

Exploitation DADS à destination des chercheurs

Création d'un fichier de production et de recherche (FPR)

Contexte

Les chercheurs exploitent actuellement le fichier DADS super-anonymisé disponible pour tout public sur le site www.insee.fr.

Il a été décidé lors du comité client produit de janvier 2010 de leur fournir un fichier avec non plus les tranches de rémunération mais les rémunérations réelles.

Objectif

Le but de la demande est d'obtenir deux tables SAS de niveau postes et salariés sur l'année 2011. Un fichier ne comprend qu'une validité.

Il s'agit des fichiers nationaux « postes » et « salariés » au 12^{ème} (échantillon au 12^{ème} constitué par tous les salariés nés au mois d'octobre).

Pour chacune des tables, mis à part pour les variables de rémunération, il s'agit de fournir des fichiers à l'image des fichiers détails insee.fr de l'année considéré.

Les descriptifs des fichiers postes et salariés sont les suivants

2011 - Fichier postes 1/12°

Informations concernant le salarié

L'âge du salarié (AGE) ;

Le sexe du salarié (SEXE) ;

La région de résidence (REGR) ;

Le département de résidence (DEPR) ;

La zone d'emploi de résidence (ZEMPR) : zone d'emploi 2011 ; ATTENTION

Le code étranger (IND_ETRANGER).

Informations sur l'établissement

La région d'implantation (REGT) ;

Le département d'implantation (DEPT) ;

La zone d'emploi d'implantation (ZEMPT) : : zone d'emploi 2011 ; ATTENTION

Le domaine d'emploi du lieu d'affectation (DOMEMPL) : **attention nouvelle définition depuis 2009 et nouvelles modalités ;**

Le domaine d'emploi de l'employeur (DOMEMPL_EMPL) : **attention nouvelle variable depuis 2009 ;**

L'activité en Naf, rév2 (A6, A17 et A38) ;

La tranche d'effectif (TREFF calculé à partir de TREFFECT).

Informations sur le poste

La date de début de période (DATDEB) ;

La date de fin de période (DATFIN) ;

La durée de la période (DUREE) ;

Le nombre d'années-travail (AT) calculé à partir de la variable DUREE (DUREE/360) ;

Le type de poste (annexe ou non annexe : FILT) ;

La convention collective du salarié (CONV_COLL) ;

La condition d'emploi (CPFD) ;

La PCS codée (PCS) ;

La catégorie socioprofessionnelle (CS) : 2 1ers caractères de la PCS ;

Le type d'emploi (TYPEEMPLOI) ;

Le nombre d'heures salariées (NBHEUR) ;

L'équivalent temps-plein (ETP) ;

Le contrat de travail (CONT_TRAV) calculé à partir de la variable CONTRAT_TRAVAIL ;

La rémunération brute (S_BRUT).

Les modalités ou valeur des variables sont les mêmes que ceux des fichiers de diffusion à l'exception :

CONT_TRAVAIL

'CDI' = CDI

'CDD' = CDD

'APP' = apprentissage

'TTP' = mission d'intérim, travail temporaire

'AUT' = autres

'ZZZ' = contrat de travail manquant

TREFF

. = Non disponible

'00' = 0 poste (salariés présents en cours d'année, mais pas au 31/12)

'01' = 1 à 9 postes

'02' = 10 à 19 postes

'03' = 20 à 49 postes

'04' = 50 à 99 postes

'05' = 100 à 249 postes

'06' = 250 postes et plus

TYPEEMPLOI

'O' = Ordinaire

'A' = Apprenti

'X' = Autre

2011 - Fichier salariés 1/12°

Informations concernant le salarié

L'âge du salarié (AGE) ;

Le sexe du salarié (SEXE) ;

La région de résidence (REGR) ;

Le département de résidence (DEPR) ;

La zone d'emploi de résidence (ZEMPR) : zone d'emploi 2011 ; ATTENTION

Informations sur l'établissement du poste principal

La région d'implantation (REGT) ;

Le département d'implantation (DEPT) ;

La zone d'emploi d'implantation (ZEMPT) : zone d'emploi 2011 ; ATTENTION

Le domaine d'emploi du lieu d'affectation (DOMEMPL) : **attention nouvelle définition depuis 2009 et nouvelles modalités** ;

L'activité en Naf, rév2 (A6, A17 et A38) ;

La tranche d'effectif (TREFF) calculé à partir de TREFFECT.

Informations sur le poste principal

La date de début de période (DATDEB) ;

La date de fin de période (DATFIN) ;

La durée de la période (DUREE) ;

La convention collective du salarié (CONV_COLL) ;

La condition d'emploi (CPFD) ;

La PCS codée (PCS) ;

La catégorie socioprofessionnelle (CS) : 2 1ers caractères de la PCS ;

Le type d'emploi (TYPEEMPLOI) ;

Le nombre d'heures salariées (NBHEUR) ;

Le contrat de travail (CONT_TRAV) calculé à partir de la variable CONTRAT_TRAVAIL.

Informations sur l'ensemble des postes effectués par le salarié

Le total du nombre d'heures (NBHEUR_TOT) ;

Le nombre de périodes effectuées (NB_PER) ;

Le nombre de périodes non annexes effectuées (NB_PER_N)

Le total des rémunérations nettes (S_NET_TOT) ;

Le total des indemnités ASSEDIC (ALCHO).

Les modalités ou valeur des variables sont les mêmes que ceux des fichiers de diffusion à l'exception :

CONT_TRAVAIL

'CDI' = CDI

'CDD' = CDD

'APP' = apprentissage

'TTP' = mission d'intérim, travail temporaire

'AUT' = autres

'ZZZ' = contrat de travail manquant

TREFF

. = Non disponible

'00' = 0 poste (salariés présents en cours d'année, mais pas au 31/12)

'01' = 1 à 9 postes

'02' = 10 à 19 postes

'03' = 20 à 49 postes

'04' = 50 à 99 postes

'05' = 100 à 249 postes

'06' = 250 postes et plus

TYPEEMPLOI

'O' = Ordinaire

'A' = Apprenti

'X' = Autre

Méthodologie

La Déclaration Annuelle des Données Sociales (DADS) est une formalité déclarative que doit accomplir toute entreprise employant des salariés, en application de l'article R243-14 du Code de la Sécurité Sociale (décret du 24 mars 1972) et des articles 87.240 et 241 de la loi 51-711 du 7 juin 1951 du Code Général des Impôts.

La DADS contient les informations destinées à l'usage de différentes institutions (URSSAF, CPAM, IRCANTEC, UNEDIC, CNRACL) et à l'INSEE pour l'observation des salaires et du volume de travail associé. Depuis 1993, l'exploitation porte sur l'ensemble des employeurs et de leurs salariés, dans le champ défini ci-dessous.

À partir de la validité 2009, sont ajoutées aux informations issues de la source DADS, celles construites à partir des fichiers de paye des agents de l'État et des fichiers de déclaration des particuliers qui emploient des salariés.

En résumé, à compter de la validité 2009, le champ de diffusion des DADS - grand format couvre l'ensemble des employeurs et de leurs salariés, sauf les activités extra-territoriales (division NAF Rev2 99).

L'exploitation des différentes sources est pour partie informatique et pour partie manuelle. Les traitements manuels sont effectués dans trois pôles, situés dans les directions régionales de Reims et de Dijon pour les salariés du secteur privé et dans la direction régionale de Rennes pour ceux des secteurs publics ; les traitements des salariés des particuliers-employeurs sont beaucoup plus automatisés et suivis uniquement à EFA. La chaîne de traitement informatique des DADS a été profondément modifiée à partir de la validité des DADS 2002 et bien aménagée pour la validité 2009. La nouvelle chaîne d'exploitation repose sur un traitement complet de chaque DADS, entreprise par entreprise, au fur et à mesure de l'arrivée de celles-ci, l'objectif étant de terminer tous les traitements d'une entreprise le plus rapidement possible. La production et la maintenance des 3 chaînes de traitement sont assurées par le CNIO.

Le fichier échantillon national « postes » (échantillon au 12ème) : données individuelles ainsi que données « établissements » redescendues au niveau « postes » ;

Le fichier échantillon national « salariés » au 12ème ;

Particularités propres aux validités 2009, 2010 et 2011

2009

❖ Extension du champ

A partir de la validité 2009, le champ est étendu à l'ensemble des salariés, à l'exception des activités extra-territoriales. Le champ est donc étendu sur l'année 2009 aux :

- particuliers employeurs
- salariés de la fonction publique d'État

Comme les fichiers des années précédentes et pour être cohérent avec les fichiers détail Insee.fr, les secteurs dont l'Apet commencent par 01 et 02 ont été retirés des fichiers FPR mis à disposition.

❖ **Cas particulier de la fonction publique d'État : ajout de la notion d'employeur**

Dans les données DADS, y compris sur la validité 2009, le SIREN de l'entreprise est toujours celui du lieu de travail du salarié. Sur les validités précédentes cette notion était confondue avec celle de SIREN employeur.

A partir de 2009, ce n'est plus le cas avec l'ajout de la fonction publique d'État. En effet, un professeur de lycée fonctionnaire sera déclaré par le SIREN de son établissement scolaire mais son employeur (Ministère de l'Éducation Nationale) possède un autre SIREN et relève d'une autre catégorie juridique.

C'est la raison pour laquelle il a été ajouté une **nouvelle variable relative à l'employeur du salarié** :

- DOMEMPL_EMPL = domaine d'emploi de l'employeur, donnée calculée notamment à partir de la catégorie juridique employeur mais au contour affiné pour certains SIREN (*voir paragraphe suivant concernant cette variable*)

Cette variable de niveau employeur est le pendant de DOMEMPL déjà disponibles dans les DADS et relative au lieu de travail. Si le SIREN employeur se confond avec le SIREN de lieu de travail, cette nouvelle variable est égale à DOMEMPL correspondant au SIREN lieu de travail.

- SIREN_EMPL = SIREN de l'entreprise employeur
- si SIREN = SIREN_EMPL alors DOMEMPL_EMPL=DOMEMPL

❖ **Nouvelle définition pour DOMEMPL**

La variable « domaine d'emploi » a été entièrement redéfinie pour mieux s'adapter au grand format DADS. Voici la nouvelle définition de DOMEMPL : (qui se calcule à partir de la catégorie juridique et du SIREN)

| DOMEMPL= | Intitulé | Filtre sur catégorie juridique et SIREN |
|----------|--|--|
| 0 | Entreprises absentes du référentiel | CJ=00 |
| 1 | Fonction Publique d'État | CJ=71** et CJ=7331 et CJ de 7382 à 7389 (à l'exception des SIREN 180014045, 180035016, 180035024, 180035032, 180035065, 180036063, 189400039) + siren 180020026 (CDC) |
| 2 | Fonction Publique Territoriale | CJ=72** et CJ=7312 à 7314 et CJ=7341 à 7363 et CJ=7371 à 7379 + siren=180014045 (CNFPT) |
| 3 | Fonction Publique Hospitalière | CJ=7364 à 7366 + siren 189400039 et 180036063 |
| 4 | Autres organismes publics administratifs | CJ=74** (sauf siren 180020026) et CJ=7321, 7322, 7323, 7381 + siren 180035016, 180035065, 180035024 et 180035032 |
| 5 | Personnes morales de droit public soumises au droit commercial | CJ=4*** |
| 6 | Entreprises individuelles | CJ=1*** (sauf source PE) et CJ=5498 |
| 7 | Particuliers Employeurs | données marquées comme issues de PE |
| 8 | Organismes privés spécialisés et groupements de droit privé | CJ=8*** et CJ=9*** |
| 9 | Autres sociétés privées | CJ=2*** et CJ=3*** et CJ=5*** (sauf 5498) et CJ=6*** |

Remarques :

- la variable DOMEMPL_EMPL présente la même définition mais pour l'employeur.

❖ **Nouveauté 2009 : codification de la PCS**

Jusqu'à la validité 2008 des DADS, la variable de mesure de la qualification était la CS-ESE à 2 chiffres (catégorie socioprofessionnelle). A partir de la validité 2009 figure désormais la codification de la profession à 4 positions codée selon la nomenclature PCS-ES (il s'agit d'une nomenclature adaptée aux DADS disponible sur www.insee.fr). Les deux premières positions de la PCS permettent de se ramener à la CS.

La finalité d'une codification en PCS-ES est de disposer d'une information plus complète, notamment en acquérant une notion de profession, ce que ne permettait pas la CS.

La variable PCS est disponible sur l'ensemble du champ salarié.

Cette PCS-ES est déterminée selon différentes méthodes en fonction de l'application source :

- dans l'application DADS (secteur privé), elle est obtenue à partir du libellé d'emploi grâce à un système de codification utilisant le logiciel Sicore et bénéficie comme pour la CS d'un traitement manuel ciblé sur une partie de l'échantillon au 1/12^{ème}.

Dans l'application SIASP (fonctions publiques), elle se déduit de la PCS déclarée par l'employeur, redressée et contrôlée ensuite par l'information du grade.

- Dans l'application PE (particuliers employeurs), la PCS s'obtient à partir du métier déclaré complété par d'autres variables comme le code d'exonération de l'employeur ou le type de personnel.

➤ **Rupture sur la partie secteur privé (source=DADS)**

Une refonte du système de codification a donc été réalisée dans l'application de gestion DADS pour parvenir à une PCS-ES comme indication de la qualification. En conséquence, il y a une **rupture de série** entre la CS 2009 (= 2 premiers chiffres de la PCS) et la CS 2008.

En particulier, la PCS 2009 code davantage d'employés (CS 5) et moins de professions intermédiaires (CS 4). Il y'a également une hausse du codage de cadres (du fait de l'utilisation du statut professionnel : variable plus fiable pour repérer les cadres).

➤ **Rupture sur le champ fonction publique territoriale et hospitalière (source=SIASP en 2009 et DADS en 2008)**

La variable PCS_1 rétropolée n'a pas pu être calculée sur ce champ, elle est donc à valeur manquante dans les fichiers DADS 2009 sur cette population. Par ailleurs, une comparaison de la structure de la CS à l'aide des fichiers DADS de diffusion 2008 et des fichiers de diffusion 2009 sur le champ FPT ou FPH amènera à un risque de rupture, car ces deux CS ont été calculées selon deux méthodes différentes (application DADS en 2008 et application SIASP en 2009).

❖ **Nomenclature agrégée**

Pour la nomenclature agrégée A6 (issue de la NAF rév 2), la modalité « GU » a été corrigée en modalité « JU » sur les deux années (variables A6 et A6_1).

❖ **ETP**

L'utilisation de la variable ETP telle que décrite dans le guide méthodologique des DADS n'est possible que sur le secteur privé (hors PE). Pour un champ plus large il faut en redéfinir une au moyen des variables disponibles sur tout le champ (nombre d'heures, durée du poste, APET, etc...).

2011

La validité 2011 des DADS est la première à être issue de déclarations réalisées au format N4DS. Cette nouvelle norme de déclaration modifie la nature des informations collectées et refond très largement l'organisation du formulaire. Elle a donc entraîné pour les éditeurs de logiciels de paie en sortie desquels les DADS sont émises d'importantes maintenances, réalisées ou pas à temps pour la collecte. Les conséquences sur la diffusion des données sont nombreuses et de différents ordres, d'autant que la nouvelle norme n'a pas été prise en compte directement par les applications SIASP et DADS (cf. infra).

1. Effet sur la déclaration

Les entreprises ont probablement profité du passage à la N4DS pour « toiletter » leur système de gestion des ressources humaines, notamment les libellés d'emplois, qui sont la brique de base pour le codage de la PCS (profession et catégorie socioprofessionnelle). On constate ainsi des changements de libellés de professions plus fréquents entre 2010 et 2011 qu'entre 2009 et 2010. Si l'on s'en tient aux salariés présents deux années consécutives dans le même établissement (les seuls pour lesquels une telle comparaison a du sens) :

- 44 % des établissements ont autant de « libellés stables » entre 2010 et 2011 qu'entre 2009 et 2010 ;
- 26 % des établissements ont plus de « libellés stables » entre 2010 et 2011 qu'entre 2009 et 2010 ;
- 30 % des établissements ont moins de « libellés stables » entre 2010 et 2011 qu'entre 2009 et 2010.

A priori, ce changement apporte une amélioration de la PCS déclarée par l'entreprise et donc de la PCS diffusée. Mais on peut constater des ruptures, de plus ou moins grande importance, entre 2010 et 2011.

2. Effet du à la recodification

Les données N4DS ne sont pas intégrées telles quelles dans la chaîne DADS qui gère le secteur privé et la chaîne SIASP qui gère le secteur public : elles sont recodifiées via une application amont appelée « frontal » pour être remises en ancienne norme, avant d'être chargées et traitées. Ainsi, certaines variables absentes de la nouvelle norme mais nécessaires au fonctionnement de l'application sont entièrement reconstituées par le frontal. La principale distorsion liée à ce mode de prise en compte de la nouvelle norme concerne la PCS. La structure de la PCS en sortie de SICORE est déformée par rapport aux validités précédentes.

Ainsi, sur l'ensemble des périodes, on observe une baisse de 1,0 point des professions intermédiaires et de 0,4 point des employés, et une hausse de 1,4 point des ouvriers entre 2010 et 2011. Cet écart peut s'expliquer au moins en partie par l'effet de déclaration et l'effet de recodification. Il se retrouve dans une moindre mesure sur la PCS diffusée après traitements manuels par les gestionnaires.

3. Effet « retard »

Le traitement de la validité 2011 par les gestionnaires a débuté avec un mois de retard par rapport à ce qui se fait habituellement, mais s'est achevé à la même date. Les données 2011 souffrent donc d'un mois de traitement gestionnaire en moins par rapport aux autres années. Cela a un effet sur la codification de la PCS (plus de PCS tirées aléatoirement) et sur les traitements de cohérence : beaucoup plus de DADS que d'ordinaire ont été redressées automatiquement, sans expertise manuelle.

De plus, une partie importante des données en provenance DGFIP (une partie des régimes spéciaux, dont la MSA) n'a été chargée que fin janvier (au lieu de septembre) : cela représente 26 700 DADS et 653 000 périodes. Ces données n'ont été traitées qu'en redressement automatique sans passer par l'expertise gestionnaire.

4. Effet « déclarations complémentaires »

Les établissements ont utilisé, à tort, la possibilité de faire des déclarations complémentaires pour pallier le fait qu'ils ne pouvaient pas faire de déclaration « annule et remplace ». Cette information a été donnée par les partenaires de la norme sans qu'il y ait eu de chiffrage précis du volume concerné.

La principale conséquence de ce problème en termes de données est que, malgré des traitements de détection des déclarations en double réalisés dans la chaîne DADS, on a pu garder à tort deux déclarations pour un même établissement. Il pourrait donc y avoir une surestimation, dans certains établissements, des salaires et temps de travail, surestimation dont il est difficile d'estimer l'ordre de grandeur.

Liste alphabétique des variables et des attributs : fichier « Postes »

| Variable | Type | Long. | Format | Informat | Libellé |
|--------------|-------|-------|--------|----------|-------------------------------------|
| A17 | Texte | 2 | \$2. | \$2. | A17 |
| A38 | Texte | 2 | \$2.\$ | 2. | A38 |
| A6 | Texte | 2 | \$2. | \$2. | A6 |
| AGE | Num. | 8 | | | AGE DU SALARIÉ |
| AT | Num. | 8 | | | ANNÉE-TRAVAIL DURÉE/JOURS |
| CONT_TRAV | Texte | 3 | | | CONTRAT DE TRAVAIL |
| CONV_COLL | Texte | 4 | | | CONVENTION COLLECTIVE NOMENCLATURE |
| CPFD | Texte | 1 | | | CPFD |
| CS | Texte | 4 | | | CS : 2 1ERS CARACTERES DE PCS |
| DATDEB | Num. | 8 | | | DATE DE DÉBUT |
| DATFIN | Num. | 8 | | | DATE DE FIN |
| DEPR | Texte | 3 | \$3. | \$3. | DÉPARTEMENT DE RÉSIDENCE |
| DEPT | Texte | 3 | \$3. | \$3. | DÉPARTEMENT DE TRAVAIL |
| DOMEMPL | Texte | 1 | \$1. | \$1. | DOMAINE D'EMPLOI DU LIEU DE TRAVAIL |
| DOMEMPL_EMPL | Texte | 1 | \$1. | \$1. | DOMAINE D'EMPLOI DE L'EMPLOYEUR |
| DUREE | Num. | 8 | | | DURÉE |
| ETP | Num. | 8 | | | ÉQUIVALENT TEMPS PLEIN |
| FILT | Texte | 1 | | | TYPE DE POSTE |
| IND_ETRANGER | Texte | 1 | | | CODE ÉTRANGER |
| NBHEUR | Num. | 8 | | | NOMBRE D'HEURES |
| PCS | Texte | 4 | | | PCS-ESE |
| REGR | Texte | 2 | \$2. | \$2. | RÉGION DE RESIDENCE |
| REGT | Texte | 2 | \$2. | \$2. | RÉGION DE TRAVAIL |
| S_BRUT | Num. | 8 | | | SALAIRE BRUT |
| SEXE | Texte | 1 | \$1. | \$1. | SEXE |
| TREFF | Texte | 2 | \$2. | | TRANCHE D'EFFECTIFS ÉTABLISSEMENT |
| TYPEMPLOI | Texte | 1 | | | TYPE EMPLOI |
| ZEMPR | Texte | 4 | \$4. | \$4. | ZONE D'EMPLOI DE RÉSIDENCE |
| ZEMPT | Texte | 4 | \$4. | \$4. | ZONE D'EMPLOI DE TRAVAIL |

Liste alphabétique des variables et des attributs : fichier « Salariés »

| Variable | Type | Long. | Format | Informat | Libellé |
|--------------|-------|-------|--------|----------|-------------------------------------|
| A6 | Texte | 2 | \$2. | \$2. | A6 |
| A17 | Texte | 2 | \$2. | \$2. | A17 |
| A38 | Texte | 2 | \$2.\$ | 2. | A38 |
| AGE | Num. | 8 | | | AGE DU SALARIÉ |
| ALCHO | Num. | 8 | | | TOTAL DES INDEMNITÉS ASSEDIC |
| CONT_TRAV | Texte | 3 | | | CONTRAT DE TRAVAIL |
| CONV_COLL | Texte | 4 | | | CONVENTION COLLECTIVE NOMENCLATURE |
| CPFD | Texte | 1 | | | CPFD |
| CS | Texte | 4 | | | CS : 2 1ERS CARACTERES DE PCS |
| DATDEB | Num. | 8 | | | DATE DE DÉBUT |
| DATFIN | Num. | 8 | | | DATE DE FIN |
| DEPR | Texte | 3 | \$3. | \$3. | DÉPARTEMENT DE RÉSIDENCE |
| DEPT | Texte | 3 | \$3. | \$3. | DÉPARTEMENT DE TRAVAIL |
| DOMEMPL | Texte | 1 | \$1. | \$1. | DOMAINE D'EMPLOI DU LIEU DE TRAVAIL |
| DOMEMPL_EMPL | Texte | 1 | \$1. | \$1. | DOMAINE D'EMPLOI DE L'EMPLOYEUR |
| DUREE | Num. | 8 | | | DURÉE |
| FILT | Texte | 1 | | | TYPE DE POSTE |
| NBHEUR | Num. | 8 | | | NOMBRE D'HEURES |
| NBHEUR_TOT | Num. | 8 | | | NOMBRE D'HEURES TOTAL |
| NB_PER | Num. | 8 | | | NOMBRE DE PÉRIODES |
| NB_PER_TOT | Num. | 8 | | | NOMBRE DE PÉRIODES TOTAL |
| PCS | Texte | 4 | | | PCS-ESE |
| REGR | Texte | 2 | \$2. | \$2. | RÉGION DE RESIDENCE |
| REGT | Texte | 2 | \$2. | \$2. | RÉGION DE TRAVAIL |
| SEXE | Texte | 1 | \$1. | \$1. | SEXE |
| S_NET_TOT | Num. | 8 | | | TOTAL DES SALAIRES NETS |
| TREFF | Texte | 2 | \$2. | | TRANCHE D'EFFECTIFS ÉTABLISSEMENT |
| TYPEEMPLOI | Texte | 1 | | | TYPE EMPLOI |
| ZEMPR | Texte | 4 | \$4. | \$4. | ZONE D'EMPLOI DE RÉSIDENCE |
| ZEMPT | Texte | 4 | \$4. | \$4. | ZONE D'EMPLOI DE TRAVAIL |