser une organisation du travail suffisamment souple pour absorber un éventuel retard dans l'instruction de la demande d'accès, afin que celui-ci ne devienne pas une contrainte.

- De s'assurer des possibilités de prise en charge financière des coûts liés à l'accès et à l'exploitation des données. La location du dispositif de sécurisation de l'exploitation et les différents services du CASD (contrôles des exports, logiciels divers, configuration des serveurs de calcul, maintenance, mesures de sécurisation<sup>19</sup>, formation, déplacements, etc.) représentent un coût pouvant s'élever à plusieurs milliers d'euros par an selon les configurations<sup>20</sup>.

- D'établir en prévision de la demande la liste précise des personnes qui seront amenées à exploiter les données dans le cadre du projet. Elles doivent obligatoirement se créer un compte sur le site du CDAP<sup>21</sup> et signer un engagement de confidentialité.

- De se rapprocher du ou de la déléguée à la protection des données de son établissement. Il ou elle sera à même de déterminer si une analyse d'impact, une déclaration ou une autorisation de la CNIL<sup>22</sup> sont obligatoires pour le traitement des données et, le cas échéant, pourra accompagner leur réalisation.

Pour plus d'informations, consulter la page dédiée aux conseils sur le site du CSS<sup>23</sup>.

## BON À SAVOIR

Les fichiers de production et de recherche (FPR) sont des jeux de données individuelles comportant un niveau de détail moindre pour quelques variables potentiellement identifiantes. La procédure d'accès à ces données s'en trouve simplifiée, bien qu'elles restent couvertes par le secret statistique<sup>24</sup>.

19. [En savoir plus sur les certifications et la sécurité au CASD](#).

20. [Accéder au simulateur de coût du CASD](#).

21. *Confidential data access portal*, Portail d'accès à des données confidentielles. [Voir le site du CDAP](#).

22. Commission nationale de l'informatique et des libertés.

23. [Accéder à la page sur le site du comité du secret statistique](#).

24. [Télécharger la liste des FPR disponibles pour chaque producteur en date du 6 octobre 2022 \(PDF, 363Ko\)](#).

## Déposer une demande initiale

Déposer une demande d'accès à des données confidentielles nécessite de passer par plusieurs étapes, détaillées dans le Guide utilisateur complet<sup>25</sup> du CDAP pour le dépôt et la gestion des demandes d'accès à des données confidentielles dont les principales sont résumées ci-après.

- Créer un compte (ou se connecter) sur l'espace dédié aux demandes relevant du CSS<sup>26</sup> sur le CDAP et transmettre un engagement de confidentialité. À noter : toutes les demandes ne sont pas forcément instruites par le CSS. Pour les données du PMSI<sup>27</sup> ou du SNDS, un autre circuit est nécessaire (détaillé également sur le CDAP).
- Créer une demande depuis la rubrique « Tableau de bord » de son compte et renseigner les informations demandées. Il s'agit alors de présenter le projet de recherche (six questions ouvertes et une fermée), de renseigner les sources dont l'accès est demandé, de justifier de leur pertinence dans le cadre du projet (une question ouverte pour chaque source), de décrire la nature du travail qui sera réalisé (recherche ou statistique publique), les domaines de recherche mobilisés, la forme des publications finales envisagées (trois questions ouvertes et une fermée) et le niveau de sécurisation de l'environnement professionnel de travail (contrôle de l'accès aux locaux, sécurité informatique, etc.). Enfin, les adresses mail de l'ensemble des participants au projet sont communiquées afin que ceux-ci valident leur participation (sous réserve de la création obligatoire d'un compte).
- Demander leur accord préalable aux différents producteurs pour l'accès à leurs données directement depuis l'espace consacré à la demande<sup>28</sup>. Les producteurs peuvent faire des commentaires, demander des éclaircissements, voire s'opposer à l'accès aux données au regard des informations transmises. Le projet peut alors être retravaillé. Cette demande est à anticiper par rapport à la date limite de transmission de la demande finalisée au secrétariat du CSS en vue de son examen (minimum conseillé de deux semaines) ; il s'agit de laisser le temps aux producteurs pour formuler un

25. [Télécharger le guide utilisateur complet \(PDF, 1,7 Mo\)](#).

26. [Découvrir l'espace dédié sur le CDAP](#).

27. Programme de médicalisation des systèmes d'information.

28. [Consulter la liste des référents chez les producteurs pour chaque source diffusée](#).

avis. Le CSS analyse la demande finalisée et arbitre après un examen de celle-ci lors d'une séance en présentiel, qui se tient quatre fois par an, ou par voie de consultation électronique – pour les demandes complémentaires après mise en œuvre d'un projet de recherche – qui ont lieu huit fois par an<sup>29</sup>. Le CSS s'assure alors de l'intérêt suffisant de l'objet d'étude, que les informations demandées ne sont pas excessives, que le demandeur présente toutes les garanties de sécurité nécessaires, etc.

- Si le projet est validé, les membres du CSS émettent un avis favorable. Chaque producteur est alors invité à signer un accord d'accès aux données. Les Archives de France signent quant à elles une dérogation globale autorisant formellement l'accès à toutes les données. Il faut compter en moyenne trois semaines pour ces formalités.

- En parallèle de ces formalités administratives, après validation par le CSS, le CASD contacte le demandeur pour entamer les démarches d'accès. Il faut se procurer et installer une SD-Box (voir ci-après), se créer une carte d'identification individuelle (suppose le relevé des empreintes digitales dans les locaux du CASD) et assister à une session d'information et de sensibilisation obligatoire d'environ deux heures (aussi appelé « séance d'enrôlement »<sup>30</sup>).

## Déposer une demande complémentaire

Si un projet nécessite d'accéder à de nouvelles données, d'intégrer de nouveaux collaborateurs, de prolonger l'autorisation d'accès aux données, etc., il est possible de déposer une demande complémentaire.

Elle est à déposer par le responsable scientifique du projet ou par la personne à l'origine de la demande initiale depuis la rubrique « Tableau de bord » de leur compte sur l'espace dédié aux demandes relevant du CSS sur le CDAP. Cette demande suivra peu ou prou la même procédure qu'une demande initiale, notamment pour l'ajout d'une nouvelle source qui suppose d'obtenir au préalable l'accord du producteur, puis entraîne les mêmes procédures administratives que suite à l'avis favorable du CSS (voir ci-dessus). Une demande de prolongation peut quant à elle être instruite par voie électronique par le CSS.

29. [Consulter le calendrier des séances du CSS](#).

30. [En savoir plus sur le déroulé de la formation](#).

## BON À SAVOIR

Le Comité du secret statistique<sup>31</sup> est présidé par un conseiller d'État et composé de représentants de la CNIL, du Sénat, de l'Assemblée nationale, du Conseil économique, social et environnemental, des représentants de l'administration, des Archives nationales, de l'INED<sup>32</sup>, des syndicats, des entreprises, des services producteurs de données, de la recherche et des collectivités territoriales. Son rôle est d'examiner et de porter un avis sur chaque question qui relève du secret statistique ainsi que des demandes de communication de données confidentielles.

## Exploiter les données avec la SD-Box

Une fois la session de formation effectuée et après l'obtention d'une carte d'identification, la SD-Box peut être installée et l'exploitation des données peut débuter. La SD-Box est un terminal sécurisé qui permet d'accéder à son environnement de travail numérique grâce à une carte d'identification individuelle<sup>33</sup>. Les utilisateurs habilités d'un même projet accèdent au même environnement de calcul grâce à leur carte d'authentification. Les terminaux SD-Box peuvent être mutualisés et utilisés par plusieurs utilisateurs, de manière non-concomitante, voire par plusieurs projets disposant

31. [Accéder au site du comité du secret statistique.](#)

32. Institut national d'études démographiques.

33. [En savoir plus sur la SD-Box.](#)

chacun de son propre environnement sécurisé de calcul (sous réserve que les équipes projets s'organisent en ce sens).

L'environnement de travail intègre les sources pour lesquelles l'accès a été autorisé, différents logiciels (dont certains sont déjà installés – Microsoft office, R, SAS, etc. – d'autres peuvent l'être de manière optionnelle sous réserve d'une demande préalable – Philcarto, Python, etc.) ainsi que la puissance de calcul nécessaire au traitement des données, conformément au besoin exprimé par le responsable du projet<sup>34</sup>. Les traitements statistiques sont réalisés à distance sur cet environnement numérique.

Chaque import de données ou export de résultats sur et à partir de l'environnement numérique (pour soumission d'un article à un comité de lecture par exemple) est validé par le CASD. Pour les imports, ce dernier vérifie que les fichiers intégrés ne contreviennent pas au respect des règles de sécurité et que leur introduction au sein de l'environnement est conforme avec les autorisations du projet. Pour les exports, il vérifie que les exploitations réalisées ne permettent pas l'identification des individus et respectent les règles de confidentialité édictées par chaque producteur de données<sup>35</sup>.

34. [En savoir plus sur l'environnement de travail accessible depuis une SD-Box.](#)

35. [Consulter les règles de confidentialités à appliquer aux sorties de résultat en fonction du producteur.](#)

## Contacts

L'équipe projet du PPR Autonomie se tient à votre disposition pour vous orienter vers des ressources ou des personnes compétentes. Des actions sont conduites afin de favoriser l'émergence et le développement d'une communauté de pratiques autour des expériences de gestion, diffusion et réutilisation des données.

**N'hésitez pas à nous suivre pour rester informé de nos actualités !**



[Site du PPR Autonomie](#)



[S'inscrire à la newsletter du PPR Autonomie](#)



[Linkedin](#)



[Contacter l'équipe projet](#)