

Utilisation des données de la Déclaration sociale nominative (DSN) à des fins de statistiques publiques ou de pilotage

Rapport

François BRUNET

Nicole ROTH

**Membres de l'Inspection générale de
l'Insee**



N°2023_24/DG75-B001

Irina SCHAPIRA

**Membre de l'Inspection générale des
affaires sociales**



N°2022-097R

Avril 2023

SYNTHÈSE

[1] L'Inspection générale des affaires sociales (IGAS) et l'Inspection générale de l'Insee ont été saisies par le cabinet du ministre du Travail, du Plein emploi et de l'Insertion (MTPEI) d'une mission relative aux usages des données issues de la Déclaration sociale nominative (DSN) à des fins de statistiques publiques et de pilotage dans le champ du travail, de l'emploi et de la formation professionnelle. Cette mission s'inscrit aussi dans le cadre de la feuille de route ministérielle des données, des algorithmes et des codes-sources de 2021, qui prévoyait la création d'un « *Labor data hub* » visant en particulier à favoriser la valorisation des données issues de la DSN. Elle propose une cartographie des usages actuels ou potentiels de ces données et met en exergue un besoin d'expertise de ces données et d'accompagnement de leurs usages dans un contexte de maturité en moyenne encore modeste des acteurs du champ travail-emploi-formation vis-à-vis de la science des données.

[2] **La DSN a été créée en 2012 dans l'objectif de simplifier la communication par les employeurs des informations nécessaires à la gestion de la protection sociale de leurs salariés.** Ce processus déclaratif s'est substitué progressivement à un nombre croissant de formalités antérieures. Son périmètre a été progressivement étendu en 2017 à la totalité des entreprises du secteur privé puis en 2022 à l'ensemble du secteur public. Seuls les salariés des particuliers-employeurs restent exclus à ce stade de ce dispositif (outre les travailleurs indépendants qui par définition ne relèvent d'aucun employeur). Les données de la DSN sont celles des flux d'information issus de ces déclarations mensuelles qui sont fondées sur l'acte de paie. Le Gip-MDS, sous la responsabilité de la Direction de la sécurité sociale (DSS), élabore le cahier des charges définissant la norme applicable en lien avec les éditeurs de logiciels de paie.

[3] **La DSN constitue aussi un gisement très riche d'informations sur le marché du travail et l'emploi, soit directement, soit par croisement avec d'autres données.** Avec 26 millions d'actes déclaratifs chaque mois et environ 150 variables par salarié, elle offre l'opportunité d'améliorer la connaissance du marché du travail (par exemple, objectiver des tensions de recrutement dans un métier, un secteur d'activité ou un territoire), d'apprécier l'impact des interventions publiques et de les ajuster plus rapidement. La DSN permet aussi de constituer de nouvelles bases statistiques enrichies grâce au croisement avec d'autres sources de données. Ces appariements contribuent au suivi des trajectoires individuelles sur le marché du travail (par exemple, pour les personnes ayant eu un contact avec le service public de l'emploi, les allocataires de minima sociaux ou encore les sortants du système éducatif). Ils se substituent pour partie aux enquêtes d'insertion professionnelle qui étaient menées auparavant sur des échantillons et qui, en raison de leur coût, étaient moins précises et/ou fréquentes.

[4] **La DSN est un traitement de données à caractère personnel au sens de la Commission nationale informatique et libertés (Cnil).** Elle contient aussi des données sensibles sur les salariés telles que la qualité de travailleur handicapé ou la perception d'indemnités journalières. Elle peut aussi révéler des données relatives à la politique des ressources humaines des employeurs (ex : politique salariale).

[5] **Les acteurs du Service statistique public (SSP) retraitent la DSN pour établir des statistiques de référence en matière d'emploi et de rémunérations salariales.** Ces retraitements consistent en particulier à corriger les principales erreurs déclaratives, gérer les risques de doublons et les déclarations retardataires ou rectificatives, et traduire les rubriques administratives en concepts et nomenclatures statistiques. La DSN permet un raccourcissement important des délais de production au sein du SSP. Le délai de l'estimation flash de l'emploi a été réduit en 2023 à moins de 40 jours après le trimestre considéré et le délai des Estimations trimestrielles de l'emploi détaillées sera réduit à 60 jours, contre 70 jours auparavant, tandis que celui de la base Tous salariés (BTS), qui fournit des informations plus détaillées sur l'emploi et les salaires, sera réduit à un an à l'horizon 2024, contre environ deux ans en 2022. La DSN a aussi constitué un levier de refonte des systèmes d'information du SSP et d'amélioration de la coordination entre les acteurs. Ainsi, depuis 2017, l'Insee, la Dares et l'Urssaf Caisse nationale produisent conjointement les estimations trimestrielles de l'emploi salarié, ce qui contribue à la cohérence des indicateurs et à l'optimisation des moyens d'expertise. La mission a néanmoins relevé des enjeux d'amélioration de la coordination dans l'utilisation de la DSN au sein du SSP, d'une part, dans le domaine du handicap (entre la Dares et l'Urssaf Caisse nationale) et, d'autre part, dans le domaine de la formation professionnelle (entre la Dares et la Caisse des dépôts et consignations).

[6] **Au-delà de ces usages à des fins de statistique publique, les usages des données de la DSN à des fins de pilotage montent progressivement en puissance.** Le recours aux données de la DSN à des fins de pilotage devrait encore progresser avec la création de « France Travail » qui vise à personnaliser davantage les services en fonction des besoins des demandeurs d'emploi et des entreprises, et à mieux mesurer l'impact sur le retour à l'emploi des dispositifs mobilisés. Sans que la cartographie établie soit nécessairement exhaustive, la mission propose une typologie des cas d'usages rencontrés en fonction de leur objectif principal :

- **analyser les parcours en emploi des bénéficiaires de dispositifs de politiques publiques :** la DGEFP a mobilisé le Gip-MDS pour suivre le parcours en emploi de cohortes de bénéficiaires du Contrat d'engagement jeunes (CEJ), le Gip « Plateforme de l'inclusion » s'est adressé à la Dnum pour suivre le parcours en emploi d'allocataires du RSA dans trois départements et, enfin, la Caisse des dépôts et consignations et la DGEFP ont déployé la plateforme « Agora » qui fournit, à partir de la DSN, des données sur le retour à l'emploi des bénéficiaires d'un dispositif de formation ;
- **cibler le contrôle et l'accompagnement des entreprises dans le respect des dispositions prévues par le code du Travail :** la DGT a mis en place en 2021 un « SI précarité » pour aider l'inspection du travail à repérer les entreprises en risque d'abuser des contrats courts ; elle utilise les effectifs des entreprises en DSN pour identifier celles dont le taux d'accidents du travail (« Dat'IA ») est hors norme ou qui ne remplissent pas leur obligation de calculer l'index d'égalité professionnelle (« EgaPro ») alors qu'elles y sont assujetties ; la DGEFP a développé l'appliquetif « OETH » pour outiller les Dreets afin de mieux repérer et accompagner les entreprises qui ne respecteraient pas leurs obligations en matière d'emploi handicapé ;
- **éclairer le marché caché de l'emploi en identifiant les entreprises susceptibles de recruter :** la start-up d'État « La bonne alternance » souhaite ainsi accéder aux données de la DSN pour identifier de façon prédictive les entreprises les plus susceptibles d'offrir des contrats en alternance ;

- **anticiper l'impact et suivre la montée en charge d'un nouveau dispositif de politique publique** : la base de données dite Dad-E du Gip-MDS a ainsi été mobilisée par la DSS pour analyser le déploiement des primes « Pepa¹ » en 2019 et « inflation » en 2021 ;
- **identifier le plus en amont possible dans un secteur donné les tensions en matière de RH et les facteurs d'attractivité des métiers** : c'est l'objet du « cube DSN »² en cours de développement en lien avec la Dares dans le SI Diamant destiné aux Agences régionales de santé (ARS), qui permettra d'identifier des marqueurs de tension RH dans le secteur sanitaire, médico-social ou social (augmentation du volume d'heures supplémentaires, de l'absentéisme, du turn-over, du recours à l'intérim...) ; cette ambition est partagée par le projet « Ramsès », développé par un prestataire pour la DGCS, afin d'analyser l'impact des conventions soumises à son agrément sur la masse salariale du secteur médico-social et social à but non lucratif ;
- **analyser l'évolution des emplois et des compétences au niveau d'un secteur, d'une branche ou d'un territoire afin de détecter des risques ou des opportunités** : c'est à cette fin que la start-up d'État « Signaux faibles », qui vise à repérer les entreprises en risque de défaillance, souhaite mobiliser les données de la DSN, tandis que la tête de réseau des Carif-Oref utilise déjà ces données dans le cadre du projet « Octopilot » pour apprécier les passerelles possibles entre métiers par région.

[7] **La feuille de route « données » du ministère en charge du Travail identifie par ailleurs le besoin de promouvoir et de valoriser l'usage des données de la DSN, en particulier avec la création d'un « Labor data hub ».** Dans la continuité du rapport « Bothorel » de décembre 2020, la circulaire du Premier ministre du 27 avril 2021 a en effet rappelé la nécessité de « *rechercher en permanence la meilleure circulation de la donnée* », en lien avec les administrateurs ministériels des données, fonction exercée par la Direction du numérique (Dnum) pour le MTPEI.

[8] **Toutefois, ce « Labor data hub », qui devait être créé à l'horizon d'un an, reste à ce jour un objet administratif non identifié.** Si le cadrage réalisé par la Dnum a permis de mieux identifier les attentes des différentes parties prenantes, il a sous-estimé les enjeux liés à la gouvernance des usages et à l'application du cadre juridique existant, au profit d'une approche orientée vers la recherche peu fructueuse de nouveaux cas d'usages et de solutions informatiques d'hébergement ou d'accès aux bases de données. Il ne s'est pas appuyé non plus sur une cartographie préalable des différentes bases de données. La DSS, le Gip-MDS ou encore le Centre sécurisé d'accès aux données (CASD), qui jouent pourtant chacun un rôle clé dans l'accès aux données de la DSN, n'ont pas été associés à ce cadrage. La mission, qui a constaté qu'il n'existe pas de modèle-type de « *hub* » de données administratives, estime que l'objet inscrit dans la feuille de route « données » du ministère en charge du Travail devrait être réorienté vers une réponse aux principaux freins rencontrés par les utilisateurs des données de la DSN.

[9] **La relative sous-utilisation des données de la DSN à des fins de pilotage des politiques publiques s'explique d'abord par une application restrictive d'un cadre juridique qui n'a pas été pensé pour cet usage.** Le décret du 28 mars 2013 ne retient pas explicitement le pilotage des politiques publiques parmi les finalités associées à l'usage des données de la DSN. En outre, ce

¹ Prime exceptionnelle de pouvoir d'achat.

² Table à multiples dimensions.

texte renvoie à un arrêté « filtre » la définition, par destinataire et par catégorie de données, des informations transmises et des finalités, souvent restrictives, selon lesquelles elles peuvent être utilisées. À titre d'exemple, la Délégation générale à l'emploi et à la formation professionnelle (DGEFP) n'est autorisée à utiliser les données de la DSN que pour « *suivre l'évolution des quotas de contrats d'apprentissage* ». Cet arrêté, construit sur une approche très granulaire des données, doit être révisé dès que la norme et les besoins des destinataires évoluent. Or ce processus de révision, imbriqué avec les évolutions annuelles de la norme DSN, repose sur des moyens limités à la DSS, ce qui ralentit l'accès aux données de la DSN. À titre d'exemple, un délai de près de deux ans s'est écoulé entre l'expression par la Direction générale du Travail (DGT) d'un besoin d'accès à des données individuelles et sa prise en compte dans une nouvelle version de l'arrêté « filtre ». D'autres acteurs qui souhaitent accéder aux données agrégées de la DSN (ex : opérateurs de compétence, start-up d'État...) ont témoigné de leur difficulté à obtenir un retour de la part de la DSS dans un délai compatible avec leurs besoins, conduisant à des solutions de contournement peu satisfaisantes (par exemple, via la Dnum qui n'est pas destinataire en droit de ces données).

[10] Un accompagnement des utilisateurs de la DSN est nécessaire pour deux raisons principales :

- ils doivent être correctement informés de la qualité des données issues de la DSN, encore inégale selon les rubriques, même si elle progresse : à titre d'exemple, les contrats aidés marchands resteraient sous-estimés en DSN d'environ deux tiers ;
- ils doivent être aiguillés vers la base de données la plus pertinente par rapport à l'usage envisagé : alors qu'il existe plusieurs bases issues des flux déclaratifs de la DSN, aucun catalogue ne précise aujourd'hui leurs avantages et inconvénients respectifs. Certaines, constituées de données brutes, sont très proches de la déclaration native tandis que d'autres, issues du SSP, contiennent des données redressées et enrichies de concepts et d'indicateurs statistiques construits afin de permettre des analyses de l'évolution du marché du travail.

[11] Or les besoins des utilisateurs des données issues de la DSN ne sont pas aujourd'hui pleinement satisfaits. En effet, la DSS tend à orienter vers le Gip-MDS l'ensemble des utilisateurs dont la demande est jugée recevable sur le plan juridique. Le Gip-MDS développe certes depuis 2019 une offre de services « *data* » qui fait partie des cinq axes stratégiques inscrits dans sa feuille de route 2022-2025. Mais le recours à ces données brutes n'est pas toujours pertinent, par exemple quand il s'agit d'apprécier l'impact d'un dispositif sur une trajectoire d'insertion professionnelle, qui suppose de mobiliser des concepts comme le retour à un emploi de qualité, ou d'identifier les tensions de recrutement dans un secteur donné. Les utilisateurs des données de la DSN à des fins de pilotage ne sont pas non plus, en l'absence de documentation suffisante et de conseil systématique, suffisamment informés de leurs limites en termes de qualité. Enfin, il n'existe pas aujourd'hui de « communauté » des utilisateurs des données de la DSN à des fins de pilotage qui favoriserait les retours d'expérience et l'émergence de nouveaux cas d'usage.

[12] Pour favoriser le recours à la DSN à des fins de pilotage des politiques publiques, la mission recommande de repenser la gouvernance de ces usages.

[13] Cette gouvernance doit d'abord reposer sur un cadre juridique assoupli. Afin d'éviter des comportements de contournements et des interprétations divergentes, la mission préconise une suppression de l'arrêté « filtre » et une révision du décret relatif à la DSN. Sur la base d'une

approche plus générique, ce décret doit définir les catégories de données redistribuées aux différents acteurs et autoriser explicitement leur usage à des fins de pilotage. Si elle devra nécessairement faire l'objet d'échanges avec la Cnil, cette évolution est jugée possible par la Direction des affaires juridiques des ministères sociaux, sous réserve que les responsables de traitement se conforment au Règlement général sur la protection des données (RGPD), en respectant notamment les finalités du traitement DSN et en n'utilisant que les données strictement nécessaires par rapport à l'objectif recherché.

[14] Cette gouvernance doit aussi reposer sur la désignation d'une entité cheffe de file avec trois schémas envisageables selon le « centre de gravité » susceptible d'apporter le plus de valeur :

- **un centre de gravité « protection sociale »** où, aujourd'hui, la DSS et, demain, la « mission interministérielle en charge de la collecte et de l'exploitation des données sociales », en cours de préfiguration, continuerait d'exercer un rôle premier en matière de régulation de l'accès aux données ;
- **un centre de gravité « data »** où la Dnum des ministères sociaux, dans la continuité du cadrage du « *Labor data hub* », et au titre de son rôle d'administratrice ministérielle des données, des algorithmes et des codes-sources (Amdac) du ministère chargé du Travail, orchestrerait ces fonctions ;
- **un centre de gravité « statistiques emploi/travail »** où la Dares, s'appuyant sur son expertise en matière d'utilisation des données de la DSN et sa connaissance des enjeux métiers du ministère, développerait une nouvelle offre de services davantage axée sur l'appui au pilotage, en mobilisant la DSN mais aussi potentiellement d'autres données administratives dans le champ travail-emploi-formation professionnelle.

[15] La mission préconise de retenir le troisième scénario avec la création d'une « cellule » spécifique, qui pourrait être désignée « *Labor data hub* », auprès de la direction de la Dares. Cette direction dispose en effet, en tant que service statistique ministériel, d'une expertise en matière de concepts et de techniques statistiques et d'une maîtrise des limites éventuelles d'usage des données sur certains champs de la déclaration. Elle dispose aussi d'une bonne connaissance « métier » de l'ensemble du champ travail-emploi-formation professionnelle ainsi que des techniques d'évaluation, la rendant à même de s'approprier les expressions de besoin des différents acteurs et de proposer les solutions adaptées. Cette cellule fédérerait les différents acteurs concernés et son comité de pilotage pourrait être présidé par la Dnum au titre de sa fonction d'administration des données.

[16] Composée de deux à trois ETP, ce qui suppose un relèvement du plafond d'emploi de la Dares, cette cellule assurerait :

- l'élaboration d'une cartographie et d'une documentation sur les bases de données issues de la DSN ;
- l'instruction des besoins d'accès aux bases DSN, en vérifiant le cas échéant la solution adaptée aux besoins des demandeurs ;
- la conception et le développement d'une offre de services d'appariement des bases DSN avec d'autres bases ;

- le conseil auprès des utilisateurs pour veiller à la conformité des usages avec le RGPD et à la bonne utilisation des données conformément à une charte qui visera à éviter un mésusage des données brutes pouvant conduire à un risque de perte de crédibilité ou de confusion des messages ;
- l'animation de la « communauté des utilisateurs » de la DSN à des fins de pilotage, en collaboration étroite avec la Dnum eu égard au rôle de cette dernière dans la stratégie de transformation numérique et de valorisation des données du ministère.

[17] Cette « communauté des utilisateurs » à des fins de pilotage sera constituée, dans un premier cercle, des acteurs qui font aujourd'hui partie des destinataires en droit des données individuelles de la DSN (DGEFP, DGT, Pôle emploi, Agence de services et de paiement, Agefiph, Caisse des dépôts et consignations, DGCS, DGOS) et, dans un deuxième cercle, d'acteurs avec une mission de service public qui souhaitent accéder à des fins de pilotage à des données agrégées, voire à des données individuelles pseudonymisées, issues de la DSN (autres membres du service public de l'emploi, administrations externes aux ministères sociaux comme la Direction générale des entreprises, start-up d'État, opérateurs de compétences...).

[18] **Un service d'appariement centré sur la mesure du retour à l'emploi pourrait être proposé par ce « Labor data hub » en lien avec le CASD.** En effet, alors qu'il s'agit de l'un des cas d'usage majeurs rencontrés par la mission et qu'il concerne de nombreux acteurs (DGEFP, Conseils départementaux, start-up d'État, Opco...), il n'existe pas d'offre standard pour un tel service. L'appariement serait défini par le « Labor data hub » (données en entrée, indicateurs en sortie...) et opéré par le CASD qui dispose d'une procédure sécurisée déjà utilisée dans le cadre de projets de recherche ou d'évaluation des politiques publiques.

[19] **Cette proposition de gouvernance contribue à tirer le meilleur parti des expertises respectives tout en évitant une dispersion des moyens.** L'ancrage de cette gouvernance au sein du MTPEI aurait pour corollaire un recentrage de l'offre de services « data » du Gip-MDS sur la fiabilisation en amont de la DSN, l'exploitation aval des données brutes ne devant s'envisager que pour des besoins de suivi très précoces et ne nécessitant pas de faire appel à des concepts relatifs à la sphère emploi-travail-formation professionnelle. Le « Labor data hub » devra aussi définir ses modalités de fonctionnement avec la mission interministérielle en cours de préfiguration à la DSS.

LISTE DES RECOMMANDATIONS

n°	Recommandations	Acteur responsable	Acteurs associés
Juridique			
1	Réviser le décret n° 2013 266 du 28 mars 2013 relatif à la DSN sur la base d'une approche plus générique des usages en introduisant la finalité de pilotage des politiques publiques et abroger l'arrêté « filtre » du 9 février 2022	DSS	DAJ Cellule LDH
2	Veiller, en lien avec le délégué à la protection des données (DPO), à la bonne alimentation du registre des traitements effectués à partir de la DSN, centraliser les Analyses d'impact relatives à la protection des données (AIPD) concernant leur usage et mieux accompagner les utilisateurs dans la prise en compte effective des règles issues du Règlement général sur la protection des données (RGPD)	Cellule LDH	DPO
Outils et accompagnement des acteurs			
3	Mettre à disposition des utilisateurs un catalogue des bases de données construites à partir de la DSN précisant les traitements effectués ainsi que leurs avantages ou inconvénients respectifs dans le cadre de différents usages	idem	Producteurs des bases de données
4	Élaborer une charte commune aux différents utilisateurs des données issues de la DSN à des fins de pilotage au sein de l'administration	idem	-
5	Créer un cadre d'échanges sur les questions éthiques soulevées par l'utilisation des données de la DSN, en particulier en cas de recours à l'intelligence artificielle	idem	-
6	Organiser des « hackathons » ou autres temps de co-construction sur la donnée, le cas échéant en mobilisant le « datalab » de la Dinum, pour faire émerger des cas d'usage de la DSN à des fins de pilotage dans le champ travail-emploi-formation	DNUM	Cellule LDH
Gouvernance			
7	Créer une cellule dénommée Labor data hub (LDH) composée de 2-3 ETP auprès du directeur de la Dares responsable, en lien avec la Dnum, d'orchestrer et de mettre en œuvre les fonctions nécessaires à une meilleure utilisation des données de la DSN à des fins de pilotage	MTPEI	DSS

n°	Recommandations	Acteur responsable	Acteurs associés
8	Concevoir un service de croisement de données (appariements) permettant de mesurer de façon efficiente et cohérente l'impact sur le marché du travail, notamment le retour à l'emploi, de dispositifs de politiques publiques	Cellule LDH	CASD
9	Recentrer l'offre de services « <i>data</i> » du Gip-MDS sur la fiabilisation amont de la déclaration et sur l'exploitation précoce de données brutes présentes en DSN ne nécessitant pas de faire appel à des concepts	DSS	DGEFP

SOMMAIRE

SYNTHESE	3
LISTE DES RECOMMANDATIONS	9
SOMMAIRE	11
INTRODUCTION	15
1 LA DSN CONSTITUE UN LEVIER POUR AMELIORER LA CONNAISSANCE DU MARCHÉ DU TRAVAIL ET LE PILOTAGE DES POLITIQUES D'EMPLOI	19
1.1 LA DSN REPRESENTE UN GISEMENT TRÈS RICHE QUOIQU'ELLE SOIT PERFECTIBLE DE DONNÉES SUR LE MARCHÉ DU TRAVAIL	19
1.1.1 <i>Les données de la DSN sont issues d'un processus déclaratif qui alimente plusieurs bases de données aux caractéristiques différentes</i>	19
1.1.2 <i>La fiabilisation des déclarations et la qualité des données, inégale selon les rubriques de la DSN, constituent des enjeux importants pour la statistique et le pilotage</i>	20
1.2 LA DSN A ENGENDRÉ UNE REFORME DES SYSTÈMES D'INFORMATION ET DE L'ORGANISATION DU SERVICE STATISTIQUE PUBLIC QUI PERMET DE RACCOURCIR LES DÉLAIS DE PRODUCTION	22
1.2.1 <i>Le service statistique public a engagé une réforme majeure de ses systèmes d'information dans un cadre de coordination accrue</i>	22
1.2.2 <i>La démarche qualité mise en œuvre par le service statistique public éclaire les fragilités des données brutes de la DSN</i>	26
1.2.3 <i>La production de bases DSN mensuelles ouvre de nombreuses perspectives d'appariements pour analyser les dynamiques d'insertion professionnelle</i>	27
1.2.4 <i>Il est envisagé d'utiliser les traitements effectués par les statisticiens publics à des fins de pilotage</i>	28
1.3 L'UTILISATION DE LA DSN À DES FINS DE PILOTAGE DANS LE CHAMP TRAVAIL-EMPLOI-FORMATION MONTE EN PUISSANCE	28
1.3.1 <i>La DSN permet d'apprécier l'impact des interventions publiques et de les ajuster plus rapidement</i>	28
1.3.2 <i>L'utilisation des données de la DSN à des fins de pilotage se diffuse progressivement</i>	29
2 LES DONNÉES DE LA DSN RESTENT SOUS-UTILISÉES EN RAISON NOTAMMENT D'UN CADRE ET D'UNE ORGANISATION INADAPTES	33
2.1 LE CADRE JURIDIQUE NE FAVORISE PAS L'ACCÈS AUX DONNÉES DE LA DSN ET FAIT DÉBAT À L'AUNE DE LA POLITIQUE PUBLIQUE DE LA DONNÉE	33
2.1.1 <i>Les textes réglementaires n'ouvrent pas suffisamment la voie à des usages des données de la DSN à des fins de pilotage</i>	33
2.1.2 <i>La politique publique de la donnée relancée en 2021 questionne tant la légalité que la pertinence de ce cadre juridique</i>	35
2.2 L'ORIENTATION ET L'ACCOMPAGNEMENT DES UTILISATEURS DES DONNÉES ISSUES DE LA DSN SONT INSUFFISAMMENT DÉVELOPPÉS	37
2.2.1 <i>Il n'existe pas de fonction d'aiguillage des utilisateurs vers la base de données la plus pertinente par rapport à leurs besoins</i>	37

2.2.2	<i>Les utilisateurs, y compris potentiels, des données issues de la DSN ont des besoins d'accompagnement</i>	38
2.3	LE « LABOR DATA HUB » RESTE A CE STADE UN OBJET ADMINISTRATIF NON IDENTIFIE	40
2.3.1	<i>Prévu par la feuille de route « données » du ministère en charge du Travail, le « Labor data hub » visait en particulier à favoriser l'utilisation de la DSN</i>	40
2.3.2	<i>Le cadrage du « Labor data hub » ne fournit pas les éléments d'appréciation nécessaires à un arbitrage sur son devenir</i>	40
2.3.3	<i>Il n'existe pas de modèle-type de « hub » ministériel de données</i>	41
2.3.4	<i>Le CASD représente aujourd'hui le principal « hub » d'accès aux données statistiques issues de la DSN</i>	42

3 POUR FAVORISER LA BONNE UTILISATION DES DONNEES DE LA DSN, UNE GOUVERNANCE SPECIFIQUE DOIT ETRE MISE EN PLACE 43

3.1	LA BONNE UTILISATION DES DONNEES DE LA DSN PEUT ETRE APPRECIEE A L'AUNE DE DIFFERENTS PRINCIPES	43
3.1.1	<i>Le cadre juridique de la DSN doit faciliter l'utilisation de ses données à des fins de pilotage</i>	43
3.1.2	<i>La DSN contient des données sensibles qui exigent un traitement conforme au droit de la protection des données personnelles</i>	44
3.1.3	<i>Le choix d'une base de données construite à partir de la DSN doit reposer sur une documentation permettant d'arbitrer selon les caractéristiques recherchées</i>	45
3.1.4	<i>Les limites des analyses réalisées à partir de données brutes issues de la DSN doivent être prises en compte dans leur communication éventuelle</i>	45
3.1.5	<i>L'utilisation des données issues de la DSN doit s'accompagner d'une réflexion éthique</i>	46
3.1.6	<i>En amont de l'accès aux données issues de la DSN, il convient de faire émerger ses utilisations les plus pertinentes</i>	46
3.2	POUR AMELIORER L'UTILISATION DES DONNEES DE LA DSN, SIX FONCTIONS DOIVENT ETRE CREEES OU RENFORCEES	47
3.2.1	<i>Instruire aux niveaux technique et juridique les besoins d'accès secondaires aux données issues de la DSN</i>	47
3.2.2	<i>Orienter et accompagner les utilisateurs dans le recours à la base de données la plus pertinente par rapport à leurs besoins</i>	47
3.2.3	<i>Contribuer à une utilisation des données de la DSN respectueuse des données personnelles et éthique</i>	47
3.2.4	<i>Animer la communauté des utilisateurs de données issues de la DSN pour favoriser les retours d'expérience et susciter de nouveaux usages</i>	47
3.2.5	<i>Veiller à l'efficacité et à la cohérence globales des traitements issus de la DSN à des fins statistiques et de pilotage</i>	48
3.2.6	<i>Recueillir les besoins d'évolution du décret relatif à la DSN liés au pilotage et s'assurer de leur prise en compte par la DSS</i>	48
3.3	TROIS SCHEMAS DE GOUVERNANCE SONT PROPOSES POUR ORCHESTRER CES FONCTIONS	48
3.3.1	<i>Il est indispensable d'identifier une direction cheffe de file pour orchestrer ces fonctions et incarner cette gouvernance</i>	48
3.3.2	<i>Une gouvernance avec un centre de gravité « protection sociale »</i>	49
3.3.3	<i>Une gouvernance avec un centre de gravité « data »</i>	49
3.3.4	<i>Une gouvernance avec un centre de gravité « statistiques emploi-travail »</i>	50

3.4 UN « <i>LABOR DATA HUB</i> », REDEFINI SUR LA BASE DE CES FONCTIONS, POURRAIT ETRE POSITIONNE AUPRES DE LA DARES ET OFFRIR UN SERVICE D'APPARIEMENT	52
3.4.1 <i>Une équipe projet constituée de deux à trois ETP positionnés auprès de la Dares pourrait constituer le « Labor data hub »</i>	52
3.4.2 <i>Un service d'appariement centré sur la mesure du retour à l'emploi pourrait être offert par ce « Labor data hub » en lien avec le CASD</i>	54
ANNEXE 1 : CARACTERISATION DES BASES DE DONNEES ISSUES DE LA DSN	57
ANNEXE 2 : L'UTILISATION DES DONNEES DE LA DSN A DES FINS DE STATISTIQUES PUBLIQUES	88
ANNEXE 3 : UTILISATION DES DONNEES DE LA DSN A DES FINS DE PILOTAGE DES POLITIQUES DU TRAVAIL, DE L'EMPLOI ET DE LA FORMATION PROFESSIONNELLE	119
ANNEXE 4 : ENCADREMENT DE L'ACCES AUX DONNEES DSN ET DE LEURS USAGES	154
LISTE DES PERSONNES RENCONTREES	181
LETTRE DE MISSION	188

RAPPORT

Introduction

[1] L'Inspection générale de l'Insee et l'Inspection générale des affaires sociales (Igas) ont été saisies en octobre 2022 par le cabinet du ministre du Travail, du Plein emploi et de l'Insertion (MTPEI) d'une mission sur les usages statistiques et à des fins de pilotage de la Déclaration sociale nominative (DSN) dans le champ du travail, de l'emploi et de la formation professionnelle³. Cette saisine constituait le prolongement d'une commande confiée initialement à la seule inspection générale de l'Insee dans un périmètre davantage centré sur le Service statistique public (SSP). En s'appuyant sur les premiers échanges qui s'étaient tenus avec les représentants du SSP, la mission a engagé ses travaux fin novembre 2022. Elle a élargi le cercle des acteurs rencontrés à ceux utilisant la DSN à des seules fins de pilotage. Elle s'est attachée à instruire également le projet de création d'un « *Labour data hub* » prévu par la feuille de route « données » 2021-2023 du MTPEI.

[2] La DSN est un dispositif déclaratif mensuel dématérialisé et assis sur la paie. Il simplifie la communication par les employeurs des informations sur leurs salariés nécessaires à la gestion de leur protection sociale. Créée en 2012⁴, la déclaration a vu son périmètre progressivement étendu à la totalité des entreprises du secteur privé en 2017 puis à l'ensemble du secteur public en 2022. Seuls les salariés des particuliers employeurs restent exclus à ce stade de ce dispositif, ainsi que les travailleurs indépendants qui par définition ne relèvent d'aucun employeur. La DSN s'est substituée à de plus en plus de formalités administratives, soit une quarantaine aujourd'hui contre une vingtaine lors de sa création. Depuis 2019, ce dispositif est aussi le vecteur de déclaration du prélèvement à la source. Enfin, il alimente, en complément du dispositif Pasrau relatif aux revenus de remplacement, le Dispositif ressources mensuelles (DRM), développé en 2019 dans le cadre de la réforme des aides au logement. Le DRM constitue un traitement de données distinct de la DSN et exclu à ce titre du périmètre de cette mission.

[3] La DSN prend la forme non pas d'une, mais de plusieurs bases de données. Ces données sont constituées des flux d'information issus des déclarations effectuées par les employeurs sur le portail « Net-entreprises » ou le portail de la CCMSA⁵. Les bases en aval de la déclaration se distinguent, d'une part, par les retraitements éventuellement effectués au préalable en vue d'améliorer la qualité des données et, d'autre part, par la fraîcheur de ces données. Alors que les destinataires de chaque formalité géraient auparavant les données qui en étaient issues, la DSN est une déclaration unifiée, dont les informations sont redistribuées aux acteurs métiers.

³ L'Insee étant sous tutelle du ministère en charge de l'Économie, cette lettre de mission aurait également dû être signée par ce cabinet.

⁴ Loi n°2012-387 du 22 mars 2012 relative à la simplification du droit et à l'allègement des démarches administratives.

⁵ Caisse centrale de la mutualité sociale agricole.

[4] La DSN constitue un gisement très riche d'informations sur le marché du travail et l'emploi. Elle représente 26 M d'actes déclaratifs par mois, contenant environ 150 variables par salarié pour une DSN du Régime général. L'exhaustivité et la fraîcheur des informations fournies sont supérieures à celles des déclarations administratives auxquelles la DSN s'est substituée, notamment la Déclaration annuelle de données sociales (DADS). Il s'agit aussi d'un gisement encore relativement jeune dont la qualité des données est inégale selon les rubriques.

[5] La DSN est un traitement de données à caractère personnel au sens de la Cnil. Elle inclut par ailleurs des données sensibles, telles que définies par la loi « informatique et libertés », concernant la qualité de travailleur handicapé ou encore la perception d'indemnités journalières. Les données sur les rémunérations salariales individuelles doivent aussi être protégées, parce qu'elles sont personnelles et relatives à la politique RH des employeurs.

[6] La lettre de mission porte sur deux types d'usages de la DSN, à des fins d'élaboration de statistiques publiques et à des fins de pilotage des politiques publiques travail-emploi-formation, ainsi que sur leur articulation. L'utilisation des données de la DSN à des fins de ciblage des contrôles et de personnalisation des services proposés peut être rattachée au pilotage des politiques publiques dans une acception large. En revanche, le recours aux données de la DSN aux fins de simplification de la gestion, par exemple en substitution d'informations qui étaient auparavant déclarées par les employeurs via d'autre canaux, est exclu du périmètre de la mission.

[7] Les statistiques publiques, définies par la loi du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques, recouvrent l'ensemble des statistiques diffusées à des fins d'information générale, et en toute indépendance professionnelle. Elles sont produites par l'Insee et les Services statistiques ministériels (SSM), qui constituent le Service statistique public (SSP)⁶. Le pilotage des politiques publiques recouvre une vaste gamme d'usages qui visent à apprécier les effets des actions engagées et à ajuster les modalités d'intervention des acteurs, sans que l'utilisation des données à cette fin ait *a priori* vocation à être rendue publique.

[8] La lettre de mission soulève les questions suivantes :

- quel est l'état des lieux des usages de la DSN à des fins de statistiques publiques ?
- quels sont les indicateurs produits sur la base de la DSN par les différents acteurs y ayant recours à des fins d'information générale ou de pilotage opérationnel ?
- faut-il faire évoluer le cadre d'utilisation de ces données à des fins de pilotage ?
- comment mieux mettre à disposition ces données , à travers par exemple le « *Labor data hub* » ?
- quelle adéquation entre les besoins et l'offre du SSP, et celle des autres opérateurs ?

⁶ Depuis 2011, l'Autorité de la statistique publique (ASP) reconnaît dans le cadre de la labellisation une sphère plus large de statistiques émanant d'organismes ayant des missions de service public. Dans le domaine de l'emploi et des rémunérations salariales, l'Urssaf Caisse nationale et la CCMSA ont ainsi été labellisées par l'ASP pour une partie de leurs productions statistiques, qui viennent en complément des statistiques du SSP et dont l'élaboration et la diffusion sont jugées par l'Autorité au Code des bonnes pratiques de la statistique européenne. L'ensemble de ces acteurs constitue ainsi le SSP élargi.

[9] La relance en 2021 d'une politique gouvernementale de la donnée a renforcé l'acuité de ces questions. La feuille de route « données » 2021-2023 du MTPEI vise en particulier une meilleure valorisation des données de la DSN pour améliorer la conception, le pilotage et l'évaluation des politiques publiques. Elle prévoit la création d'un « *Labor data hub* » défini comme « *une plateforme de données et de services du travail, de l'emploi et de la formation pour faciliter le partage et promouvoir l'utilisation des données au sein de la sphère sociale* ». Ce projet de plateforme a fait l'objet d'un cadrage par la direction du numérique des ministères sociaux, qui assume la fonction d'administration des données pour le MTPEI. L'enjeu d'un pilotage renforcé et plus réactif de l'impact des politiques d'emploi est aussi au cœur du projet de création de « France Travail ».

[10] Pour répondre à ces différentes questions, la mission a échangé avec les acteurs participant à la production des données issues de la DSN, à la régulation de l'accès à ces données ou à leur utilisation à des fins de statistiques publiques ou de pilotage. Pour identifier les usages de la DSN, mais aussi cerner son niveau de « sous-utilisation », elle a notamment rencontré :

- le Gip-MDS, responsable du dispositif déclaratif dans son ensemble, et la CNAV qui assure la redistribution des flux dont sont issues les données individuelles de la DSN, ainsi que la Direction de la sécurité sociale (DSS), en charge de la tutelle de ces deux opérateurs. La mission s'est également appuyée sur le rapport de l'Igas sur la gouvernance du Gip-MDS⁷ qui a contribué à la création d'une mission interministérielle en charge des projets de collecte et d'exploitation des données nécessaires à la gestion des prestations sociales⁸, qui sera positionnée dans les prochaines semaines auprès de la Direction de la sécurité sociale ;
- l'Insee et la Dares, en charge de produire les principaux livrables statistiques fondés sur la DSN mais aussi, pour la seconde, en tant que SSM ayant des fonctions d'appui du ministère dans le suivi et l'évaluation des politiques, ainsi que l'Urssaf Caisse nationale, qui contribue à la production de statistiques publiques à partir de la DSN ;
- le Centre d'accès sécurisé aux données (CASD) qui constitue la principale plateforme de mise à disposition de données statistiques issues de la DSN et qui joue un rôle spécifique en matière d'appariements avec d'autres bases de données ;
- les principales directions des ministères sociaux utilisant les données de la DSN à des fins de pilotage (Direction générale du Travail, Délégation générale à l'emploi et à la formation professionnelle, Direction générale de la cohésion sociale, Direction générale de l'offre de soins), la Direction du numérique mais aussi d'autres entités publiques qui utilisent ou cherchent à utiliser ces données : opérateurs tels que Pôle emploi, autres gestionnaires de régimes et dispositifs tels que la Caisse des dépôts et consignations, start-ups d'État (« Plateforme de l'inclusion », « Place des entreprises », « Immersion facilitée », « La Bonne

⁷ Rapport 2022-021R d'Eric Ginesy et Laurent Gratieux, « *La gouvernance du Gip-MDS et des dispositifs de déclaration sociales* », juillet 2022.

⁸ Selon ce rapport, les sujets d'arbitrage par l'État restent fréquents et complexes, ce qui justifie la restauration d'une maîtrise d'ouvrage stratégique interministérielle, positionnée auprès de la DSS qui assure de facto cette fonction depuis 2018, mais ne peut durablement arbitrer seule des évolutions qui concernent un nombre croissant de départements ministériels, en particulier la sphère emploi-travail-formation professionnelle. Un avis de vacance du poste d'expert de haut niveau chargé du pilotage des systèmes de collecte et d'utilisation des données sociales, qui serait responsable de cette nouvelle mission, a été publié le 12 février 2023 au Journal officiel.

alternance », « Diagorienté », « Signaux faibles ») ou encore, hors des ministères sociaux, la Direction générale des entreprises ;

- des acteurs extérieurs à l'administration intervenant dans le champ travail-emploi-formation professionnelle et utilisant ou souhaitant utiliser les données de la DSN, en particulier des opérateurs de compétences (Opcv) et des observatoires de branches professionnelles ainsi que le réseau des Carif-Oref ;
- France Stratégie et des utilisateurs de la sphère études-recherche, dont un chercheur participant à l'évaluation des politiques publiques dans le champ du travail, de l'emploi et de la formation professionnelle.

[11] Enfin, la mission s'est attachée à collecter des éléments de comparaison en matière de valorisation des données. Elle n'a pas identifié de modèle étranger dans le champ travail-emploi-formation. Néanmoins, elle a échangé avec la Direction interministérielle du numérique (Dinum) en charge de favoriser la circulation et l'utilisation des données dans le secteur public. Elle a également rencontré les représentants de trois « hubs » de données au sein de l'État (« *Health data hub* », « *Bercy hub* », projet IDEE porté par le ministère de l'Éducation nationale), ce qui lui a permis de conclure à l'absence d'un modèle type qui s'imposerait *de facto*.

[12] Le présent rapport, qui s'accompagne de quatre annexes⁹, est organisé en trois parties :

- la première met en avant le potentiel que représente la DSN pour améliorer la connaissance et le fonctionnement du marché du travail grâce à ses usages à des fins de statistiques publiques et de pilotage ;
- la deuxième montre que la relative sous-utilisation de la DSN à des fins de pilotage s'explique en particulier par une approche restrictive du cadre juridique et une organisation inadaptée ;
- enfin, la troisième propose de créer ou renforcer des fonctions déclinables dans trois schémas possibles de gouvernance afin de favoriser une bonne utilisation des données de la DSN.

⁹ Annexes relatives à la caractérisation des bases de données issues de la DSN (n°1), à l'utilisation des données de la DSN à des fins de statistiques publiques (n°2), à l'utilisation de la DSN à des fins de pilotage (n°3), à l'encadrement de l'accès aux données de la DSN et à leur utilisation (n°4).

1 La DSN constitue un levier pour améliorer la connaissance du marché du travail et le pilotage des politiques d'emploi

1.1 La DSN représente un gisement très riche quoique perfectible de données sur le marché du travail

1.1.1 Les données de la DSN sont issues d'un processus déclaratif qui alimente plusieurs bases de données aux caractéristiques différentes

[13] **Les données issues de la DSN proviennent de flux déclaratifs fondés sur une norme qui est à la fois un cahier des charges pour les déclarants et une clé de lecture pour leurs utilisateurs.** Les employeurs transmettent au 5 ou au 15 de chaque mois via leur logiciel de paie un flux de données correspondant au mois M-1. Ce flux est construit à partir d'une norme dite NEODeS¹⁰. Elle est portée par un cahier technique qui est mis à jour chaque année en fonction des besoins des utilisateurs. Ce cahier définit la nomenclature et le format des données, qui sont regroupées dans des rubriques, ainsi que les contrôles, bloquants ou non, opérés dans le cadre du processus déclaratif. Il apporte donc des indications sur la sémantique des données brutes et les règles applicables aux déclarants, qui sont particulièrement utiles aux statisticiens même si elles ne leur sont pas destinées en première intention.

[14] **La DSN contient à la fois des données agrégées au niveau de l'établissement et des données individuelles relatives aux salariés, qui sont réparties entre les organismes utilisateurs.** Une fois les déclarations effectuées par les employeurs sur le portail Net-entreprises ou celui de la CCMSA, ces flux de données alimentent des « blocs », c'est-à-dire des modules techniques articulés entre eux. Ces blocs composent l'architecture du système DSN. Les données agrégées au niveau de l'établissement constituent le bloc 1. Elles sont destinées aux organismes chargés du recouvrement. Les données individuelles, qui constituent le bloc 3, sont adressées à la Cnav pour redistribution aux organismes destinataires¹¹. Elles incluent des données nominatives relatives, en particulier, au lieu d'activité du salarié, aux caractéristiques de son contrat de travail, au montant de sa rémunération, à sa durée de travail, aux cotisations et contributions versées et à tout événement relatif à sa situation en emploi (embauche, licenciement, arrêt maladie, etc.).

[15] **Les organismes destinataires de données de la DSN opèrent des traitements métiers et constituent en aval des bases qui présentent différentes caractéristiques en lien avec leurs usages respectifs.** La mission a analysé sept bases de données utilisées à des fins de statistiques publiques ou de pilotage (cf. annexe n°1). Elle met en évidence des critères permettant d'apprécier la pertinence d'une base de données à l'aune des besoins des utilisateurs, parmi lesquels :

- sa finalité, certaines bases étant conçues pour un usage à des fins de statistiques publiques ou d'évaluation ;

¹⁰ Norme d'échanges optimisée des données sociales.

¹¹ La Cnav assure en outre la certification du Nir (numéro de sécurité sociale) et la conservation des données pendant cinq ans.

- le périmètre couvert, qui peut concerner les seuls salariés et agents couverts par la DSN ou être étendu, via des retraitements spécifiques, aux salariés des particuliers-employeurs ;
- les unités élémentaires utilisées pour compiler les données de la DSN, qui sont, à titre principal, des contrats, des postes, des salariés, des établissements ou des entreprises ;
- la part des données expertisées, c'est-à-dire ayant fait l'objet de contrôles de cohérence et de retraitements permettant d'améliorer la qualité des informations véhiculées ;
- l'accessibilité de la documentation relative au contenu de la base et à ses méthodes de traitement ;
- la fraîcheur des données, le retraitement des données effectué dans certaines bases impliquant, un délai de mise à disposition supplémentaire (par rapport aux bases ne contenant que des données brutes) allant de quelques jours à plus d'un an ;
- les concepts accessibles, qui correspondent à des informations supplémentaires par rapport à celles contenues dans la DSN, comme l'emploi, le cas échéant en équivalent temps plein, la masse salariale ou encore le salaire individuel (net, brut, horaire, annuel...).

1.1.2 La fiabilisation des déclarations et la qualité des données, inégale selon les rubriques de la DSN, constituent des enjeux importants pour la statistique et le pilotage

[16] **La qualité des données issues de la DSN doit être appréciée à l'aune de leurs usages.** Le Gip-MDS analyse la qualité de la DSN sous l'angle de la conformité des DSN déposées par rapport à la norme NEODES. Il estime ainsi le taux global de conformité de la DSN à 94 %¹². En pratique, ce taux de conformité renvoie aux contrôles de la norme embarqués dans le processus déclaratif. Pour faire progresser ce taux de conformité, le Gip-MDS anime des ateliers de surveillance de la qualité. Les axes de surveillance mis en place visent à vérifier que la structure de la déclaration est correcte et conforme aux valeurs attendues ainsi qu'aux contrôles internes portés par le cahier technique. Toutefois, pour ne pas compromettre le processus déclaratif, ces contrôles ne sont pas nécessairement bloquants, ni donc susceptibles de repérer certaines erreurs déclaratives. À titre d'exemple, la déclaration en DSN d'une entreprise qui omettrait de déclarer l'embauche d'un salarié sous un statut d'apprenti, ou encore une fin de contrat de travail, peut pour autant être conforme à la norme.

[17] **La fiabilisation des déclarations DSN constitue un enjeu commun à l'ensemble des utilisateurs.** Comme l'a souligné le Haut conseil au financement de la protection sociale¹³, la fiabilisation des données individuelles est d'autant plus importante qu'elles contribuent de façon croissante à déterminer les droits sociaux des assurés, qu'ils soient contributifs (ex : chômage, retraite) ou non contributifs (ex : allocations logement). Le développement de la solidarité à la source justifie la mise en place de démarches de fiabilisation auprès des employeurs, y compris en aval du processus déclaratif, telles que prévues par la loi de financement de la sécurité sociale

¹² Ce taux correspond au nombre de DSN déposées conformes au niveau du point de dépôt du système DSN par rapport au nombre total de DSN déposées sur la période.

¹³ HCFiPS, « L'avenir du recouvrement social », juillet 2022.

pour 2020¹⁴. Pour autant, une démarche de fiabilisation ne garantit pas une amélioration automatique de la qualité des données utilisées à des fins de statistiques publiques ou de pilotage. D'une part, toutes les entreprises ne modifieront pas leurs comportements déclaratifs et, d'autre part, l'évolution régulière de la norme DSN peut engendrer de nouvelles erreurs.

[18] **S'il tendrait selon les utilisateurs à progresser, le niveau de qualité des données déclaratives est globalement peu documenté.** Il n'appartenait pas à la mission d'analyser de façon exhaustive la qualité des données brutes de la DSN. Néanmoins, sans prétendre à l'exhaustivité, elle a relevé en particulier les points d'attention suivants¹⁵ :

- **les données relatives à certains « dispositifs de politiques publiques »** seraient très partielles, d'autres systèmes d'information étant utilisés pour le paiement des aides : selon les analyses de la Dares, corroborées par l'ASP et la DGEFP, mi-2022, les comparaisons avec les autres systèmes d'information montrent que seuls les contrats aidés du secteur non marchand (PEC) seraient correctement déclarés en DSN (sous-estimation de l'ordre de 10 % depuis 2020-21, en amélioration par rapport aux années précédentes). Les niveaux en contrats aidés marchands (CIE) seraient en revanche très sous-estimés en déclaration DSN (d'environ deux tiers), de même que ceux des Ateliers et chantiers d'insertion (environ deux tiers), et des Entreprises d'insertion (environ trois quarts) ;
- **les données relatives aux contrats de travail** présenteraient aussi plusieurs limites tenant, soit à l'absence de déclaration des modifications intervenues, notamment des fins de contrat (par exemple, transfert d'un contrat de travail à un autre établissement, notamment dans le cas de l'intérim qui donne lieu à la création d'un nouveau contrat sans fermeture du précédent), soit à un défaut de cohérence entre la nature du contrat, le niveau de rémunération et la quotité de temps de travail déclarés, ou encore à des erreurs dans le code de convention collective déclaré ;
- **les données relatives à la localisation des emplois** peuvent aussi comporter des erreurs en cas de déclaration groupée par un seul établissement, par exemple celui de l'établissement payeur, tout comme **les données sur les professions et catégories socioprofessionnelles (PCS)**¹⁶ dont environ 20 % seraient redressées par l'Insee ;

¹⁴ La LFSS pour 2020 prévoit la création de comptes rendus métiers normalisés (CRM) permettant de signaler aux entreprises déclarantes les anomalies constatées et, à partir de 2024, la possibilité pour les organismes de protection sociale de se substituer aux déclarants en cas d'absence de correction des erreurs signalée.

¹⁵ Ces différents points ont été portés à la connaissance de la mission par les acteurs de la statistique publique. La Dares et l'Insee estiment par ailleurs qu'en l'absence de contrôles bloquants et dès lors que l'entreprise n'a pas un intérêt direct à déclarer correctement une donnée, il sera difficile d'en garantir la qualité initiale. Des expertises resteraient ainsi nécessaires, même en régime de croisière. La mission a par ailleurs noté qu'une partie des utilisateurs rencontrés n'étaient pas en mesure d'apprécier le niveau de qualité des données utilisées.

¹⁶ La nomenclature des professions et catégories socioprofessionnelles (PCS-ESE) est une nomenclature placée sous la responsabilité de l'Insee et utilisée en DSN par les employeurs pour codifier la profession de leurs salariés. Elle se présente sous la forme d'un code à quatre chiffres, le 4^{ème} chiffre correspondant à environ 400 métiers. [Nomenclatures des professions et catégories socioprofessionnelles des emplois salariés des employeurs privés et publics | Insee](#)

- **les données relatives au Finess des établissements** des champs sanitaire, médico-social et social¹⁷ du secteur public ne seraient encore renseignées qu'à hauteur d'environ 30 %.

1.2 La DSN a engendré une refonte des systèmes d'information et de l'organisation du service statistique public qui permet de raccourcir les délais de production

1.2.1 Le service statistique public a engagé une refonte majeure de ses systèmes d'information dans un cadre de coordination accrue

[19] **La DSN a constitué un « choc déclaratif » pour le service statistique public qui utilisait auparavant quatre sources administratives distinctes dans ses travaux sur l'emploi et les salaires (cf. annexe n° 2).** Les acteurs concernés ne se limitent pas à l'Insee et la Dares qui relèvent du Service statistique public (SSP) au sens de la loi fondatrice de 1951¹⁸. Il s'agit également de l'Urssaf Caisse nationale, qui contribue, avec le SSP, à « *l'exploitation à des fins d'information générale de données collectées par des administrations, des organismes publics ou des organismes privés chargés d'une mission de service public* » selon les termes de cette loi. Avant la création de la DSN, chacun de ces acteurs exploitait différemment une ou deux des quatre déclarations administratives produites par les entreprises¹⁹.

[20] **L'unification du processus déclaratif engendrée par la DSN a d'abord nécessité une refonte des systèmes d'information statistiques respectifs de ces acteurs.** En effet, si chacun est désormais destinataire d'une partie conséquente des données de la DSN, l'usage qui en est fait n'est pas le même. C'est ce qui explique le maintien de systèmes d'information distincts, même si des passerelles ont été développées pour la définition de certains concepts. La refonte de ces systèmes d'information, qui est en cours d'aboutissement, a permis d'intégrer des retraitements au fil de l'eau des données issues de la DSN :

- le Système d'information sur l'emploi et les revenus d'activités (Siera), piloté par l'Insee, permet de produire, à partir de la DSN, des données retraitées chaque trimestre, sur l'emploi et le coût du travail (salaires nets ou bruts) et, chaque année, sur l'emploi (corrigé de la multiactivité) et les salaires individuels, par le biais d'un exercice de compilation avec d'autres sources pour intégrer, jusqu'en 2022, les données du secteur public et, à ce jour encore, celles des salariés des particuliers employeurs (ainsi que par ailleurs, les données des non-salariés) ;

¹⁷ Finess (Fichier national des établissements sanitaires et sociaux) est le répertoire national de référence des personnes morales intervenant dans les domaines sanitaire, médico-social et social et de l'enseignement des professions sanitaires et sociales. Selon le prestataire Invénis, en charge du projet « Ramsès », le taux de renseignement du Finess en DSN serait de l'ordre de 30 %, ce qui complexifie la possibilité d'identifier en DSN un établissement.

¹⁸ Loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques.

¹⁹ Les déclarations et enquêtes sur les mouvements de main d'œuvre ainsi que les relevés mensuels de mission d'intérim étaient traités par la Dares, les bordereaux récapitulatifs de cotisations par l'Acoss et l'Insee, et les déclarations annuelles de données sociales par l'Insee.

- le Système d'information sur les mouvements de main d'œuvre (Sismmo), piloté par la Dares, livre des données mensuelles sur l'intérim et des données trimestrielles sur l'ensemble des embauches et des fins de contrats de travail au niveau des établissements ;
- enfin, le système d'information TTE1, piloté par l'Urssaf Caisse nationale, fournit chaque mois et chaque trimestre les effectifs des entreprises du secteur privé au sens de la Sécurité sociale ainsi que des indicateurs sur la masse salariale et les exonérations de cotisations et contributions sociales.

[21] **La DSN a contribué à une meilleure coordination entre les acteurs participant à la production de statistiques trimestrielles sur l'emploi.** En effet, depuis 2017, les estimations trimestrielles d'emploi salarié sont co-produites par l'Insee, la Dares et l'Urssaf Caisse nationale²⁰. Cette co-production a été rendue possible par un rapprochement préalable des traitements effectués sur les données issues de la DSN (par exemple, la correction des variations saisonnières) et par l'adoption d'une définition commune de l'emploi²¹. Elle donne aujourd'hui accès chaque trimestre à un diagnostic consolidé et cohérent de l'emploi sur l'ensemble du champ salarié.

[22] **Néanmoins, l'articulation entre les acteurs de la statistique publique dans leur utilisation des données de la DSN reste à préciser sur certains points.** Ils concernent en particulier la production de données sur l'emploi handicapé, qui relève historiquement de la Dares dans le cadre de l'ancienne déclaration OETH²² mais qui fait aussi l'objet de travaux de la part de l'Urssaf Caisse nationale depuis l'intégration en DSN de cette déclaration. Il s'agit aussi des rôles respectifs de l'Insee et de la Dares en matière de production de données relatives aux salaires. S'il est reconnu par tous que la mesure du salaire total (y compris primes et autres éléments de rémunération variables) doit porter sur un cycle annuel complet, des besoins existent sur des périodes infra-annuelles, notamment pour qualifier la qualité de l'emploi. L'Insee et la Dares ont depuis peu convenu de retenir dans ce cas le concept de salaire de base, qui fournit une mesure approximative de la qualité des emplois. D'autres pistes semblent envisageables en infra-annuel pour assurer un suivi de la distribution des masses salariales (primes comprises) par tête, par ETP ou par déciles. Ces indicateurs permettraient d'apprécier les évolutions dans les périodes où elles sont plus dynamiques, comme actuellement en période de forte inflation²³.

²⁰ L'Urssaf Caisse nationale produit les évolutions trimestrielles de postes dans le secteur privé (hors intérim, agriculture et particuliers employeurs), la Dares les évolutions trimestrielles de l'emploi intérimaire et, enfin, l'Insee évolutions de postes sur le champ de la fonction publique, des salariés de particuliers employeurs et des affiliés au régime agricole.

²¹ Dans le cadre de ces estimations, l'emploi est défini au sens du Bureau international du travail (BIT) qui nécessite une « traduction » de la déclaration DSN pour établir précisément le fait d'occuper un emploi à une date T ou non. Au sens du BIT, toute personne dont le contrat de travail tel que déclaré par son employeur est ouvert le dernier jour du mois relève de l'emploi salarié. Cet effectif statistique diffère par ailleurs des effectifs réglementaires que l'Urssaf Caisse nationale calcule pour l'application des seuils de cotisations (effectif moyen mensuel, effectif moyen annuel).

²² Obligation d'emploi des travailleurs handicapés. Le recouvrement de la contribution en cas de non-respect de cette obligation, auparavant assurée par l'Agefiph, a été transféré aux Urssaf en 2022.

²³ Enfin, et de façon plus collatérale s'agissant de l'usage de la DSN, est apparu un nouvel acteur dans le champ de la formation professionnelle, la Caisse des dépôts et consignations, qui met en œuvre le nouveau système d'information Agora sur la formation professionnelle et le Compte personnel de formation. La DSN est utilisée à la fois pour alimenter les comptes personnels et pour assurer un suivi de la situation en emploi des personnes ayant bénéficié d'une formation. L'émergence de ce nouveau SI pose des questions de

1.2.1.1 Ces évolutions ont permis de réduire les délais de publication des livrables statistiques fondés sur la DSN

[23] **Différents livrables statistiques fondés sur la DSN sont publiés à échéances régulières et avec des délais variables par rapport à la période de déclaration considérée.** Le tableau 1 ci-dessous regroupe ces livrables par principal producteur. Les livrables mensuels ou trimestriels correspondent à des analyses conjoncturelles. Ils comportent des informations moins détaillées et des concepts moins élaborés que les livrables annuels qui renvoient eux à des analyses plus structurelles sur l'emploi et les salaires. Comme le montre ce tableau, les principaux livrables conjoncturels fondés sur la DSN sont aujourd'hui publiés entre 40 et 70 jours après la fin du trimestre considéré. Néanmoins, ils font l'objet d'estimations provisoires, dites « flash », dans un délai de 40 à 45 jours suivant la fin du trimestre. La base Tous salariés (BTS), qui constitue le principal livrable structurel issu de la DSN, avec des informations à la fois sur les emplois et les salaires déclinées à différentes mailles (poste, salarié, établissement, entreprise), était en 2022 publiée par l'Insee avec un décalage d'environ dix-huit mois par rapport à l'année considérée²⁴.

[24] **Si ces délais s'expliquent essentiellement par les retraitements effectués sur les données pour en améliorer la qualité, le SSP s'efforce de les réduire.** En effet, les utilisateurs de ces données tendent d'autant plus à questionner leur délai de mise à disposition par le SSP que leur source est dorénavant rafraîchie chaque mois. Par ailleurs, le service statistique européen, Eurostat, impose aux instituts statistiques nationaux un calendrier de plus en plus court de fourniture des données trimestrielles. Les estimations « flash » pour les statistiques trimestrielles visent à répondre à cette attente de données plus fraîches, mais en contrepartie d'un statut provisoire, qui implique un caractère révisable. De façon plus ponctuelle, la crise sanitaire, parce qu'elle a engendré des évolutions brusques et majeures sur le marché du travail, a aussi été l'occasion pour le SSP d'expérimenter des usages à plus « haute fréquence » de la DSN. En effet, les évolutions mensuelles étaient alors d'une ampleur nettement plus forte que le risque de « bruit » statistique. S'agissant des données structurelles produites à un rythme annuel, grâce en particulier à l'automatisation croissante et au fil de l'eau des retraitements effectués, les délais de publication devraient se réduire notablement à horizon 2024 en passant à environ un an.

coordination des productions avec celles de la Dares, de labellisation éventuelle par l'ASP des séries produites, et le cas échéant d'expertises conjointes.

²⁴ Ce délai s'est élevé jusqu'à 36 mois au moment de la mise en place de la DSN.

Tableau 1 : Délais actuels et cibles de publication des principaux livrables statistiques fondés sur la DSN

Producteurs	Livrables	Fréquence	Délai actuel de publication	Délai cible de publication
Insee	Estimations d'emploi salarié	Trimestrielle	T + 70 jours après estimation flash à T + 45 jours ²⁵	T + 60 jours Estimation flash à T + 30 jours
	Indice du coût du travail ²⁶	Trimestrielle	T + 75 jours	T+ 65 jours Estimation flash à T+45 jours-
	Base Tous salariés ²⁷	Annuelle	N + 24 mois en 2022	N + 12 mois
	Fichier localisé sur les rémunérations et l'emploi ²⁸	Annuelle	N + 24 mois en 2022	N + 13 mois
Dares	Suivi des contrats de travail dans le champ privé ²⁹	Trimestrielle	T + 90 jours	-
	Suivi de l'emploi intérimaire	Mensuelle	T + 40 jours	-
Urssaf Caisse nationale	Effectifs salariés du secteur privé ³⁰	Trimestrielle	T + 70 jours	T + 60 jours
	Masse salariale ³¹	Trimestrielle	T + 70 jours après estimation flash à T + 60 jours	T + 60 jours

Source : Mission.

²⁵ Les estimations trimestrielles localisées sont livrées à T + 90 jours avec un objectif à T + 80 jours.

1.2.2 La démarche qualité mise en œuvre par le service statistique public éclaire les fragilités des données brutes de la DSN

[25] **Les données brutes issues de la DSN présentent parfois des limites qui peuvent compromettre leur qualité métrologique dans le cadre de certains usages, notamment pour le suivi et l'évaluation des politiques publiques.** Ces anomalies peuvent être liées à des erreurs, des omissions ou des retards de déclaration par les employeurs. Parfois, il est nécessaire de transformer les données brutes afin de décrire les unités statistiques d'intérêt, comme le poste de travail. Enfin, la norme DSN ne permet pas directement, sans traduction, de poser les concepts pertinents pour l'analyse du marché du travail, comme le taux d'emploi au sens du Bureau international du travail, la mesure du retour à l'emploi ou à un emploi durable, les nomenclatures officielles des professions occupées (PCS) ou encore le salaire net perçu par unité de temps de travail rémunéré ou en équivalent temps plein. La stabilité des concepts, et donc de leur mesure, doit par ailleurs être recherchée, alors que la déclaration évolue dans le temps au gré de la réglementation sociale.

[26] **A contrario, les limites des données brutes ne sont pas forcément rédhibitoires.** Dans certains cas, il est d'ailleurs nécessaire de s'appuyer sur des déclarations brutes, les seules opposables aux déclarants. Quand les évolutions sont fortes (montée en charge d'un dispositif ou pendant la crise sanitaire), les délais optimaux de l'action publique sont tels qu'une information raisonnablement bruitée est aussi préférable à une information retraitée mais plus tardive.

[27] **Le service statistique public assure une mission globale de qualification des données issues de la DSN qui contribue à en améliorer la qualité.** Cette mission repose en particulier sur les traitements suivants :

- **une appréciation, voire une correction, des principales erreurs déclaratives** grâce à des contrôles de cohérence externe : à titre d'exemple, les licenciements économiques déclarés en DSN apparaissent sous-estimés par rapport aux inscriptions à Pôle Emploi à la suite d'un licenciement ;
- **une gestion des risques de doublons** liés à des changements non déclarés sur un contrat de travail qui peuvent conduire à compter à tort une embauche et une fin de contrat: selon la Dares, environ 200 000 contrats disparaissent en moyenne chaque mois de la DSN, c'est-à-dire qu'ils ont été déclarés en DSN puis n'apparaissent plus, mais sans qu'aucune fin de contrat ait été signalée, ce qui nécessite de chaîner les contrats entre plusieurs DSN et

²⁶ Il s'agit de la somme de la masse salariale, des cotisations sociales et des autres coûts (nets des exonérations), rapportée au volume horaire de travail. L'ICT « salaires seuls » est un indice de coût du travail hors cotisations sociales.

²⁷ Il s'agit d'une synthèse annuelle sur les postes salariés du privé et de la fonction publique, désormais issus de la DSN, et des déclarations des particuliers employeurs.

²⁸ Il s'agit d'une synthèse produite à partir de la Base tous salariés qui a été enrichie d'informations relatives aux établissements.

²⁹ Hors agriculture, intérim et particuliers employeurs.

³⁰ Hors agriculture et par secteur d'activité et régions.

³¹ Assiette déplafonnée soumise à cotisations et contributions dans le secteur privé hors agriculture.

d'imputer éventuellement un motif de fin de contrat en tenant compte en particulier de la durée du contrat ;

- **une gestion des déclarations retardataires ou rectificatives** permettant de passer de la date de la déclaration à la date de référence : à M+1, environ 5 % des salariés n'ont pas encore été déclarés et environ 30 % des mouvements de main d'œuvre sont mal déclarés et rectifiés ultérieurement, si bien que le chaînage des déclarations mensuelles permet d'améliorer l'exhaustivité et la précision des informations se rapportant à une période déterminée ;
- **une traduction des rubriques administratives en concepts statistiques** permettant de mesurer des agrégats économiques d'intérêt, comme l'emploi, les professions occupées par les salariés ou encore le salaire perçu : s'agissant de ce dernier point, la DSN représentant la paie du mois, des redressements sont nécessaires pour prendre en compte les décalages de paie ou encore les rappels de salaire³².

1.2.3 La production de bases DSN mensuelles ouvre de nombreuses perspectives d'appariements pour analyser les dynamiques d'insertion professionnelle

[28] **La création de la DSN a ouvert des possibilités d'appariement avec d'autres bases de données.** Ces appariements sont devenus un mode de collecte d'information statistique complémentaire particulièrement économe puisqu'ils utilisent une information déjà produite. Dans le cadre du SSP ils sont soumis, comme les enquêtes statistiques, à l'approbation préalable du Conseil national de l'information statistique (Cnis). La création en 2016 d'un « code statistique non signifiant »³³ transverse au SSP, qui se substitue au Nir (le numéro de sécurité sociale, dont l'usage par les statisticiens est encadré et nécessite un décret en Conseil d'État), a eu pour objet de faciliter les appariements au sein du SSP. Ces appariements sont souvent réalisés via le Centre d'accès sécurisé aux données (CASD) qui héberge les bases concernées. L'exhaustivité des DSN sur leur champ, la qualité des identifiants individuels et la possibilité d'avoir une information à différents horizons temporels ont démultiplié l'intérêt de ces enrichissements, pour des usages à la fois de statistiques publiques et de pilotage. La Dares joue un rôle pivot dans la réalisation de ces appariements en lien avec ses partenaires, en particulier Pôle emploi et la Cnaf.

[29] **Ces appariements entre la DSN et d'autres bases permettent de mieux suivre les trajectoires individuelles sur le marché du travail.** Ce suivi représente un enjeu en particulier à l'issue d'une formation, ou du bénéfice d'une mesure de politique de l'emploi (par exemple, un contrat aidé), ou encore après un épisode de chômage ou un problème de santé. Les appariements se substituent ainsi progressivement aux enquêtes d'insertion professionnelle qui étaient menées préalablement sur un échantillon d'individus. Les dispositifs d'appariement

³² [Les résultats des contrôles effectués par les métiers statistiques, comme par les autres métiers en charge de la gestion des droits ou du recouvrement, peuvent être remontés auprès du Gip-MDS dans le cadre d'une boucle de rétroaction pour fiabiliser autant que possible la déclaration en amont (évolution de la norme et des contrôles embarqués, ciblage de la communication auprès des éditeurs et déclarants sur les erreurs les plus fréquentes).

³³ La création de ce code est prévue par l'article 34 de la loi n° 2016-1321 « Pour une République numérique » du 7 octobre 2016.

suivants (cf. annexe 2 pour plus de précisions), qui mobilisent tous les données de la DSN, fournissent des éclairages différents pour répondre à cet enjeu :

- **ForCe** reconstitue les trajectoires des personnes ayant eu un contact avec le service public de l'emploi ou ayant suivi une formation financée totalement ou partiellement par les pouvoirs publics ;
- **InserJeunes** rassemble des données sur la situation sur le marché du travail de jeunes sortis du système éducatif en filières professionnelles, sachant qu'un projet comparable, **InserSup**, est en cours de développement afin de suivre selon le même principe l'insertion professionnelle des sortants de l'enseignement supérieur ;
- enfin, **Midas**, reconstitue les trajectoires professionnelles des inscrits à Pôle emploi et des bénéficiaires de minima sociaux.

1.2.4 Il est envisagé d'utiliser les traitements effectués par les statisticiens publics à des fins de pilotage

[30] **Les conditions dans lesquelles les données administratives retraitées par le SSP, comme celles issues de la DSN, pourraient être utilisées à des fins de pilotage font depuis 2022 l'objet de travaux sous l'égide de l'Insee.** Deux types de retraitements pourraient être distingués, ceux qui visent uniquement à corriger les données sources, dont les résultats peuvent être ouverts à des fins de pilotage lorsque cela s'avère pertinent, et ceux de nature « statistique publique », qui resteraient sous le couvert du secret statistique ou qui comportent des imputations les rendant inadaptées à des usages à des fins de pilotage. Comme le souligne l'Autorité de la statistique publique³⁴, une telle offre de service, mettant à profit l'expertise et l'expérience acquises par les statisticiens publics en matière de structuration et de traitement des données, participerait de l'extension de la fonction d'appui à la conduite des politiques ministérielles prévue par la charte des services statistiques ministériels de 2019.

1.3 L'utilisation de la DSN à des fins de pilotage dans le champ travail-emploi-formation monte en puissance

1.3.1 La DSN permet d'apprécier l'impact des interventions publiques et de les ajuster plus rapidement

[31] **Les données issues de la DSN constituent un actif précieux pour comprendre les mutations en cours sur le marché du travail.** Elles peuvent plus particulièrement contribuer à davantage objectiver :

- les tensions de recrutement par branche, secteur, métier ou territoire ;
- les spécificités de la situation sur le marché du travail de certaines catégories d'actifs (femmes, seniors, travailleurs en situation de handicap, par exemple) ;

³⁴ Délibéré du 31 janvier 2022.

- la qualité des emplois (nature et durée des contrats de travail, niveau relatif de rémunération, *turn-over*, absentéisme, sinistralité...) par territoire, branche, secteur ou métier ;
- les mobilités professionnelles entre secteurs ou métiers, ainsi que les mobilités géographiques des salariés.

[32] **La stratégie « data »³⁵ du ministère en charge du Travail identifie la DSN comme une source de données à mieux mobiliser dans le cadre du pilotage des politiques publiques.** Cette feuille de route décline une stratégie interministérielle³⁶ dont la relance en 2021 fait suite au rapport dit Bothorel³⁷. Ce rapport soulignait l'utilité, démontrée en particulier pendant la crise sanitaire, de mobiliser plus massivement et plus rapidement les données dans le cadre du pilotage des politiques publiques. S'agissant de la DSN, elle a fait l'objet de la mise en place d'une « *taskforce* » sous l'égide de la Dnum des ministères sociaux visant à développer ses usages.

1.3.2 L'utilisation des données de la DSN à des fins de pilotage se diffuse progressivement

1.3.2.1 Une typologie des cas d'usage des données de la DSN à des fins de pilotage peut être établie selon leur objectif principal

[33] **Il n'existe pas de recensement des différents usages des données de la DSN à des fins de pilotage.** En effet, aucun acteur ne centralise les demandes d'accès aux données à cette fin, ni les traitements réalisés. La mission s'est donc efforcée, au fil de ses entretiens, d'identifier ces demandes, sans être pour autant en mesure de garantir leur exhaustivité. La mission a recensé les cas d'usage effectifs, c'est-à-dire des acteurs qui ont obtenu l'accès aux données de la DSN à des fins de pilotage. Elle a aussi identifié des cas d'usage potentiels signalés par des interlocuteurs ayant fait une démarche d'accès à ces données qui n'a pas encore abouti. Elle a distingué les usages mobilisant les données individuelles, le cas échéant pseudonymisées, et ceux reposant sur des données agrégées (au niveau de l'entreprise ou de l'établissement). L'annexe n°3 propose une cartographie détaillée et illustrée de ces cas d'usage précisant à chaque fois le « guichet » mobilisé par les acteurs.

[34] **Ces cas d'usage, dont un échantillon est présenté ici, peuvent être regroupés selon la typologie suivante construite à partir de l'objectif principal recherché :**

- **analyser les parcours en emploi des bénéficiaires de dispositifs de politiques publiques :** la DGEFP a mobilisé le Gip-MDS pour suivre le parcours en emploi de cohortes de bénéficiaires du Contrat d'engagement jeunes (CEJ), le Gip « Plateforme de l'inclusion » s'est adressé à la Dnum pour suivre le parcours en emploi d'allocataires du RSA dans trois départements et,

³⁵ Feuille de route 2021-2023 des données, des algorithmes et des codes sources du MTPEI datée du 22 septembre 2021.

³⁶ Circulaire du Premier ministre datée du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources.

³⁷ Eric Bothorel, Stéphanie Combes, Renaud Vedel, « Pour une politique publique de la donnée », décembre 2020.

enfin, la Caisse des dépôts et consignations et la DGEFP ont déployé la plateforme « Agora » qui fournit, à partir de la DSN, des données sur le retour à l'emploi des bénéficiaires d'un dispositif de formation ;

- **cibler le contrôle et l'accompagnement des entreprises dans le respect des dispositions prévues par le code du Travail** : la DGT a mis en place en 2021 un « SI précarité » pour aider l'inspection du travail à repérer les entreprises en risque d'abuser des contrats courts ; elle utilise les effectifs des entreprises en DSN pour identifier celles dont le taux d'accidents du travail (« Dat'IA ») est hors norme ou qui ne remplissent pas leur obligation de calculer l'index d'égalité professionnelle (« EgaPro ») alors qu'elles y sont assujetties ; la DGEFP a développé l'appliquetif « OETH » pour outiller les Dreetes afin de mieux repérer et accompagner les entreprises qui ne respecteraient pas leurs obligations en matière d'emploi handicapé ;
- **éclairer le marché caché de l'emploi en identifiant les entreprises susceptibles de recruter** : la start-up d'État « La bonne alternance » souhaite ainsi accéder aux données de la DSN pour identifier de façon prédictive les entreprises les plus susceptibles d'offrir des contrats en alternance ;
- **anticiper l'impact et suivre la montée en charge d'un nouveau dispositif de politique publique** : la base de données dite Dad-E du Gip-MDS a ainsi été mobilisée par la DSS pour analyser le déploiement des primes « Pepa³⁸ » en 2019 et « inflation » en 2021 ;
- **identifier le plus en amont possible dans un secteur donné les tensions en matière de RH et les facteurs d'attractivité des métiers** : c'est l'objet du « cube DSN »³⁹ en cours de développement en lien avec la Dares dans le SI Diamant destiné aux Agences régionales de santé (ARS), qui permettra d'identifier des marqueurs de tension RH dans le secteur sanitaire, médico-social ou social (augmentation du volume d'heures supplémentaires, de l'absentéisme, du turn-over, du recours à l'intérim...) ; cette ambition est partagée par le projet « Ramsès », développé par un prestataire pour la DGCS, afin d'analyser l'impact des conventions soumises à son agrément sur la masse salariale du secteur médico-social et social à but non lucratif ;
- **analyser l'évolution des emplois et des compétences au niveau d'un secteur, d'une branche ou d'un territoire afin de détecter des risques ou des opportunités** : c'est à cette fin que la start-up d'État « Signaux faibles », qui vise à repérer les entreprises en risque de défaillance, souhaite mobiliser les données de la DSN, tandis que la tête de réseau des Carif-Oref utilise déjà ces données dans le cadre du projet « Octopilot » pour apprécier les passerelles possibles entre métiers par région.

³⁸ Prime exceptionnelle de pouvoir d'achat.

³⁹ Table à multiples dimensions.

Utilisation de la DSN pour cibler les entreprises avec un recours abusif aux contrats courts

La DGT a mis en service en septembre 2021 un outil dit « SI précarité » à destination de l'inspection du travail pour cibler dans leurs contrôles les entreprises présentant une part élevée de contrats précaires. Un contrat précaire est défini dans l'outil comme un CDD ou une mission d'intérim. Les données issues de la DSN correspondent à des données agrégées à l'échelle d'une entreprise à partir des données compilées par la Dares dans Sismmo. Elles sont transmises à M+3 par rapport à la déclaration en DSN. L'outil donne la possibilité de réaliser différents filtres (par région, département, niveau d'effectif souhaité...) et de calculer plusieurs indicateurs pour apprécier le taux de recours aux contrats précaires en fonction de la taille de l'entreprise, dans le temps et par rapport au secteur d'activité ou à la région d'implantation.

La Dnum, en lien avec la DGT et la Dares, élabore un outil pour l'inspection du travail mobilisant cette fois les données individuelles et nominatives de la DSN. Ce projet est soutenu par Étalab qui finance partiellement trois Entrepreneurs d'intérêt général (EIG) mis à disposition de la Dnum pendant douze mois depuis septembre 2022. Selon la Dnum, l'objectif premier de ce projet, inscrit dans la feuille de route « données » du ministère en charge du Travail, est de prouver la valeur des données issues de la DSN en les rendant intelligibles par les utilisateurs, en l'espèce les services de l'inspection du travail. Le cas d'usage identifié concerne, comme le « SI précarité », la lutte contre le recours abusif aux contrats courts. Cependant, contrairement au SI précarité, ce traitement devrait mobiliser un algorithme susceptible d'apprécier la probabilité d'un recours abusif aux contrats courts. Grâce aux données individuelles et nominatives (qui devraient être, selon la Dnum, anonymisées ou pseudonymisées), il permettrait de descendre au niveau du poste de travail pour, par exemple, apprécier le respect du délai de carence entre deux CDD.

Utilisation de la DSN pour identifier des alertes RH dans le secteur sanitaire

La Dnum, et en son sein le service à compétence nationale en charge des SI des Agences régionales de santé (ARS), développe, en lien avec la Dares et la Direction générale de l'offre de soins (DGOS), un cube DSN dans le SI « Diamant ». « Diamant » est un outil décisionnel de pilotage par la donnée centrée sur la régulation de l'offre de soins. Il est utilisé par les ARS, la DGOS, la Haute autorité de santé (HAS) et les 135 établissements supports de Groupements hospitaliers de territoire ». L'objectif de ce cube, qui devrait être mis en service en juin 2023, est d'utiliser les données mensuelles à M+1 agrégées de la DSN, retraitées par la Dares, pour mesurer, et si possible anticiper, des tensions sur le plan des ressources humaines au niveau d'un établissement du champ sanitaire ou médico-social ainsi que dans les maisons et centres de santé. Ces tensions pourront être appréciées sur la base de données construites à partir de la DSN telles que l'évolution du volume d'heures supplémentaires, de l'absentéisme ou encore du taux de turn-over. Les facteurs d'attractivité des établissements (par exemple, l'impact de la mise en œuvre d'une nouvelle prime ou encore les différences de niveau et structure de rémunération) pourront aussi être mieux appréciés sachant que jusqu'ici, les données utilisées par les ARS étaient celles du bilan social annuel des établissements, moins fines et surtout moins fraîches. Une formation est prévue pour accompagner le déploiement, avec notamment une alerte sur la prudence à avoir vis-à-vis des données ayant un recul de moins de trois mois.

1.3.2.2 Ces cas d'usages souvent sous-traités traduisent la maturité encore assez modeste des acteurs vis-à-vis de la science des données

[35] Le recours aux données de la DSN à des fins de pilotage par les acteurs rencontrés repose souvent sur des compétences externalisées, celles d'un prestataire ou celles de la Dnum, ce qui ne favorise pas la montée en compétences des administrations.

[36] **Les cas d'usage recensés montrent que l'usage des données issues de la DSN est encore aujourd'hui assez « classique » :**

- les données mobilisées visent majoritairement à produire des indicateurs agrégés au niveau des établissements ;
- les données de la DSN les plus utilisées sont les effectifs des entreprises qui servent au calcul de ratios ;
- au niveau individuel, les cas d'usage les plus fréquents consistent en l'appariement d'une cohorte de bénéficiaires d'un dispositif de politique publique avec les données de la DSN pour un suivi du parcours en emploi salarié ;
- en dehors de « La bonne alternance », qui veut alimenter son algorithme avec les données de la DSN, les autres cas d'usages ne mobilisent pas la DSN à des fins prédictives.

[37] **Cette approche de la donnée traduit un niveau de maturité encore assez modeste de la part de la majorité des acteurs du champ travail-emploi-formation.** Dans le cadre de la feuille de route « données », la Dnum a inscrit des actions pour améliorer la culture de la donnée au sein du ministère en charge du Travail. Elle a impulsé le projet « Champollion », lancé en 2022, qui mobilise en particulier la DSN dans le cadre d'un algorithme visant à identifier les entreprises à risque d'abus des contrats précaires, avec un soutien du programme « entrepreneurs d'intérêt général » d'Étalab⁴⁰, afin notamment de démontrer la valeur de cette approche. Pôle emploi, qui développe depuis 2018 des projets à vocation prédictive fondés sur la valorisation des données fait plutôt figure d'exception⁴¹, tout en reconnaissant le caractère encore expérimental de ces démarches dont la valeur doit être évaluée.

⁴⁰ Étalab est un département de la Dnum assurant, au titre des missions de l'Administrateur général des données, des algorithmes et des codes sources, la coordination, la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée. Il est aussi responsable du programme « Entrepreneurs d'intérêt général » qui recrute chaque année pendant dix mois des talents du numérique pour résoudre des défis d'intérêt général aux côtés des administrations.

⁴¹ Financé par le Fonds de transformation de l'action publique (FTAP), le programme « Intelligence emploi » inclut différents projets fondés sur des algorithmes comme, pour les demandeurs d'emploi, un service d'analyse automatique du CV et, pour les entreprises, un service de prédiction sur le risque qu'une offre d'emploi ne soit pas pourvue dans un délai de 30 jours après son dépôt.

2 Les données de la DSN restent sous-utilisées en raison notamment d'un cadre et d'une organisation inadaptés

2.1 Le cadre juridique ne favorise pas l'accès aux données de la DSN et fait débat à l'aune de la politique publique de la donnée

2.1.1 Les textes réglementaires n'ouvrent pas suffisamment la voie à des usages des données de la DSN à des fins de pilotage

2.1.1.1 L'architecture et le contenu des textes réglementaires relatifs à la DSN freinent l'accès à ses données à des fins de pilotage

[38] **Le cadre juridique d'accès aux données issues de la DSN est calqué sur ses objectifs premiers (cf. annexe n° 4).** La DSN constitue en effet un traitement de données à caractère personnel⁴² à des fins de gestion se substituant à plusieurs déclarations ou formalités antérieures. La loi⁴³ a donc prévu que ces données servent essentiellement dans le cadre des actes de gestion afférents (recouvrement des cotisations, des contributions sociales et de certaines impositions, vérification de leur montant, ouverture et calcul des droits des salariés en matière d'assurance sociale...). Néanmoins, la loi ouvre la voie à un usage plus général des données pour « *l'accomplissement par les administrations et organismes destinataires de leurs missions* ».

[39] **L'accès aux données de la DSN repose en outre sur une architecture juridique relativement complexe et peu adaptée à leur usage à des fins de pilotage.** Le décret du 28 mars 2013⁴⁴ définit les finalités de ce traitement de données, toujours en lien essentiellement avec l'objectif de substitution à des démarches antérieures. Si deux des finalités distinguées peuvent renvoyer à du « pilotage », le terme est absent du décret, alors qu'il est évoqué dans celui relatif à l'utilisation du Nir⁴⁵. Le décret de 2013 établit aussi la liste des treize organismes attributaires de données issues de la DSN⁴⁶, qui sont ceux utilisant ces données en lieu et place des informations collectées précédemment via des formalités spécifiques. Ce décret renvoie ensuite à un arrêté, dit arrêté « filtre », la définition, par destinataire et par catégorie de données, des informations transmises

⁴² La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés définit un traitement de données à caractère personnel comme « toute opération ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé », sachant que constitue une donnée à caractère personnel, toujours au sens de cette loi, « toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres ».

⁴³ Loi n°2012-387 du 22 mars 2012 relative à la simplification du droit et à l'allégement des démarches administratives.

⁴⁴ Décret n°2013 266 du 28 mars 2013 relatif à la DSN.

⁴⁵ Décret n° 2019-341 du 19 avril 2019 relatif à la mise en œuvre de traitements comportant l'usage du numéro d'inscription au répertoire national d'identification des personnes physiques ou nécessitant la consultation de ce répertoire.

⁴⁶ Sur ces 13 destinataires, sept relèvent du champ travail-emploi-formation professionnelle : Délégation générale à l'emploi et à la formation professionnelle (DGEFP), Direction générale du Travail (DGT), Pôle emploi, Agefiph, Caisse des dépôts et consignations, Agence de services et de paiement (ASP), Dares.

ainsi que les finalités selon lesquelles elles peuvent être utilisées. Cet arrêté qui décrit de façon très granulaire les rubriques retransmises doit être révisé dès que la norme et les besoins des destinataires évoluent. Il a ainsi été révisé à sept reprises depuis la création de la DSN. En outre, les traitements autorisés sont particulièrement restrictifs. À titre d'exemple, la DGEFP n'est autorisée à utiliser les données de la DSN que pour « *suivre l'évolution des quotas de contrat d'apprentissage dans les entreprises de 250 salariés et plus* » et la Caisse des dépôts et consignations pour « *alimenter le compte personnel de formation des salariés concernés* », alors que, de facto, elle mobilise aussi la DSN pour suivre le retour à l'emploi des personnes à l'issue d'une formation financée.

2.1.1.2 L'accès aux données de la DSN est trop lent car il est imbriqué avec le processus d'évolution de la norme et repose sur des moyens limités

[40] **Le processus de révision de l'arrêté « filtre » est alourdi par son imbrication avec les évolutions annuelles de la norme DSN associées à une nouvelle version du cahier technique.** La DSS, en tant que tutelle du Gip-MDS, instruit et arbitre, en lien avec le Gip-MDS, les demandes d'évolution de la norme⁴⁷. Ces évolutions sont retracées chaque année dans un cahier technique publié un an avant son entrée en vigueur. Pour élaborer chaque nouvelle version de ce cahier, la DSS interroge, avant le mois de mars N-2 par rapport à l'entrée en vigueur de ladite version, l'ensemble des membres du Gip-MDS sur leurs besoins⁴⁸. Ce recueil porte à la fois sur les besoins d'évolution de la norme DSN et sur ceux d'accès à des données individuelles déjà présentes en DSN (l'accès à des données agrégées ne requiert pas d'évolution du cadre juridique). En effet, la prise en compte de ce double besoin nécessite l'évolution d'un même texte, l'arrêté « filtre », qui est publié après recueil de l'avis de la Cnil.

[41] **Le calendrier de révision de l'arrêté « filtre » est en outre tributaire de la faible taille de l'équipe de la DSS qui en assure la responsabilité, ce qui ralentit un peu plus l'accès aux données individuelles.** Une même équipe constituée de deux ETP au sein du bureau en charge du recouvrement à la DSS instruit à la fois le besoin d'évolution de la norme et les demandes d'accès aux données⁴⁹. La DSS a indiqué à la mission avoir des difficultés à tenir l'objectif d'une révision annuelle de l'arrêté « filtre ». Révisé chaque année entre 2013 et 2016, il n'a ensuite été révisé qu'à deux reprises, en 2019 puis en 2022, avec un impact à la hausse sur le délai entre la demande initiale et l'accès aux données⁵⁰. À titre d'exemple, un délai de deux ans s'est écoulé entre

⁴⁷ La gouvernance de la norme DSN pourrait évoluer dans les prochains mois en lien avec la mise en place d'une maîtrise d'ouvrage stratégique interministérielle, positionnée auprès de la DSS, et la réactivation du Comité de normalisation des données sociales (CNDS) telles que recommandées par le rapport de l'Igas sur la gouvernance du Gip-MDS (juillet 2022).

⁴⁸ La DGEFP coordonne la réponse à cette sollicitation de la DSS pour les organismes et administrations relevant du MTPEI.

⁴⁹ Ce dossier distingue la transmission de nouvelles données déclarées en DSN ; la transmission de données DSN à des organismes ou administrations qui ne les recevaient pas auparavant ; l'extension des données transmises à des organismes et administrations qui recevaient déjà des données de la DSN ; les précisions apportées aux données de la DSN déjà transmises aux organismes et administrations.

⁵⁰ La DSS a indiqué à la mission qu'une nouvelle version de l'arrêté « filtre » devrait paraître au 2^{ème} trimestre 2023.

l'expression du besoin d'accès à des données individuelles de la DSN par la DGT et sa prise en compte dans l'arrêté « filtre » du 7 mars 2019.

2.1.2 La politique publique de la donnée relancée en 2021 questionne tant la légalité que la pertinence de ce cadre juridique

2.1.2.1 La valorisation des données administratives, « carburant » de l'intelligence artificielle, est conditionnée à leur meilleure circulation

[42] Depuis une quinzaine d'années, la circulation des données administratives est promue à travers en particulier une obligation d'échange entre administrations de « *toutes les informations ou données strictement nécessaires pour traiter une demande présentée par le public ou une déclaration transmise par celui-ci en application d'un texte législatif ou réglementaire* »⁵¹.

[43] En effet, un meilleur partage des données administratives est considéré comme un préalable à leur valorisation aux fins de modernisation de l'action publique. Dans son étude sur l'intelligence artificielle et l'action publique⁵², le Conseil d'État souligne que l'administration n'est pas propriétaire des données qu'elle détient : « *elle n'est est que la gardienne, et son devoir est de les partager pour qu'en soit tirée toujours plus de valeur, dans l'intérêt général* »⁵³. Dans la continuité du rapport Bothorel, la circulaire du Premier ministre du 27 avril 2021 demande aux administrations de « *rechercher en permanence la meilleure circulation de la donnée* » et de mettre en place dans chaque ministère un administrateur des données.

2.1.2.2 Dans ce contexte, le cadre juridique d'accès aux données de la DSN tend à être de plus en plus contourné

[44] L'encadrement juridique de l'accès aux données de la DSN fait l'objet d'interprétations divergentes. Le directeur interministériel du numérique (Dinum), en sa qualité d'administrateur général des données⁵⁴ (AGD)⁵⁵ a estimé que les dispositions issues de la loi pour une République numérique, ainsi que le principe de responsabilisation des administrations créant des traitements de données, rendaient illégal l'arrêté « filtre »⁵⁶. De son côté, la Direction des affaires juridiques

⁵¹ Loi n°2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit.

⁵² « Intelligence artificielle et action publique : construire la confiance, servir la performance », étude du Conseil d'État adoptée en assemblée générale plénière du 31 mars 2022.

⁵³ Le Conseil d'État précise que « *ce partage ne doit avoir d'autre borne que la nécessaire confidentialité de certaines données et, en surplomb, la nécessité de ne pas fragiliser la confiance dans les institutions et collectivités publiques auxquelles les données sont remises par des tiers* ».

⁵⁴ La circulaire n°6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources a confié au directeur interministériel du numérique, en sa qualité d'administrateur général des données, algorithmes et codes sources, la coordination du réseau des administrateurs ministériels des données.

⁵⁵ Ce courrier du 16 novembre 2021 fait suite à une saisine de l'administratrice des données du ministère en charge du Travail datée du 11 octobre 2021 qui interrogerait l'AGD sur les contraintes d'ordre réglementaire à l'extension de l'exploitation de la DSN prévue par la feuille de route ministérielle relative aux données.

⁵⁶ Dans ce courrier, l'AGD écrit en particulier qu'« *il apparaît que le législateur a clairement entendu organiser un régime de circulation large des données de la DSN puisque pouvaient y avoir accès non seulement les*

(DAJ) des ministères sociaux, ainsi que la DSS, rappellent que la DSN implique la mise en œuvre d'un traitement de données à caractère personnel qui doit être conforme au RGPD et à la loi « informatique et liberté » du 6 janvier 1978, ce que ne remet pas en cause la loi pour une République numérique. Ce traitement exige par conséquent un encadrement au niveau réglementaire afin de préciser les catégories de données traitées et la liste de leurs destinataires⁵⁷. La DAJ précise néanmoins que le décret en Conseil d'État serait suffisant, sans renvoi à un arrêté « filtre ».

[45] **L'encadrement juridique de l'accès aux données de la DSN fait déjà l'objet de contournements autorisés, en particulier sur le fondement du code du Travail.** Ainsi, les membres du service public de l'emploi peuvent accéder à une partie des données à caractère personnel issues de la DSN via le « système d'information concernant les demandeurs d'emploi et salariés ». Prévu par le code du Travail⁵⁸, ce système d'information, placé sous la responsabilité de Pôle emploi, est alimenté en particulier par les données de la DSN auquel l'opérateur a accès conformément à l'arrêté « filtre ». C'est sur la base de ces dispositions que l'Unédic, qui demande par ailleurs un accès direct aux données individuelles de la DSN, accède aujourd'hui à ces données. Le code du Travail prévoit aussi un accès indirect aux données de la DSN dans le cadre du dispositif « Agora », géré par la Caisse des dépôts et consignations, qui assure un partage d'informations sur les parcours des bénéficiaires d'une formation professionnelle entre les organismes financeurs⁵⁹.

[46] **Avec l'arrêté « Champollion », la Dnum a manifesté sa volonté d'aller plus loin dans le contournement du cadre juridique d'accès aux données de la DSN.** Cet arrêté du 21 septembre 2022 a créé un traitement de données à caractère personnel ayant pour finalité de « *structurer et de mettre à disposition des administrations relevant du champ du travail, de l'emploi et de l'insertion, des données pour l'accomplissement de leurs missions* ». Les données à caractère personnel issues de la DSN font partie de celles pouvant être enregistrées dans ce traitement. Les finalités de nature générique définies par le texte excèdent largement celles prévues par le décret du 28 mars 2013 sur la DSN⁶⁰ et contreviennent au principe de minimisation prévu par le Règlement général sur la protection des données (RGPD). Élaboré avec un faible degré

organismes de protection sociale mais également l'ensemble des services de l'État » et que « eu égard au cadre juridique (...), aux orientations définies par le Premier ministre, et par la simple application de la hiérarchie des normes, il m'apparaît clairement que l'arrêté « filtre » ne peut avoir pour effet de limiter le cadre de la circulation des données prévues par la loi ».

⁵⁷ Cet encadrement est nécessaire pour respecter les principes de limitation des finalités et de minimisation des données prévus à l'article 5 du RGPD qui dispose que « *les données à caractère personnel (...) doivent être collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités, (...) [et] adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées* ».

⁵⁸ Article R. 5312-38 du code du travail

⁵⁹ Ces données, issues de la DSN, correspondent à celles dont la Caisse des dépôts et consignations est destinataire, conformément à l'arrêté « filtre » au titre de sa mission d'alimentation du Compte personnel de formation (CPF). Bien que la plateforme Agora, présentée par la Caisse des dépôts et consignations comme le « data hub de la formation », soit à ce stade logée au sein du SI CPF, l'utilisation des données de la DSN à une fin non prévue par les textes relatifs à la DSN justifierait, selon la DGEFP, une mise en cohérence de ces derniers avec les dispositions du code du Travail.

⁶⁰ Simplifier les démarches administratives ; favoriser le recours à l'intelligence artificielle et son développement ; participer à des expérimentations relatives à l'utilisation des données.

d'association de la DSS, et sans consultation préalable de la Dares, ce texte soulève selon la DAJ des points d'attention⁶¹.

2.2 L'orientation et l'accompagnement des utilisateurs des données issues de la DSN sont insuffisamment développés

2.2.1 Il n'existe pas de fonction d'aiguillage des utilisateurs vers la base de données la plus pertinente par rapport à leurs besoins

2.2.1.1 Le Gip-MDS a développé une offre de services « data » qui tend à être mobilisée par la DSS sans prise en considération d'autres sources existantes

[47] **Le Gip-MDS a développé depuis 2019 une offre de services « data » à partir de l'exploitation des données brutes de la DSN.** Le développement des usages autour de la DSN fait partie des cinq axes stratégiques de sa feuille de route 2022-2025. Cette offre repose sur quatre à cinq ETP. Elle s'appuie sur une plateforme dite PIQ (Projet d'indicateurs qualité) dont la finalité première est de permettre une analyse de la qualité déclarative de la DSN, et qui peut aussi fournir des indicateurs dits décisionnels⁶². Le Gip-MDS souligne qu'il ne s'agit pas d'une offre de services statistique dans la mesure où il n'effectue quasiment aucun retraitement sur les données brutes de la DSN et parce que les indicateurs n'ont pas vocation à être publiés⁶³. Cette offre de services consiste à répondre aux sollicitations des acteurs qui l'identifient comme un guichet d'accès aux données de la DSN et à leur fournir à un appui dans leur expression de besoin. Le Gip-MDS comptabilise une quinzaine de sollicitations dans les deux dernières années.

[48] **Par ailleurs, de façon plus proactive, le Gip-MDS a aussi identifié des indicateurs qui pourraient être produits de façon automatisée avec les données brutes de la DSN,** comme une partie des indicateurs contribuant au calcul de l'index égalité femmes-hommes⁶⁴. Le Gip-MDS propose en particulier de calculer à partir de la DSN des indicateurs RH (ratio d'absentéisme, part des employés seniors, répartition des ETP par nature de contrat...). La DSS ne s'est pas prononcée

⁶¹ En effet, si elle reçoit les flux DSN au titre de sa fonction de direction des SI, la Dnum n'est pas, en droit, destinataire des données individuelles issues de la DSN. Par ailleurs, la DAJ estime que les finalités du traitement et les catégories de données restent définies de façon trop large, en particulier au regard du RGPD. Enfin, elle rappelle que les traitements ultérieurs mis en œuvre par les destinataires des données avec les données de « Champollion » devront faire l'objet, le cas échéant, du texte réglementaire et de la formalité adéquats.

⁶² Des outils décisionnels y sont également adossés afin de faciliter l'exploitation, le requêtage et la visualisation des données.

⁶³ La proposition initiale du Gip-MDS inscrite dans un document de travail daté du 26 mars 2019 avait été assimilée à une « offre de service statistique » (en lien sans doute avec son intitulé « mise à disposition de données type indicateurs et statistiques « décisionnelles ») et suscité une opposition de la part de la Dares, l'Insee et l'Acoss. Dans un courrier du 11 avril 2019 adressé à la DSS, ces trois organismes avaient alerté sur le fait que ce projet risquait de générer de la confusion dans le débat public et n'allait pas dans le sens d'une rationalisation de l'offre statistique.

⁶⁴ Cette offre de services est détaillée dans une note « Proposition d'un cadre mutualisé en matière d'indicateurs décisionnels » datée du 27 octobre 2022 et adressée par le Gip-MDS à la DSS.

sur la pertinence de cette offre de services « data » mais a indiqué à la mission qu'elle ne devait pas mobiliser plus que les moyens actuels.

[49] **La DSS régule l'accès aux données de la DSN de la part des acteurs en sous-traitant une partie de l'instruction au Gip-MDS.** En pratique, le Gip-MDS est destinataire d'une partie des demandes d'accès à la DSN de la part d'acteurs, extérieurs au MTPEI (Direction générale des entreprises, start-up d'État...), qui l'identifient comme un « guichet ». Il pré-instruit ces demandes pour le compte la DSS en appréciant leur faisabilité et leur compatibilité avec le cadre juridique puis il exécute les travaux après accord de la DSS.

[50] **La DSS ne dispose pas des moyens ni de l'expertise pour apprécier correctement ces demandes d'accès aux données, ni pour les aiguiller vers les bases de données pertinentes.** En 2022, elle aurait reçu une quinzaine de demandes d'accès aux données de la DSN. Faute d'informations sur les dates de réception de ces demandes, la mission n'a pas pu apprécier le délai de réponse moyen de la DSS à ces demandes qui n'impliquent pas de modifier les textes. Néanmoins, à titre d'exemple, un délai de plus de six mois a été constaté entre la demande de la DGEFP, pourtant soutenue par le cabinet de la ministre en charge du Travail, d'accéder à des données agrégées pour le suivi des bénéficiaires du Contrat d'engagement jeune (CEJ), et l'accord donné par la DSS. Ce délai peut s'expliquer par la difficulté pour la DSS à apprécier la compatibilité de certaines demandes avec les finalités prévues par le décret du 28 mars 2013, malgré le travail de pré-instruction parfois réalisé par le Gip-MDS. Quand une demande est jugée recevable, la DSS ne dispose pas non plus de la connaissance d'autres voies ou sources alternatives en aval susceptibles de satisfaire le besoin. Son exécution est donc sous-traitée mécaniquement au Gip-MDS.

[51] **Une partie des demandes d'accès aux données de la DSN échappe en pratique à la DSS.** Ce phénomène s'explique sans doute par le délai de réponse de la DSS aux demandes d'accès. En l'absence de retour, une partie des demandeurs se tournent alors vers d'autres acteurs disposant également des données issues de la DSN. À titre d'exemple, le Gip « Plateforme de l'inclusion », qui avait au départ adressé à la DSS sa demande d'accès aux données individuelles de la DSN, s'est finalement tourné vers la Dnum des ministères sociaux qui lui a communiqué les données attendues. La responsable de la start-up d'État « Signaux Faibles », qui avait également soumis à la DSS une demande d'accès aux données issues de la DSN, a finalement obtenu ces données par l'intermédiaire de la Dares, grâce au soutien du cabinet de la ministre en charge du Travail.

2.2.2 Les utilisateurs, y compris potentiels, des données issues de la DSN ont des besoins d'accompagnement

[52] **Il n'existe pas aujourd'hui de dispositif garantissant la prise en compte des besoins non satisfaits des acteurs utilisant ou souhaitant potentiellement utiliser les données de la DSN.** L'exemple des Opco et des observatoires de branches professionnelles est à cet égard éclairant. Une partie de ces acteurs ont identifié la DSN comme un gisement de données particulièrement utile pour construire une stratégie d'achat de formation, éclairer les partenaires sociaux sur la situation comparée de l'emploi et des salaires dans une branche ou sur un territoire et, plus généralement, répondre aux demandes d'informations qui leur sont adressées de façon croissante par différentes institutions (État, Conseils régionaux, agglomérations...) sur des sujets spécifiques (sur les seniors, l'emploi des femmes, les travailleurs handicapés...). Plusieurs Opco ont donc manifesté depuis 2019, auprès du Gip-MDS, de la DGEFP et de France compétences, leur souhait

d'accéder directement à des données issues de la DSN, y compris des données individuelles. À défaut d'accord de la part de la DSS, France compétences a travaillé depuis juillet 2022 avec la Dares pour mieux répondre aux besoins des Opco en adaptant dans un premier temps les portraits de branche publiés chaque année par la Dares⁶⁵, qui prévoit par ailleurs de publier courant 2023 des estimations trimestrielles d'emploi à l'échelle de chaque branche.

[53] Quand ils accèdent aux données de la DSN, certains utilisateurs ne bénéficient pas de l'accompagnement dont ils auraient besoin pour :

- identifier les fragilités des données qu'ils mobilisent à l'aune de l'utilisation prévue ;
- vérifier que le traitement envisagé est conforme aux règles relatives à la protection des données personnelles ;
- repérer les données qui pourraient être utilisées à titre complémentaire par rapport à la DSN pour enrichir leurs analyses ;
- s'assurer que les concepts utilisés dans le cadre de leur exploitation des données de la DSN sont cohérents par rapport à l'existant.

[54] Deux cas d'usage peuvent être mobilisés, à titre d'exemples, pour illustrer cette différence d'approche :

- la DGEFP a demandé en 2022 au Gip-MDS, avec l'autorisation de la DSS, de réaliser un suivi de la situation à 3, 6 et 12 mois de cohortes d'allocataires du CEJ : aucun échange n'a eu lieu au préalable avec la Dares pour apprécier si les données brutes étaient plus pertinentes pour ce suivi que les données retraitées par le service statistique, disponibles avec un décalage de trois mois, ni pour échanger sur la façon de définir une personne en emploi (par exemple, en fonction du nombre d'heures de travail) ;
- à l'inverse, le service à compétence nationale en charge des SI des ARS, en lien avec la DGOS, a bénéficié d'un accompagnement de la Dares pour construire le « cube DSN » du SI Diamant, qui permettra à horizon juin 2023 aux ARS d'identifier les signaux d'une tension RH au niveau des établissements sanitaires, médico-sociaux et sociaux sur la base des données de la DSN préalablement retraitées par la Dares, au sein desquelles ont été sélectionnées celles qui permettaient de calculer les indicateurs les plus pertinents.

[55] Enfin, il n'existe aujourd'hui aucune « communauté » des utilisateurs de la DSN à des fins de statistiques publiques ou de pilotage. Les cas d'usage restent particulièrement peu débattus et mutualisés. Or tous les utilisateurs des données de la DSN rencontrés par la mission ont exprimé ce souhait de pouvoir échanger avec d'autres sur la qualité respective des bases de données, les cas d'usage identifiés ou encore l'intérêt de mobiliser les données de la DSN en lien avec la transformation numérique des services. Cette communauté pourrait s'étendre aux acteurs qui mobilisent déjà, ou cherchent à mobiliser, la DSN dans le cadre de leur gestion prévisionnelle des emplois et des compétences, comme les Opco et les observatoires de branche professionnelle.

⁶⁵ Ces portraits concernent uniquement les branches de plus de 5 000 salariés. Les portraits 2020 ont été publiés en juillet 2022 et les portraits 2021 sont attendus en juillet 2023.

2.3 Le « *Labor data hub* » reste à ce stade un objet administratif non identifié

2.3.1 Prévu par la feuille de route « données » du ministère en charge du Travail, le « *Labor data hub* » visait en particulier à favoriser l'utilisation de la DSN

[56] **Cette feuille de route adoptée le 22 septembre 2021 bénéficiait d'un portage politique fort pour développer l'usage des données, en particulier la DSN.** C'est dans ce cadre qu'est prévue la création d'un « *Labor data hub* », défini comme « *la plateforme de données et de services du travail, de l'emploi et de la formation pour faciliter le partage et promouvoir l'utilisation des données au sein de la sphère sociale (...) il réunit à la fois des éléments techniques (entrepôt), des services et des outils pour former un espace de travail commun autour des données, et plus particulièrement pour creuser les usages de l'intelligence artificielle (...)* », l'objectif fixé étant d'établir le cadre juridique et de choisir la plateforme technique de ce « *hub* » à un an avec, comme cas d'usage prioritaire, l'évaluation de l'assurance chômage à travers l'appariement des données de Pôle emploi, de la DSN et de la Cnaf.

[57] **Le bilan de la feuille de route, réalisé en octobre 2022 par la Dnum, traduit un net essoufflement d'une partie des chantiers relatifs à la valorisation des données.** Des actions prévues par la feuille de route restent inabouties, telles que la création d'un catalogue des données et des algorithmes, la mise en place de modalités d'échanges normées et partagées⁶⁶, et celle du « *Labor data hub* ». Par ailleurs, le comité ministériel des données, piloté par la Dnum, ne s'est pas réuni entre avril 2022 et mars 2023. La Dnum considère aujourd'hui qu'il est nécessaire de mieux imbriquer les chantiers « données » inscrits dans cette feuille de route avec les besoins métiers des directions et opérateurs du ministère. Une réflexion a été engagée en début d'année, en lien avec les directions du ministère en charge du Travail, sur le devenir de cette stratégie.

2.3.2 Le cadrage du « *Labor data hub* » ne fournit pas les éléments d'appréciation nécessaires à un arbitrage sur son devenir

[58] **À partir de l'objectif fixé par la feuille de route « données », la Dnum, appuyée par un prestataire⁶⁷, a réalisé un cadrage du « *Labor data hub* » (LDH) au premier trimestre 2022.** Les documents communiqués par la Dnum à la mission montrent que ce cadrage s'est centré :

- d'une part, sur la recherche de cas d'usages des données, en particulier celles de la DSN, à travers des entretiens réalisés avec les directions métiers du ministère et Pôle emploi ;

⁶⁶ S'agissant de ces deux dernières actions, le document de la Dnum renvoie à la création à venir d'une « *taskforce cartographie* ».

⁶⁷ Le prestataire sélectionné par la Dnum, Octo Technology, a facturé cette prestation d'une durée de trois mois au montant de 99 960,00€ TTC.

- d'autre part, sur l'instruction de différents scénarios d'architecture pour favoriser la mise à disposition de ces données, du plus centralisé (la création d'un « lac de données »⁶⁸ sous la responsabilité de la Dnum) au moins centralisé (un « *data mesh* »⁶⁹).

[59] **Si ce cadrage a permis de mieux identifier les attentes des différentes parties prenantes, il n'est pas suffisant pour fonder un arbitrage sur la création du LDH.** En effet, l'examen du dossier de cadrage communiqué à la mission soulève plusieurs points d'attention :

- ni la DSS, ni le Gip-MDS, ni le CASD n'ont été consultés, ce qui explique le fait que le cadrage fasse en partie l'impasse sur les modalités actuelles d'accès à la DSN, qu'il s'agisse du cadre juridique régulé par la DSS, ou de l'accès à ces données via les bases sous la responsabilité de la DSS ou du SSP ;
- les acteurs rencontrés n'ont pas identifié de cas d'usage évidents, ce qui peut témoigner d'un faible niveau de maturité sur l'usage de la donnée dans les politiques publiques : ainsi, le cas d'usage initialement identifié pour constituer le « *Minimum viable product* » (MVP)⁷⁰ du LDH, qui visait à utiliser plusieurs catégories de données pour identifier les entreprises ayant un recours illégal au travail détaché, s'est avéré finalement ne pas correspondre aux besoins de la DGT ;
- en termes de gouvernance, le cadrage s'est davantage centré sur les questions liées à l'architecture des données, sans faire au préalable une cartographie des différentes bases de données existantes, et au détriment d'une analyse des fonctions attendues du LDH et de leur répartition entre acteurs.

2.3.3 Il n'existe pas de modèle-type de « *hub* » ministériel de données

[60] **La mission n'a pas identifié de modèle de « *hub* » ministériel de données sur la base duquel pourrait être construit le « *Labor data hub* » :**

- aucun modèle « type » de « *hub* » n'a été diffusé auprès des administrateurs de données, par exemple par la Dinum ;
- aucune cartographie, et *a fortiori* aucune analyse comparative, des « *hubs* » existants n'a été identifiée par la mission ;
- aucun exemple à l'étranger de « *hub* » de données administratives sur l'emploi n'a non plus été communiqué par les interlocuteurs interrogés à ce sujet⁷¹.

[61] **Les trois exemples de « *hub* » identifiés par la mission traduisent bien l'hétérogénéité des objets que ce terme tend à masquer :**

⁶⁸ Un lac de données (ou « *data lake* ») est un espace de stockage global des données présentes au sein d'une organisation.

⁶⁹ Il s'agit d'un modèle dans lequel les données conservent chacune leur propre infrastructure avec un partage entre acteurs favorisés par des standards communs et des règles d'interopérabilité.

⁷⁰ Un MVP ou produit minimum viable est la version la plus minimaliste d'un produit conçu avec un minimum de moyens pour tester le produit auprès d'un maximum d'utilisateurs.

⁷¹ La mission a en particulier interrogé le département de l'Insee en charge des relations internationales.

- le « **Health data hub** », créé sous la forme de Gip, vise à favoriser l'accès aux données de santé aux projets d'intérêt public, y compris ceux portés par des acteurs privés : si l'objectif était initialement que les données soient hébergées par une plateforme gérée par ce « *hub* », les données restent à ce stade dans des bases principalement accessibles sous la responsabilité de l'assurance maladie et du CASD⁷² ;
- le « **Bercy hub** », sous la responsabilité de l'administrateur général des données, positionné auprès du secrétariat général des ministères économiques et financiers, joue essentiellement un rôle d'animation en matière de valorisation des données des différentes directions, lesquelles présentent des moyens et un niveau de maturité hétérogènes dans ce domaine : une plateforme dite « Nubonyxia » qui, sans héberger les données, fournirait un espace sécurisé pour les valoriser, est en cours de création ;
- enfin, le **projet « Innovations, données et expérimentations en éducation » (IDEE)**, porté par le service statistique du ministère en charge de l'Éducation nationale, se présente aussi comme un « *hub* », plutôt ciblé sur le monde de la recherche, en ce qu'il vise à favoriser l'accès aux données administratives détenues par le ministère via un centre sécurisé, avec un accent mis sur leur documentation et leur structuration ainsi que sur l'animation de la communauté des utilisateurs.

2.3.4 Le CASD représente aujourd'hui le principal « *hub* » d'accès aux données statistiques issues de la DSN

[62] **S'il s'adresse plus particulièrement aujourd'hui aux chercheurs, le Centre d'accès sécurisé aux données (CASD) est aujourd'hui le principal « *hub* » de données statistiques issues de la DSN.** Constitué sous forme de Gip le CASD héberge en effet une grande partie des bases statistiques issues de la DSN mises à disposition par le SSP, qui font partie des sources les plus utilisées. L'accès à ces données suppose l'accord préalable du comité du secret statistique qui apprécie en particulier l'opportunité, du point de vue de l'intérêt général, du projet d'étude et de la capacité des demandeurs à protéger les données confidentielles. Une fois obtenu l'accord de ce comité, le CASD met à disposition des utilisateurs, via une procédure payante de boîtier informatique et de bulle sécurisée, un environnement pour utiliser les données de la DSN. Le CASD offre aussi la possibilité d'apparier les données issues de la DSN avec d'autres bases de données, en particulier les bases fiscales (cf. annexe n°2). Par ailleurs, le CASD est la solution retenue par la Dares et la Drees pour héberger leur plateforme informatique, projet qui n'a pas pu aboutir au sein de l'infrastructure des ministères sociaux.

[63] **La mobilisation des données hébergées par le CASD, dont celles issues de la DSN, tend à se développer.** Entre 2018 et 2022, la hausse des demandes d'accès aux données du CASD s'est établie en moyenne à 7 % par an. Selon les données communiquées par le CASD, sur environ 500 sources disponibles sur cette plateforme, les données issues de la DSN sont les plus mobilisées, avec plus de deux cents projets de recherche en cours et près de la moitié des

⁷² Le « *Health data hub* » (HDH) joue néanmoins le rôle de guichet unique pour l'accès aux données. Un comité éthique et scientifique (Cesrees) est sollicité pour avis. Selon le HDH, une moyenne de 300 projets sont soumis au Cesrees chaque année. Le projet initial de créer une plateforme permettant d'héberger de façon centralisée l'ensemble des données de santé a été mis en suspens en 2020 en raison de la recherche d'un hébergeur autre que le « *cloud* » de Microsoft qui soulevait des questions en termes de souveraineté.

utilisateurs de ce « *hub* » qui les mobilisent. Cet usage se développe à des fins qui dépassent de plus en plus la recherche et l'évaluation au sens strict. Ainsi, c'est via le CASD que la tête de réseau des Carif-Oref a obtenu l'accès aux données de la DSN utilisées pour développer l'outil « Octopilot » (cf. section 1.3.2.1 *supra*). L'Opco Akto a aussi mobilisé des données de la DSN via le CASD pour réaliser une étude sur les travailleurs saisonniers qui sera publiée en 2023. D'autres Opco ont indiqué à la mission envisager de réaliser une démarche similaire pour obtenir des données mobilisables plus facilement en fonction de leurs besoins que les données par branche mises à disposition en « *open data* » par l'Insee ou la Dares.

3 Pour favoriser la bonne utilisation des données de la DSN, une gouvernance spécifique doit être mise en place

3.1 La bonne utilisation des données de la DSN peut être appréciée à l'aune de différents principes

3.1.1 Le cadre juridique de la DSN doit faciliter l'utilisation de ses données à des fins de pilotage

[64] **Le cadre juridique doit faciliter l'accès aux données issues de la DSN à des fins de pilotage tout en veillant au respect du RGPD, en particulier le principe de minimisation⁷³.** La feuille de route de la Dinum considère à cet égard que la politique publique de la donnée est encore trop limitée par les incertitudes quant à l'interprétation des différents cadres juridiques qui induisent, chez les détenteurs, « *une résistance à la mise à disposition des données* ». Sous réserve d'une analyse complémentaire par la DAJ des ministères sociaux et d'échanges avec la Cnil, la mission recommande de réviser le décret n° 2013-266 du 28 mars 2013 relatif à la DSN pour :

- inclure explicitement parmi les finalités du traitement DSN l'utilisation de ses données à des fins de pilotage des politiques publiques ;
- harmoniser les droits d'accès aux données au sein du ministère en charge du Travail pour simplifier les projets de valorisation des données que les directions pourraient mener en partenariat ;
- identifier mieux les acteurs du service public de l'emploi, aujourd'hui, et de France Travail, demain, parmi les destinataires des données de la DSN ;
- définir les catégories de données à transmettre aux organismes attributaires sur la base d'une approche plus générique.

[65] Concernant l'arrêté « filtre », son abrogation, jugée possible sur le plan juridique par la DAJ des ministères sociaux, doit être conditionnée à la possibilité de redistribuer, à partir du seul décret, les flux de la DSN aux organismes attributaires au sens du décret. Selon la DAJ, un acte

⁷³ Selon la Cnil, le principe de minimisation prévoit que les données à caractère personnel doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées.

réglementaire resterait nécessaire, au sein d'une administration de l'État, pour la création d'un traitement de données utilisant les données issues de la DSN, ce texte pouvant *a priori* être un arrêté dès lors que le traitement relève de la compétence d'un chef de service⁷⁴ ; la liste exacte des données de la DSN utilisées dans le cadre de chaque traitement de données créé par une administration devrait ensuite figurer dans l'analyse d'impact relative à la protection des données (AIPD) réalisée, le cas échéant, par le responsable de traitement.

L'analyse d'impact relative à la protection des données

Prévue par le RGPD, l'AIPD est un outil de responsabilisation des organismes qui créent un traitement de données. Elle est obligatoire pour les traitements susceptibles d'engendrer des risques élevés. Elle se décompose en trois parties : une description détaillée du traitement mis en œuvre ; l'évaluation de la nécessité et de la proportionnalité du traitement (finalité, données et durée de conservation, information et droits des personnes...) ; l'étude des risques sur la sécurité des données et leurs impacts potentiels sur la vie privée. Selon la Cnil, l'AIPD est un outil qui permet de construire un traitement conforme au RGPD et respectueux de la vie privée. Elle concerne les traitements de données personnelles qui sont susceptibles d'engendrer un risque élevé pour les droits et libertés des personnes concernées.

[66] Enfin, l'arrêté « Champollion », qui soulève des questions sur le plan juridique et excède le cadre de la DSN, devrait être abrogé.

Recommandation n°1 Réviser le décret n° 2013-266 du 28 mars 2013 relatif à la DSN sur la base d'une approche plus générique des usages en introduisant la finalité de pilotage des politiques publiques et abroger l'arrêté « filtre » du 9 février 2022

3.1.2 La DSN contient des données sensibles qui exigent un traitement conforme au droit de la protection des données personnelles

[67] **Les standards en termes de respect du RGPD des différents utilisateurs de la DSN sont très hétérogènes.** Si le principe du RGPD est celui d'une responsabilisation de chaque créateur de traitement de données, force est de constater qu'il n'y a pas aujourd'hui de vision globale sur la conformité des traitements issus de la DSN au RGPD. La déléguée à la protection des données (DPO⁷⁵) des ministères sociaux a indiqué à la mission ne pas avoir accès à l'ensemble des analyses d'impact relatives à la protection des données (AIPD) produites, le cas échéant, par les utilisateurs des données issues de la DSN, même au sein de ces ministères. En pratique, la mission a constaté que la production d'AIPD en cas d'utilisation des données issues de la DSN était systématique au sein du SSP, et plus aléatoire chez les autres acteurs. Elle a noté enfin l'absence de cadre d'échanges à ce sujet qui favoriserait un alignement des standards entre acteurs et améliorerait la conformité des usages de ces données, dont une partie sont sensibles, au RGPD.

⁷⁴ La DAJ des ministères sociaux a indiqué à la mission qu'un groupe de travail, piloté par le Secrétariat général du Gouvernement (SGG), étudiait à l'heure actuelle la possibilité d'assouplir les formalités administratives nécessaires pour créer un traitement de données. À ce stade, un décret simple, soumis à l'avis de la Cnil, reste nécessaire pour créer un traitement de données relevant de plusieurs chefs de service ne relevant pas du même ministre.

⁷⁵ « Data protection officer ».

Recommandation n°2 Veiller, en lien avec le délégué à la protection des données (DPO), à la bonne alimentation du registre des traitements effectués à partir de la DSN, centraliser les Analyses d'impact relatives à la protection des données (AIPD) concernant leur usage et mieux accompagner les utilisateurs dans la prise en compte effective des règles issues du Règlement général sur la protection des données (RGPD)

3.1.3 Le choix d'une base de données construite à partir de la DSN doit reposer sur une documentation permettant d'arbitrer selon les caractéristiques recherchées

[68] **L'intérêt d'une base de données ne peut pas se définir indépendamment du cas d'usage envisagé.** Il existe une base DSN déclarative mensuelle et des bases de données « aval » issues de la DSN et retraitées qui présentent différentes caractéristiques et niveaux de qualité. Il manque aujourd'hui une cartographie de ces bases de données, et *a fortiori* une analyse de leurs avantages et inconvénients respectifs à l'aune des objectifs poursuivis.

[69] **Le choix d'une base de données adéquate doit reposer sur un arbitrage entre différents critères.** Cet arbitrage dépend en particulier de la pondération des caractéristiques respectives de chaque base dans les besoins de l'utilisateur, en particulier celle des critères respectivement relatifs à la fraîcheur et la robustesse des données. Les utilisateurs de ces données, en particulier à des fins de pilotage, doivent pouvoir accéder à la documentation nécessaire pour réaliser cet arbitrage. Ils doivent aussi, parfois, être accompagnés par des acteurs disposant d'une expertise sur les caractéristiques des différentes bases de données ou sur la norme DSN. Ainsi, le recours à la base PIQ du Gip-MDS devrait être privilégié quand des données précoces (brutes), sont indispensables pour suivre la montée en charge d'un nouveau dispositif (ex : prime inflation, prime de partage de la valeur). En revanche, dès lors qu'il s'agit de travailler sur des concepts du marché du travail (emploi, salaires, catégories de salariés...), ou d'apprécier le parcours en emploi de bénéficiaires d'un dispositif de politiques publiques (ex : bénéficiaires du CEJ, allocataires du RSA), la mobilisation de bases davantage retraitées, comme Sismmo, est préférable.

Recommandation n°3 Mettre à disposition des utilisateurs un catalogue des bases de données construites à partir de la DSN précisant les traitements effectués ainsi que leurs avantages ou inconvénients respectifs dans le cadre de différents usages

3.1.4 Les limites des analyses réalisées à partir de données brutes issues de la DSN doivent être prises en compte dans leur communication éventuelle

[70] **L'utilisation de données brutes issues de la DSN est susceptible de générer du « bruit » ou de l'« entropie » si elles sont utilisées à mauvais escient, sans expertise et prise en compte des risques encourus.** Pour préserver la confiance du public dans les messages qui lui sont communiqués, il convient que les entités utilisatrices soient clairement informées des limites associées à l'utilisation de données brutes, qu'elles sont ensuite tout à fait légitimes à mobiliser, si ces données sont estimées pertinentes, pour leurs besoins de pilotage, comme la montée en charge de nouveaux dispositifs de politique publique. Toutefois, la publication de ces données, si elles s'avèrent différentes des données de référence diffusées en général ultérieurement par le SSP, pourrait créer de la confusion dans le débat public. Ce risque est soulevé par les acteurs de

la sphère statistique publique, qui doivent régulièrement justifier de leurs chiffrages consolidés établis avec un certain recul quand ils diffèrent de ceux issus des données brutes.

[71] **Pour prévenir ce risque, une charte de « bonne utilisation » de la DSN, incluant des conseils de communication, pourrait être élaborée.** L'élaboration de cette charte devrait associer les acteurs du SSP et le Gip-MDS. Elle devrait s'appliquer à l'ensemble des autres acteurs qui utilisent les données de la DSN à des fins de pilotage.

Recommandation n°4 Élaborer une charte commune aux différents utilisateurs des données issues de la DSN à des fins de pilotage au sein de l'administration

3.1.5 L'utilisation des données issues de la DSN doit s'accompagner d'une réflexion éthique

[72] **La question de l'éthique dans l'utilisation des données issues de la DSN est un sujet à explorer, surtout dans la perspective du développement du recours à l'intelligence artificielle.** Si la mission n'a pas instruit à proprement parler cette question qui n'était pas dans son périmètre, elle estime indispensable que ce sujet fasse l'objet d'échanges entre les acteurs concernés. Par exemple, en cas d'utilisation de ces données dans le cadre d'un algorithme, comment encadrer leur finalité, éviter les biais discriminatoires, assurer un niveau de transparence, et donc de confiance, suffisant ? Il s'agit, d'une part, d'éviter une confiance excessive de la part de l'utilisateur en lui fournissant des outils pour comprendre les résultats et contrer l'effet « boîte noire » et, d'autre part de savoir calculer les biais liés à la qualité des données sur lesquelles l'algorithme s'appuie et de faire valider l'éthique des modèles développés⁷⁶.

Recommandation n°5 Créer un cadre d'échanges sur les questions éthiques soulevées par l'utilisation des données de la DSN, en particulier en cas de recours à l'intelligence artificielle

3.1.6 En amont de l'accès aux données issues de la DSN, il convient de faire émerger ses utilisations les plus pertinentes

[73] **Il est indispensable de partir des besoins « métiers » pour tirer le meilleur potentiel des données issues de la DSN.** À cet égard, la feuille de route « données » du MTPEI, qui arrive à échéance en 2023, pourrait mieux identifier les cas d'usage en lien avec les directions concernées. Ces directions pourraient être davantage accompagnées à cette fin par la Dnum, le cas échéant en lien avec la Dinum. La feuille de route de la Dinum⁷⁷, publiée en mars 2023, souligne la nécessité d'accélérer le déploiement de nouveaux cas d'usage des données administratives « *en donnant accès aux ressources et aux données nécessaires* ». Un incubateur de projets *data*, le « *datalab* », sera créé au sein de la Dinum, en particulier pour accompagner les porteurs de projet de l'administration « *vers un impact métier réel en exploitant à plein les données existantes* ». Les projets de valorisation des données pourront par ailleurs bénéficier d'un financement au travers d'un guichet dédié du Fonds de transformation de l'action publique (FTAP) doté de 10 M€.

⁷⁶ Voir, à titre d'exemple, la [charte de Pôle emploi pour une intelligence artificielle éthique](#).

⁷⁷ [Feuille-de-route-DINUM.pdf \(numerique.gouv.fr\)](#).

Recommandation n°6 Organiser des « *hackathons* » ou autres temps de co-construction sur la donnée, le cas échéant en mobilisant le « *datalab* » de la Dinum, pour faire émerger des cas d’usage de la DSN à des fins de pilotage dans le champ travail-emploi-formation

3.2 Pour améliorer l’utilisation des données de la DSN, six fonctions doivent être créées ou renforcées

3.2.1 Instruire aux niveaux technique et juridique les besoins d’accès secondaires aux données issues de la DSN

[74] Quel que soit le cadre juridique, même en élargissant les finalités relatives aux destinataires, il restera nécessaire de réguler et d’instruire les demandes d’accès secondaire à ces données de la part d’acteurs ne faisant pas partie des attributaires primaires au sens du décret (exemple : Opco, administrations en dehors du champ social, start-up d’État...). Ces expressions de besoin doivent être plus ou moins accompagnées dans leur élaboration. Elles doivent ensuite être instruites sur le plan juridique (conformité de la demande d’accès par rapport aux finalités du traitement DSN adéquat) et technique (faisabilité au regard des données disponibles).

3.2.2 Orienter et accompagner les utilisateurs dans le recours à la base de données la plus pertinente par rapport à leurs besoins

[75] Les utilisateurs des données issues de la DSN, en particulier ceux qui y recourent de façon ponctuelle, ont besoin d’être aiguillés vers la base de données la plus pertinente par rapport à leurs besoins. Au-delà de l’identification de la « bonne » base de données, certains acteurs peuvent avoir besoin d’une intermédiation pour que le responsable de la base de données leur ouvre l’accès aux données attendues, dans le respect des dispositions et procédures existantes pour garantir la confidentialité des données.

[76] Au-delà du choix de la base de données, certains utilisateurs ont besoin d’être accompagnés dans la compréhension de ces données en lien avec la norme DSN, le processus déclaratif, les retraitements effectués ou non qui ont un impact sur leur qualité.

3.2.3 Contribuer à une utilisation des données de la DSN respectueuse des données personnelles et éthique

[77] Là encore, selon la familiarité des utilisateurs avec le cadre RGPD et l’utilisation de la donnée, un accompagnement peut être nécessaire pour qu’ils intègrent certains standards et adaptent leurs modalités d’utilisation des données.

3.2.4 Animer la communauté des utilisateurs de données issues de la DSN pour favoriser les retours d’expérience et susciter de nouveaux usages

[78] Les cas d’usage de la DSN à des fins de pilotage gagnent à être mutualisés et à faire l’objet d’échanges pour aller vers une utilisation plus avisée et efficace de ces données et susciter de nouvelles utilisations en lien avec la transformation numérique de l’action publique. Des actions

de sensibilisation à l'attention des managers permettraient également de susciter un réflexe d'utilisation des données administratives, comme la DSN, à des fins de pilotage.

3.2.5 Veiller à l'efficacité et à la cohérence globales des traitements issus de la DSN à des fins statistiques et de pilotage

[79] Le fonctionnement actuel conduit à une dispersion des moyens et à la redondance de certains traitements, qui gagneraient à être mutualisés, à la fois pour gagner en efficacité (par exemple sur les appariements) et en cohérence, pour homogénéiser les méthodes lorsque cela est possible, sans réinventer à chaque fois des solutions différentes, qui risquent de surcroît de susciter de la confusion dans le débat public en cas de communication (cf. section 3.1.4 *supra*).

3.2.6 Recueillir les besoins d'évolution du décret relatif à la DSN liés au pilotage et s'assurer de leur prise en compte par la DSS

[80] Même si l'arrêté « filtre » est supprimé et si le décret DSN est révisé sur la base d'une approche plus générique, il restera toujours nécessaire, même de façon moins fréquente, d'adapter le cadre juridique en lien avec l'utilisation de la norme DSN et les besoins.

3.3 Trois schémas de gouvernance sont proposés pour orchestrer ces fonctions

3.3.1 Il est indispensable d'identifier une direction cheffe de file pour orchestrer ces fonctions et incarner cette gouvernance

[81] **Les fonctions décrites *supra*, qui conditionnent une meilleure utilisation des données de la DSN, en particulier à des fins de pilotage, ne sont à ce stade pas identifiées par les utilisateurs.** Certaines, comme l'adaptation du cadre juridique ou l'instruction des besoins des utilisateurs, sont en pratique exercées principalement par la DSS. D'autres, comme l'accompagnement des utilisateurs ou leur orientation vers les bases de données pertinentes, sont exercées de façon plus aléatoire par différents acteurs (Dares et Dnum en particulier). Enfin, l'animation de la communauté des utilisateurs de la DSN à des fins de pilotage est inexistante.

[82] **Si ces fonctions nécessiteront toujours l'expertise de plusieurs acteurs, elles doivent trouver leur place au sein d'une gouvernance sous l'égide d'une direction cheffe de file.** En effet, au regard du niveau de maturité encore assez faible de la plupart des utilisateurs des données de la DSN, la mise en place d'une gouvernance spécifique, même si elle doit rester légère, conditionne le développement des usages selon les principes proposés par la mission. La mise en place de cette gouvernance spécifique a pour objectifs de :

- mieux identifier les besoins des utilisateurs des données de la DSN à des fins de pilotage ;
- les orienter et les accompagner vers les solutions les plus appropriées ;
- faire remonter d'éventuels besoins complémentaires aux producteurs de bases de données ;

- donner de la visibilité aux usages de la DSN à des fins de pilotage pour jouer un rôle de catalyseur.

[83] **Différents schémas de gouvernance sont envisageables, qui diffèrent par leur « centre de gravité », ou encore par le positionnement de la direction à qui serait confié le rôle de chef de file.** Chacun présente des forces et des faiblesses selon l'analyse proposée ci-dessous qui s'appuie sur une appréciation de l'expertise dont disposent les principaux acteurs pour orchestrer les fonctions nécessaires à la bonne utilisation des données de la DSN. Le tableau 2 ci-dessous synthétise cette analyse.

3.3.2 Une gouvernance avec un centre de gravité « protection sociale »

[84] Si on ne peut pas évoquer, à proprement parler, de gouvernance des usages de la DSN, force est de constater que la DSS joue aujourd'hui, en lien avec le Gip-MDS, un rôle de régulation de l'accès aux données brutes de la DSN⁷⁸. Une nouvelle entité, intitulée « *mission interministérielle en charge des projets de collecte et d'exploitation des données nécessaires à la gestion des prestations sociales* », placée sous l'autorité du DSS, est en cours de préfiguration. L'avis de vacance du poste de responsable de cette mission, paru le 12 février dernier, précise qu'« *il (elle) est destinataire de l'ensemble des demandes d'accès à des données nouvelles ou existantes collectées ou détenues dans ces dispositifs examinés avec les équipes de la DSS, du GIP-MDS ou des organismes et administrations concernés* ». La fonction de chef de file pourrait être confiée à cette mission interministérielle en charge des projets de collecte et d'exploitation des données nécessaires à la gestion des prestations sociales.

[85] Ce scénario a pour avantage de relier la question de l'accès aux données de la DSN à des fins de pilotage à celle de l'évolution de la norme DSN ainsi qu'au processus de fiabilisation des données, dont la DSS assure la maîtrise d'ouvrage stratégique. Cependant, sa principale faiblesse tient au fait que cette mission n'est pas compétente sur les autres bases de données aval construites à partir de la DSN produites notamment par les acteurs de la statistique publique, et n'a pas l'expertise correspondante. La mission serait moins compétente pour appréhender correctement les besoins des utilisateurs de la sphère travail-emploi-formation et à les aiguiller, voire les accompagner, dans l'utilisation des bases de données les plus appropriées, hors de la sphère déclaration de la DSN. Par ailleurs, selon la mission, les moyens de cette mission devraient être fortement mobilisés par le chantier de la solidarité à la source, ce qui pourrait justifier que la question de l'accès aux données à des fins de pilotage dans la sphère emploi soit traitée par une autre instance.

3.3.3 Une gouvernance avec un centre de gravité « data »

[86] La fonction d'administration des données (Amdac) pour le ministère en charge du Travail a été confiée à la Dnum, tandis que, dans le champ santé-solidarités, elle a été attribuée à la Drees. La Dnum joue un rôle pivot dans la mise en œuvre de la feuille de route « données » du ministère en charge du Travail. Elle a réalisé le cadrage du « *Labor data hub* » et impulsé le projet « *Champion* ». En tant que DSI du ministère, elle héberge une base de données issue de la DSN

⁷⁸ Ce constat renvoie au diagnostic posé par la mission Igas sur la gouvernance du Gip-MDS et des dispositifs de déclaration sociale (juillet 2022).

même si, en droit, elle n'est pas destinataire de ces données pour son propre usage. Enfin, la Dnum dispose de compétences en matière d'accompagnement à la transformation numérique. Ce sont autant de raisons qui pourraient justifier que la Dnum assure un rôle de cheffe de file dans la gouvernance des usages de la DSN à des fins de pilotage.

[87] Néanmoins, le degré de collaboration de la Dnum avec la DSS et le Gip-MDS en matière de valorisation des données, et notamment de celles de la DSN, est aujourd'hui faible, comme l'illustre le précédent de l'arrêté « Champollion ». La Dnum n'est pas familière non plus des problématiques de « qualité » des données, à apprécier en fonction des objectifs poursuivis, et, plus globalement, elle est moins à même d'éclairer les utilisateurs de ces données sur leurs fragilités en lien avec le processus déclaratif. Enfin, si la Dnum a une vision transversale des priorités métiers de l'ensemble des ministères sociaux, elle n'est pas systématiquement représentée dans les instances de pilotage des dispositifs de politique publique du champ travail-emploi-formation.

3.3.4 Une gouvernance avec un centre de gravité « statistiques emploi-travail »

[88] La Dares contribue déjà, comme tous les SSM, à un appui aux administrations dans l'accomplissement de leurs missions, pouvant aller jusqu'à un appui à l'aide à la décision et au pilotage des politiques publiques, comme le prévoit la charte des SSM publiée en 2019⁷⁹. Elle est à ce titre représentée dans la majorité des instances de pilotage des réformes du champ travail-emploi-formation (CEJ, assurance chômage, France Travail, accompagnement renforcé des bénéficiaires du RSA...). Ces activités d'aide à la décision ou au pilotage ne donnent d'ailleurs pas lieu à publication, contrairement aux livrables statistiques. La Dares participe par ailleurs au réseau des SSM sous l'égide de l'Insee, ce qui lui confère une ouverture large sur les problématiques sectorielles des autres ministères sur la thématique emploi-formation.

[89] Elle dispose ainsi de compétences la dispensant de recourir à des prestations externes, même en matière de valorisation des données qui dépassent ses missions statistiques. En effet, comme l'a rappelé le rapport Dinum-Insee sur l'évaluation des besoins de l'État en compétences et expertises en matière de données (juin 2021)⁸⁰, une grande partie des compétences disponibles au sein de l'État en matière de données se situe en dehors de la filière numérique. Ce vivier de compétences justifiait, selon les auteurs, de développer une offre de services « donnée » des SSM adaptées à chaque ministère et articulée avec celle des services chargés du numérique.

[90] En pratique, la Dares exploite déjà aujourd'hui en dehors du cadre statistique les données de la DSN de Sismmo, disponibles mensuellement avec un délai situé entre un et trois mois selon le niveau de retraitement souhaité. Un exemple significatif est celui de la collaboration entre la Dares et le service à compétence nationale en charge du SI des ARS en vue du développement d'un « cube DSN » dans le SI Diamant, qui permettra à terme de mieux identifier les alertes RH dans le secteur sanitaire, médico-social et social. Si les données utilisées dans ce cube seront proches des données brutes, afin que les ARS identifient les alertes en temps quasi-réel, elles ont fait l'objet d'échanges ayant permis de bien cerner leurs fragilités (en particulier, le faible niveau

⁷⁹ [Charte des services statistiques ministériels \(developpement-durable.gouv.fr\)](https://developpement-durable.gouv.fr/).

⁸⁰ [RAPPORT-besoins-competences-donnee.pdf \(numerique.gouv.fr\)](#).

de renseignement du Finess) et de construire les indicateurs de tension RH les plus pertinents par rapport aux besoins exprimés par les ARS et la DGOS.

[91] Toutefois, cette fonction d'appui aux utilisateurs des données de la DSN à des fins de pilotage n'est pas identifiée dans l'organisation et l'offre de services de la Dares. Elle est aussi fonction des autres priorités de la Dares, par ailleurs engagée dans des opérations de production de statistiques publiques et de leur diffusion. La Dares n'est pas non plus toujours clairement identifiée comme un « guichet » pouvant aider les utilisateurs des données de la DSN, y compris extérieurs au ministère (ex : Opco, branches) à identifier les données pertinentes et à les mobiliser en fonction de leurs besoins. Elle ne dispose pas de moyens spécifiques à cette fin et les initiatives prises pour répondre aux besoins d'utilisateurs comme les Opco, ou les ARS, reposent sur sa capacité à redéployer des moyens, en l'espèce ceux de son « département DSN », qui assure aussi la production de livrables statistiques (en particulier les estimations trimestrielles d'emploi et les statistiques sur les mouvements de main-d'œuvre). Enfin, la Dares dépend en grande partie des moyens de la Dnum pour réaliser des activités de « *datavisualization* »⁸¹.

Tableau 2 : Appréciation des points forts/faibles des acteurs en lien avec l'utilisation des données DSN à des fins de pilotage dans le champ travail-emploi-formation (- à ++)

Expertises	DSS ⁸²	Dnum	Dares
Cadre juridique DSN	++	-	-
Norme DSN	++	-	+
Bases de données DSN et qualité des données	+	-	++
Problématiques métiers travail-emploi-formation	-	+	++
Méthodologie de suivi, pilotage et d'évaluation	+	+	++
Valorisation des données à des fins de transformation numérique	-	++	+

Source : *Mission*.

⁸¹ La « *datavisualization* » ou la *dataviz* est une discipline de la « *data science* » qui emploie des outils graphiques afin de traduire des quantités importantes de données en visuels compréhensibles. Elle rend intelligibles les données collectées afin de transmettre des informations via des représentations accessibles à tous.

⁸² Sous-direction en charge du recouvrement aujourd'hui et, à court terme, mission interministérielle en charge des projets de collecte et d'exploitation des données nécessaires à la gestion des prestations sociales.

3.4 Un « Labor data hub », redéfini sur la base de ces fonctions, pourrait être positionné auprès de la Dares et offrir un service d'appariement

[20] Le « *Labor data hub* » pourrait être relancé sous une forme différente en l'orientant vers la prise en charge de ces fonctions. Il serait positionné auprès de la Dares, compte tenu de son expertise statistique et métier dans le domaine travail-emploi-formation professionnelle, tout en associant l'ensemble des acteurs identifiés *supra*. Son articulation avec la mission interministérielle en cours de préfiguration auprès de la DSS devra notamment être précisée.

3.4.1 Une équipe projet constituée de deux à trois ETP positionnés auprès de la Dares pourrait constituer le « *Labor data hub* »

[92] Pour incarner la gouvernance des usages des données DSN à des fins de pilotage et orchestrer les fonctions nécessaires à cette fin, la mission recommande de créer, auprès du directeur de la Dares, une cellule composée de deux à trois ETP. À titre de comparaison, l'offre de services « *data* » du Gip-MDS, qui inclut un travail sur la conformité de la DSN à la norme, repose aujourd'hui sur environ quatre ETP. Il est difficile pour la mission d'estimer de façon robuste les moyens dont aurait besoin ce « *hub* » en l'absence d'éléments prévisionnels de volumétrie sur les sollicitations qui lui seront adressées. L'instruction des besoins d'accès peut être plus ou moins chronophage selon la capacité des demandeurs à formuler leurs attentes. En revanche, une charge fixe peut être identifiée en lien avec le travail de catalogage des bases de données et d'élaboration d'un service d'appariements. Ces ETP dévolus à la Dares devraient tous conduire à une augmentation de son plafond d'emplois. En termes de profils, il est indispensable que cette équipe maîtrise les enjeux liés à la qualité des données en lien avec les bases existantes, la stratégie de transformation numérique du ministère mais aussi qu'elle soit en mesure de prendre en compte les besoins « métiers » des utilisateurs du champ travail-emploi-formation.

[93] Le département de la Dares en charge de la production statistique à partir de la DSN ne serait pas partie intégrante de cette cellule. Il continuerait à produire la base de données issue de la DSN (Sismmo) et ses livrables.

[94] Pour donner de la visibilité à cette cellule, y compris auprès d'autres administrations qui identifient l'objet « *hub* » comme un guichet favorisant l'accès à des données, elle pourrait s'intituler « *Labor data hub* » en cohérence avec le projet prévu par la feuille de route « données » du ministère en charge du Travail. Elle aurait pour rôle de coordonner les différents acteurs dont l'expertise, et les moyens, sont nécessaires au développement des fonctions sur lesquelles serait assise cette gouvernance :

- la DSS continuerait d'assurer la responsabilité du cadre juridique relatif à la DSN, y compris dans ses dispositions relatives à l'accès aux données, et déléguerait au LDH l'examen de la conformité des demandes d'accès aux données à ce cadre ;
- le Gip-MDS, maître d'ouvrage opérationnel du processus déclaratif amont sous la responsabilité stratégique de la mission interministérielle à la DSS, mettrait à disposition de cette cellule, et des utilisateurs qui la sollicitent, son expertise sur la norme DSN et le processus déclaratif, ainsi que la base de données PIQ utile pour suivre la montée en charge de dispositifs pour lesquels des données les plus précoces sont indispensables (ex : prime inflation) ;

- la DAJ, représentée par la déléguée à la protection des données, jouerait un rôle de conseil auprès des utilisateurs sur le respect du RGPD ;
- la Dnum ferait le lien entre les utilisations de ces données à des fins de pilotage et, en tant qu'administratrice ministérielle des données, des algorithmes et des codes sources (Amdac), sa stratégie globale de transformation numérique et de valorisation des données : l'Amdac assurerait en outre une fonction de présidence de la comitologie du LDH⁸³, dont le secrétariat serait assuré par la Dares ;
- le CASD offrirait un espace sécurisé pour la réalisation de croisements (ou appariements) entre les données issues de la DSN et d'autres bases de données statistiques (bases fiscales ou de l'Enseignement supérieur, bases relatives aux allocataires des minima sociaux, aux demandeurs d'emploi...).

[95] La cellule a besoin d'un minimum de deux à trois ETP pour assurer, d'une part, la coordination de ces fonctions exercées directement par les acteurs concernés et, d'autre part, la mise en œuvre des actions qu'elle exercerait en propre, à savoir :

- l'élaboration d'une cartographie et d'une documentation sur les bases de données issues de la DSN ;
- l'instruction des besoins d'accès à la DSN, le conseil auprès des demandeurs pour définir une solution adaptée à leurs besoins et l'aiguillage vers les bases de données pertinentes ;
- la conception et le développement d'une offre de services d'appariement (cf. *infra*) ;
- le conseil auprès des utilisateurs pour veiller à la conformité des usages au RGPD et favoriser une approche éthique des usages ;
- l'animation de la communauté des utilisateurs pour partager les retours d'expérience et identifier les nouveaux besoins afin d'enrichir l'offre existante. Cette communauté sera constituée, dans un premier cercle, des acteurs qui font aujourd'hui partie des destinataires en droit des données individuelles de la DSN (DGEFP, DGT, Pôle emploi, Agence de services et de paiement, Agefiph, Caisse des dépôts et consignations, DGCS, DGOS) et, dans un deuxième cercle, d'acteurs avec une mission de service public qui souhaitent accéder à des fins de pilotage à des données agrégées, voire à des données individuelles pseudonymisées, issues de la DSN (autres membres du service public de l'emploi, administrations externes aux ministères sociaux comme la Direction générale des entreprises, start-up d'État, opérateurs de compétences...). En revanche, cette cellule n'aurait pas vocation à répondre aux demandes d'accès aux données à des fins de recherche ou d'évaluation, qui continueront d'être traitées par le CASD, ni à des demandes réalisées dans un but lucratif par des acteurs privés qui pourraient être néanmoins être satisfaites dans le cadre de la stratégie d'« open data ».

[96] En revanche, ce LDH n'aurait pas vocation à exploiter lui-même les données issues de la DSN, en particulier à réaliser les appariements attendus.

⁸³ Une articulation devra être recherchée entre cette communauté des utilisateurs et la comitologie « data » qui pourrait être mise en place dans le cadre de France Travail.

[97] Pour financer les investissements qui devraient accompagner la mise en place de cette cellule (par exemple, la création d'un portail présentant les bases de données existantes, le cadre juridique, la procédure d'accès, etc.), une demande de financement pourrait être adressée au Fonds de financement de la transformation publique (FTAP) qui vient d'être doté de 10 M€ pour faire émerger des projets de valorisation des données. À titre de comparaison, la Direction générale des finances publiques, mais aussi les Douanes, ont bénéficié ces dernières années d'un soutien du FTAP pour développer leurs démarches de valorisation des données. La Dares et le SSM de l'Éducation nationale ont également obtenu un soutien du FTAP pour le projet InserJeunes.

[98] Un bilan pourrait être réalisé à horizon trois ans afin d'estimer l'adéquation de cette organisation par rapport aux besoins des utilisateurs des données de la DSN à des fins de pilotage.

Recommandation n°7 Créer une cellule dénommée Labor data hub (LDH) composée de 2-3 ETP auprès du directeur de la Dares responsable, en lien avec la Dnum, d'orchestrer et de mettre en œuvre les fonctions nécessaires à une meilleure utilisation des données de la DSN à des fins de pilotage

3.4.2 Un service d'appariement centré sur la mesure du retour à l'emploi pourrait être offert par ce « Labor data hub » en lien avec le CASD

[99] Un des cas d'usage majeurs rencontrés par la mission concerne la réalisation d'appariements avec les données de la DSN pour réaliser des études d'impact de certains dispositifs de politiques publiques (ex : CEJ, RSA, alternance...). Or il n'existe pas d'offre standard pour un tel service, alors que ce besoin concerne de nombreux acteurs (DGEFP, Conseils départementaux, start-up d'État, Opco...).

[100] Il serait efficient d'organiser un tel service en un seul point d'entrée, avec une méthodologie standardisée. La mission recommande de confier la conception et l'organisation de ce service au « Labor data hub ». En effet, son positionnement au sein de la Dares lui permet de bénéficier de son expertise en matière d'appariements et d'évaluation des politiques publiques. En pratique, l'appariement serait défini par le « hub » (données en entrée, indicateurs en sortie...) puis opéré par le CASD⁸⁴. Il dispose d'un savoir-faire et d'une procédure sécurisée qui lui permet d'effectuer régulièrement des appariements pour des projets de recherche ou d'évaluation des politiques publiques. Pour les appariements utilisant un Nir haché, il joue de plus en plus un rôle de tiers de confiance⁸⁵.

⁸⁴ Le CASD fait des appariements à partir de traits d'identité ou en faisant le hachage du NIR : le coût d'un projet ponctuel d'appariement, estimé actuellement à 20 000 euros (30 jours), est surtout administratif (gestion du projet, assistance juridique et RGPD, conventionnement), la réalisation technique ne prenant quant à elle que 2 à 3 jours. Dans le cadre d'une offre de service structurée, certains coûts pourraient être mutualisés.

⁸⁵ <https://www.insee.fr/fr/statistiques/fichier/4254170/courstat-3.pdf> 2019 (p. 85-87). En revanche, le CASD ne pourrait pas piloter la conception du service (indicateurs de suivi...) ni l'instruction des demandes techniques en raison de son manque d'expertise sur les problématiques emploi-formation professionnelle, les concepts d'emploi et la DSN.

Recommandation n°8 Concevoir un service de croisement de données (appariements) permettant de mesurer de façon efficiente et cohérente l'impact sur le marché du travail, notamment le retour à l'emploi, de dispositifs de politiques publiques

[101] De façon plus générale, l'existence de diverses bases de données aval pose la question de l'efficience globale des moyens et de la redondance de certains traitements. La mission a ainsi relevé que les bases de données produites par le Gip MDS (Dad-e et PIQ) étaient sollicitées pour répondre à des demandes dans le domaine du marché du travail, avec un risque de redondance avec d'autres acteurs plus familiers des concepts pertinents dans le domaine de l'analyse du marché du travail. La mission recommande de limiter ce type d'activité et de revoir l'offre « data » inscrite dans la feuille de route du GIP-MDS, cette production d'indicateurs nécessitant une expertise plus large que la seule expertise de la norme et de la déclaration.

Recommandation n°9 Recentrer l'offre de services « data » du Gip-MDS sur la fiabilisation de la déclaration et sur l'exploitation précoce de données brutes présentes en DSN ne nécessitant pas de faire appel à des concepts

François BRUNET

Irina SCHAPIRA

Nicole ROTH

ANNEXE 1 : Caractérisation des bases de données issues de la DSN

ANNEXE 1 : CARACTERISATION DES BASES DE DONNEES ISSUES DE LA DSN	57
1 LES FLUX DSN ALIMENTENT PLUSIEURS BASES DE DONNEES AVAL PLUS OU MOINS ELABOREES	59
1.1 LA DSN EST CONSTITUEE A PARTIR DES FLUX DECLARATIFS DES EMPLOYEURS	59
1.1.1 <i>Vue simplifiée des flux amont</i>	59
1.1.2 <i>Description succincte des blocs de la DSN</i>	60
1.1.3 <i>Le Bloc 1 est le point de dépôt ou de rejet</i>	61
1.1.4 <i>Le Bloc 3 est la source des données nominatives</i>	62
1.1.5 <i>Le déclarant peut différer la correction d'anomalies</i>	62
1.2 UN APERÇU DES SEPT BASES AVAL COUVRANT LA PLUPART DES SALAIRES DE LA DSN	63
1.2.1 <i>Le Gip-MDS effectue quelques traitements pour constituer les bases miroirs Dad-e et PIQ</i>	63
1.2.2 <i>A partir de DSNP3 (données brutes), la Dares construit Sismmo et Post-Sismmo, qui répondent à des besoins différents</i>	64
1.2.3 <i>L'Insee produit la base statistique BTS et ses dérivés</i>	66
1.2.4 <i>L'Urssaf CN construit la base « Tables Tout en un » (TTE1)</i>	66
1.3 TROIS BASES AVAL SUR UNE PARTIE DES SALAIRES DE LA DSN	67
1.3.1 <i>DSNPE : les intérimaires et demandeurs d'emploi</i>	67
1.3.2 <i>GOETH : les travailleurs handicapés</i>	68
1.3.3 <i>Agora : les personnes ayant bénéficié d'une formation professionnelle et leur suivi</i>	68
1.4 L'EXISTENCE DE BASES MULTIPLES INTERROGE SUR UN RENFORCEMENT DE LA COORDINATION ET SUR L'INFORMATION DE LEURS UTILISATEURS	68
2 LES DIMENSIONS A PRENDRE EN COMPTE POUR UNE CARTOGRAPHIE DES DIFFERENTES BASES DE DONNEES AVAL DE LA DSN	69
2.1 LES REFERENCES SUR LA QUALITE DES DONNEES : DES SIMILITUDES ENTRE LES DOMAINES DONNEES ET STATISTIQUES	69
2.1.1 <i>Références dans le domaine de la donnée</i>	69
2.1.2 <i>Références dans le domaine statistique</i>	70
2.2 DIMENSIONS POUR APPREHENDER LA QUALITE DES BASES DE DONNEES	71
2.2.1 <i>Les caractéristiques globales : finalité, périmètre couvert, délai de mise à disposition et opposabilité</i>	72
2.2.2 <i>Les principaux retraitements : unités élémentaires, informations expertisées et concepts construits</i>	72
2.2.3 <i>Les autres attributs : comparabilité temporelle, recul historique, nomenclatures et interopérabilité</i>	73
2.3 DIMENSIONS CONCERNANT L'ACCESSIBILITE AUX BASES DE DONNEES	73

2.3.1	Modalité d'accès à la base.....	73
2.3.2	Documentation sur les données, leurs traitements constitutifs, leurs usages et limites ainsi que l'investissement pour les utiliser	74
2.3.3	Critères généraux : taux de satisfaction des utilisateurs et usages actuels	74
2.4	LA QUALITE D'UNE BASE DEPEND DE L'OBJECTIF POURSUIVI.....	74
3	UNE TENTATIVE DE CARTOGRAPHIE DES BASES ISSUES DE LA DSN	75
3.1	LE « BLOC 3 » DE LA DSN NATIVE EST TRES DOCUMENTE	75
3.2	DAD-E ET PIQ : DES BASES MENSUELLES DU GIP-MDS PROCHES DU BLOC 3	76
3.2.1	Dad-e, base très proche du Bloc 3	76
3.2.2	PIQ, base anonymisée et enrichie	77
3.3	TABLES TOUT EN UN (TTE1) DE L'URSSAF CN, UNE BASE MENSUELLE EXPERTISEE ET ENRICHE	78
3.4	DSNP3-SISMMO DES BASES MENSUELLES PLUS OU MOINS ENRICHIES.....	79
3.4.1	DSNP3 de la Dnum, une base mensuelle très proche du Bloc 3	79
3.4.2	Sismmo de la Dares, une base mensuelle proche du Bloc 3 chaînant les contrats... ..	80
3.4.3	Post-SISMMO, une base mensuelle expertisée et enrichie par la Dares.....	81
3.5	BASE TOUS SALAIRES DE L'INSEE, UNE BASE ANNUELLE EXPERTISEE TRES RICHE	83
3.6	DES POINTS A AMELIORER RELATIFS A L'ACCES AUX BASES DE DONNEES	85
3.6.1	Les informations sur les traitements des bases de données aval	85
3.6.2	Les informations sur la qualité des données et leurs usages.....	85
3.6.3	Les informations sur la procédure pour obtenir les données	85
3.7	UNE CARTOGRAPHIE A AFFINER ET FAIRE CONNAITRE DANS UN CADRE PLUS LARGE.....	86
3.7.1	Une cartographie à affiner et à actualiser régulièrement	86
3.7.2	Une cartographie à faire connaître dans un cadre plus large	86

1 Les flux DSN alimentent plusieurs bases de données aval plus ou moins élaborées

[1] Il existe plusieurs bases « DSN » : en effet le flux de déclarations sociales est transmis à un grand nombre d'acteurs dont plusieurs se servent pour constituer une base de données. Ici seules les bases de données utilisées à des fins de pilotage et de statistiques publiques sont présentées et analysées.

1.1 La DSN est constituée à partir des flux déclaratifs des employeurs

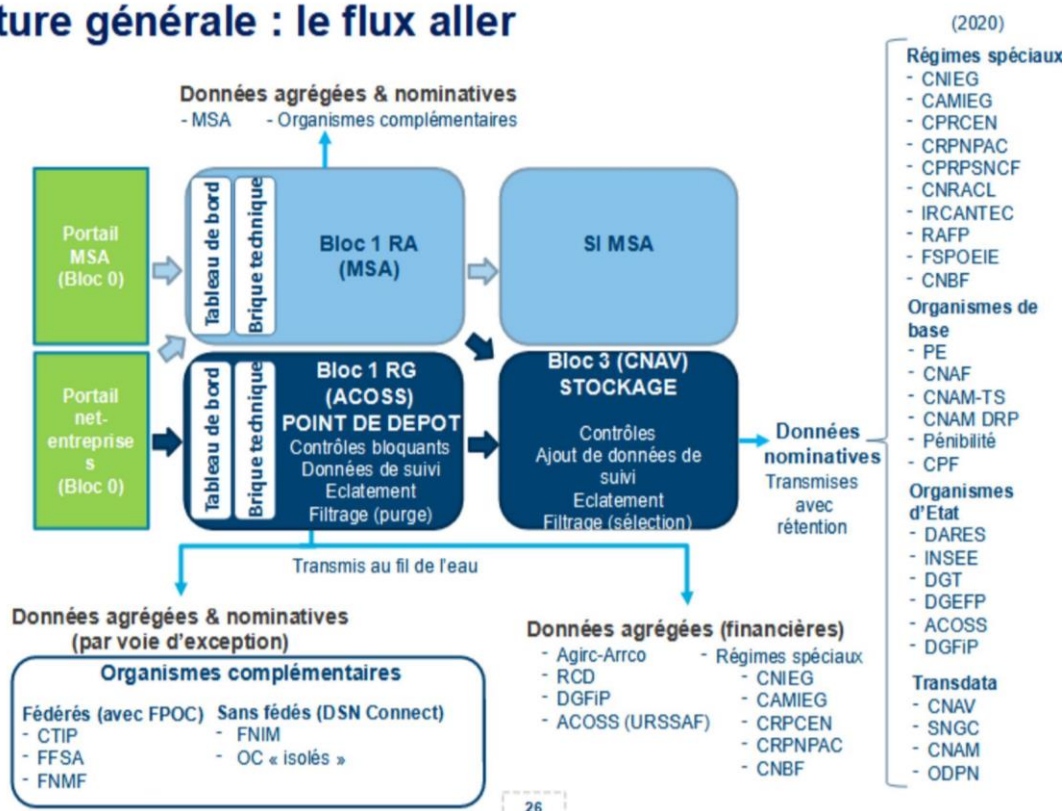
[2] En amont des bases de données permettant d'accéder à des informations provenant de la DSN, les entreprises font leurs déclarations à partir de leur système d'information sur le portail de net-entreprises, développé par le Gip-MDS, ou sur celui de la MSA. En fonction de leur taille, elles le font au plus tard le 5 ou le 15 du mois M+1 pour les données relatives au mois M.

1.1.1 Vue simplifiée des flux amont

[3] De façon agrégée, le schéma ci-après décrit les grands flux de données à l'aller, du déclarant vers l'administration hors traitement des anomalies au regard de la norme. Un schéma au point 1.1.4 indique les flux quand les contrôles signalent des anomalies.

Schéma 1 : Architecture des flux DSN

Architecture générale : le flux aller



Source : Gip-MDS

[4] **Ce schéma liste les organismes et administrations qui sont destinataires de la DSN selon deux niveaux de granularité : données agrégées (au niveau des entreprises) et/ou données nominatives.** Les données nominatives sont transmises à chaque destinataire pour les seules informations le concernant d'après la réglementation. Le schéma signale les principaux « blocs » de la DSN : 0, 1 et 3 (concernant la notion de « bloc », voir ci-après). Il en existe d'autres qui sont décrits brièvement dans le tableau 1 *infra*.

1.1.2 Description succincte des blocs de la DSN

[5] **L'architecture DSN est composée de « blocs », c'est-à-dire de modules techniques s'articulant les uns avec les autres, et composant l'ensemble du système DSN.** A ces blocs, de tailles très variables, sont associées des responsabilités spécifiques en termes de maîtrise d'œuvre et d'exploitation. La description succincte des blocs et les rôles des acteurs figurent dans le tableau ci-dessous :

Tableau 1 : Description des processus associés aux blocs de la DSN

BLOC	Description du processus	MOE	Exploitation
Bloc 0	Inscription du déclarant (employeurs ou tiers-déclarant) sur net-entreprises.fr	GIP MDS	GIP MDS
RCD	Vérification du SIRET	URSSAF CN	URSSAF CN
Bloc 1	Point de dépôt ou de rejet	URSSAF CN/MSA	URSSAF CN//MSA
Brique de contrôle	Vérification Technique	CNAV	URSSAF CN//MSA
API DSN	Bloc exclusivement technique	GIP MDS	URSSAF CN//MSA
Gestion mandats	Vérification des coordonnées bancaires	GIP MDS	GIP MDS
DSN connect	Envoi des DSN aux Organismes Complémentaires « isolés »	GIP MDS	GIP MDS
Bloc 3	Point de stockage	CNAV	CNAV
Dad-e	Extraction mensuelle des données dans une base miroir construite autour de l'identité du salarié pour permettre un accès optimisé en performance aux informations individuelles	GIP MDS	CNAV
CID	Contrôles inter-déclarations pour s'assurer de la cohérence dans le temps entre deux DSN	CNAV	CNAV
DSN distrib	Envois aux organismes ne réceptionnant pas au fil de l'eau	GIP MDS	GIP MDS
PAS CRM	Mise à disposition des comptes rendus métier nominatifs et financiers pour le déclarant concerné	GIP MDS	GIP MDS
Net-CRM	Mise à disposition des taux d'accidents du travail pour le déclarant concerné	GIP MDS	GIP MDS
Nomenclatures	Mise à disposition de référentiels	GIP MDS	GIP MDS
FPOC	Mise à disposition des fiches de paramétrage des OC	GIP MDS	GIP MDS
Hypervision	Bloc exclusivement technique	GIP MDS	GIP MDS

BLOC	Description du processus	MOE	Exploitation
Requêtage (PIQ)	Extraction de Dad-e afin d'avoir une base de requêtage anonymisée pour optimiser la qualité des données	GIP MDS	CNAV

Source : DSS

[6] **La DSS assure la maîtrise d'ouvrage stratégique (MOAS) de l'ensemble de la DSN et le GIP MDS la MOA opérationnelle.**

1.1.3 Le Bloc 1 est le point de dépôt ou de rejet

[7] **Le Bloc 1 est le point de dépôt dans lequel sont effectués des contrôles notamment du déclarant, de caractéristiques générales du fichier (taille, nom et format), des Siret (via le RCD, répertoire commun des déclarants) et de conformité à la norme⁸⁶.** Les contrôles de conformité à la norme sont de cinq types, ceux des deux premiers types sont explicités dans le cahier des charges de la norme (NÉODES - Norme d'Échanges Optimisée des Données Sociales)⁸⁷ :

- contrôles de cohérence bloquants (uniquement dans le bloc 1) : si un seul contrôle est erroné, la DSN est rejetée ;
- contrôles de signalement non bloquants : la DSN est rejetée à partir de 1 000 anomalies de ce type ;
- contrôles des référentiels externes : vérifient que les valeurs utilisées appartiennent bien aux nomenclatures utilisées par le cahier technique ;
- contrôles de syntaxe : vérifient le format de la rubrique (longueur, nature, expression régulière) ;
- contrôles de structure : vérifient que l'enchaînement des rubriques et la structure du message sont conformes au modèle.

[8] **Si les contrôles s'avèrent non-bloquants, le système envoie au déclarant un accusé d'enregistrement électronique et transmet sa déclaration au Bloc 3.** En parallèle, un flux de données agrégées en sortie du Bloc 1 est transmis à l'Urssaf CN ou à la CCMSA et à d'autres acteurs mais pas au ministère du travail ni à l'Insee.

[9] Si les contrôles s'avèrent bloquants, alors le système transmet au déclarant un avis de rejet de la DSN. Celle-ci n'est pas transmise aux organismes et est à refaire par le déclarant avant la date limite qui reste la même.

⁸⁶ La norme découle de la définition d'un modèle conceptuel de données commun, qui a constitué la clé qui a permis de faire fonctionner la DSN

⁸⁷ Un certain nombre d'outils sont en ligne pour faciliter l'accès à la norme, outre le cahier technique : un outil de contrôle DSN VAL qui teste un fichier DSN avant de le déposer, une brique de contrôle DSN qui intègre l'outil de contrôle dans un logiciel éditeur, et des tables de nomenclatures.

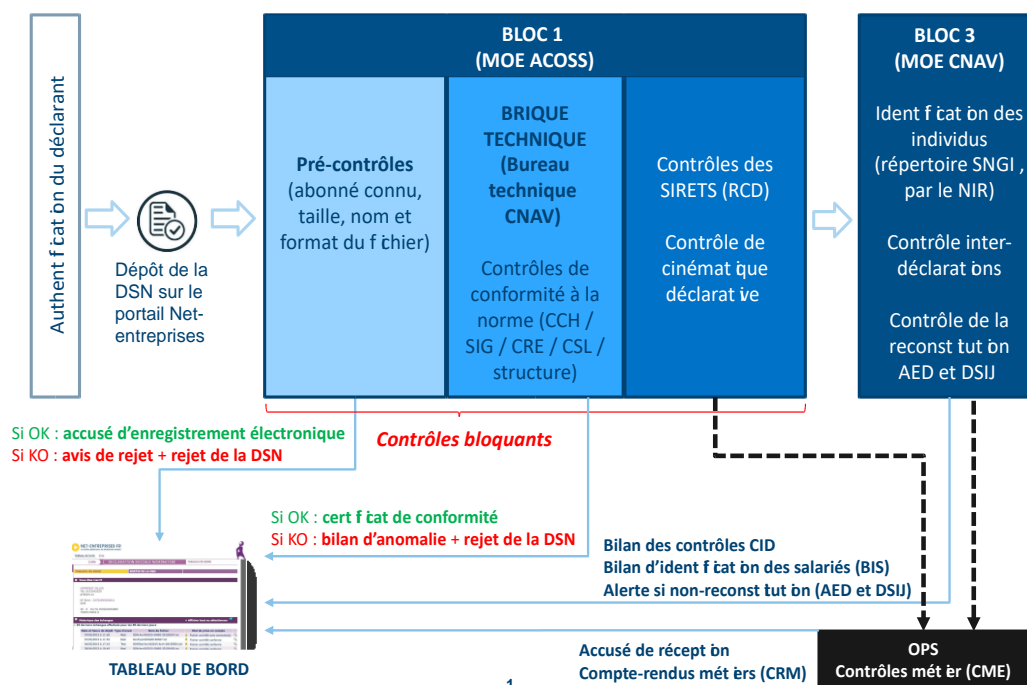
1.1.4 Le Bloc 3 est la source des données nominatives

[10] **Le Bloc 3 réceptionne les déclarations qui ont passé les contrôles du Bloc 1, du régime agricole et du régime général des autres régimes.** Il est le point de stockage dans lequel l'identification des individus est vérifiée et le NIR certifié par la Cnav, qui assure la maîtrise d'œuvre de ces traitements (répertoire SNGI-système national de gestion des identifiants). D'autres contrôles non-bloquants sont faits, dont les contrôles inter-déclarations (qui vérifient la cohérence entre deux déclarations déposées successivement par un même établissement) et le contrôle de la déclaration de salaire pour le versement des Indemnités Journalières. Un flux de données nominatives en sortie du Bloc 3 est transmis aux nombreux destinataires prévus par le décret, dont l'Urssaf CN, le ministère du travail (Dares, DGT et DGEFP), l'Insee, Pôle Emploi et la Caisse des dépôts au titre des divers régimes et dispositifs qu'elle gère (CNRACL, IRCANTEC, CPF...).

1.1.5 Le déclarant peut différer la correction d'anomalies

[11] **Le déclarant accède à un tableau de bord qui lui indique pendant trois mois le résultat des contrôles effectués dans le Bloc 1, dans le Bloc 3 et post-Bloc 3 par sept organismes sur leur champ (DGFIP, Cnav, Cnam, Urssaf, Agirc-Arrco...).** Seul le bloc 1 comporte des contrôles bloquants exigeant une nouvelle déclaration. Pour les autres anomalies signalées, en cas d'erreur, une rectification est admise dans la paie du mois suivant.

Schéma 2 : Flux DSN en cas d'anomalie



Source : Gip-MDS⁸⁸

⁸⁸ AED (Attestation employeur destinée à Pôle Emploi), DSIJ (Déclaration des indemnités journalières)

[12] Le Bloc 3 est soumis à la norme de qualité NÉODeS (contrôles bloquants, non-bloquants ou pas de contrôle). La définition des contrôles suit un équilibre entre fiabilité des données et capacité des entreprises à répondre. Des contraintes trop fortes pourraient empêcher une entreprise de déclarer une situation particulière (un salaire atypique...). Des contraintes trop molles fourniraient des données de qualité insuffisante. L'enjeu est important car toutes les bases en aval du Bloc 3 sont tributaires de la qualité déclarative du Bloc 3.

1.2 Un aperçu des sept bases aval couvrant la plupart des salariés de la DSN

[13] Parmi les bases en aval de la DSN qui servent des usages à des fins de pilotage ou d'information générale (statistiques publiques), certaines couvrent tous les salariés ou presque, alors que d'autres n'en couvrent qu'une partie minoritaire (voir 1.3). Les sources généralistes qui couvrent la plupart des salariés sont produites par le Gip-MDS, le ministère du travail (Dares et Dnum), l'Urssaf CN et l'Insee.

1.2.1 Le Gip-MDS effectue quelques traitements pour constituer les bases miroirs Dad-e et PIQ

[14] Le Gip-MDS construit chaque mois deux bases à partir du Bloc 3. Elles couvrent tous les salariés de la DSN.

- **Base Dad-e**

[15] **A l'origine, il s'agissait de construire une base miroir de la DSN sur laquelle appuyer des webservices en attendant l'existence d'une base miroir de la DRM** (dispositif de ressources mensuelles, alimenté par la DSN et la base Pasrau des revenus de remplacement), **afin que les salariés puissent exercer leur droit d'accès à leurs données** (droit RGPD demandé par la Cnil). Cette base constitue une image proche du Bloc 3 à quelques différences près : suppression de plusieurs éléments de cotisation, suppression des doublons de versement (même montant, même numéro de versement) et gestion de changements implicites (par exemple un contrat dont le numéro change sans déclaration explicite de ce changement).

[16] Dad-e sert au Gip également pour vérifier rapidement la qualité des déclarations afin d'intervenir au plus vite auprès des éditeurs.

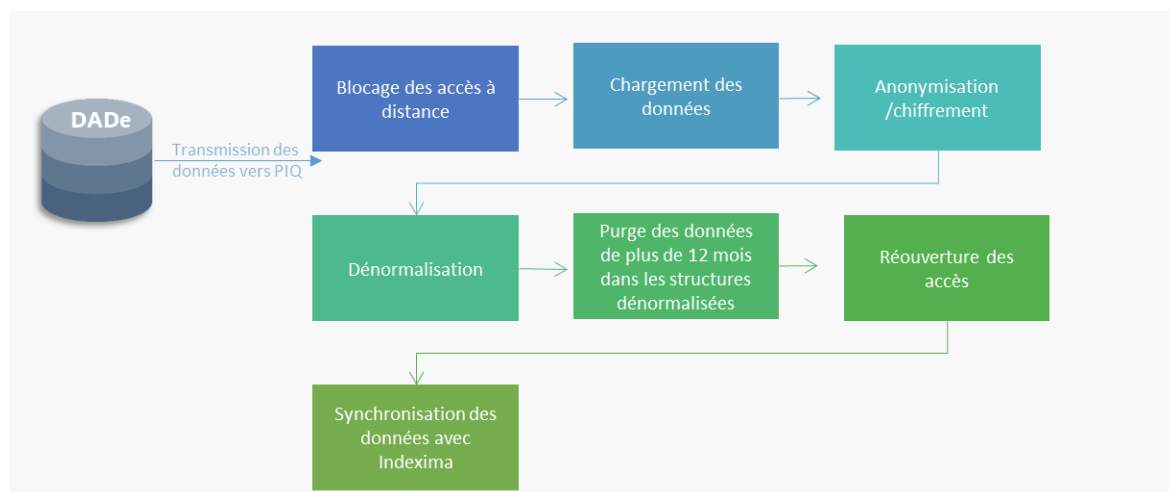
[17] Le Gip-MDS utilise aussi Dad-e pour tester la pertinence du besoin quand il reçoit une demande d'évolution de la DSN. L'usage de Dad-e se justifie enfin par la difficulté importante à effectuer une pseudonymisation directement à partir du Bloc 3, car le stockage des données personnelles est disséminé dans le Bloc 3 alors qu'il est clairement organisé en fonction de l'individu dans Dad-e.

- **Base PIQ (Production d'Indicateurs Qualité)**

[18] **PIQ est une image de Dad-e qui est pseudonymisée : c'est-à-dire qu'elle est expurgée des éléments permettant d'identifier directement une personne** (Nir, nom, adresse personnelle, jour et mois de naissance) mais il reste des éléments qui peuvent permettre une identification indirecte des personnes (année de naissance, sexe, Siret de l'employeur...).

[19] Au départ, PIQ est un outil de production d'indicateurs qualité à usage interne du Gip-MDS. C'est une base de requêtage qui lui permet d'assurer les missions de suivi quantitatif et qualitatif de la DSN. Ainsi le Gip peut s'assurer au fil des mois du caractère satisfaisant des réceptions de la DSN. Cette base permet aussi des usages qualifiés de « décisionnels » par le Gip-MDS (aide à la décision par des éléments chiffrés imprécis) et l'établissement de statistiques en réponse à des demandes d'indicateurs par des tiers pour du suivi ou du pilotage de politique publique.

Schéma 3 : Processus des traitements des données via PIQ

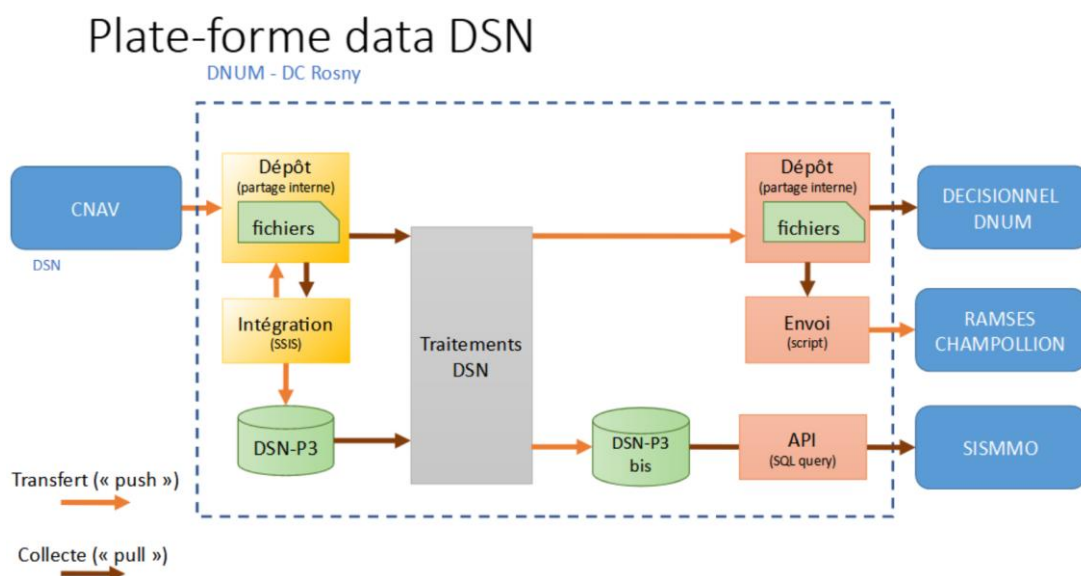


Source : Gip-MDS

1.2.2 A partir de DSNP3 (données brutes), la Dares construit Sismmo et Post-Sismmo, qui répondent à des besoins différents

[20] **DSNP3 (DSN phase 3) est la copie de l'extrait du Bloc 3 reçu par les ministères sociaux,** contenant les informations auxquelles les directions ont droit. En tant que direction informatique pour les directions destinataires de la DSN (Dares, DGT, DGEFP, etc.), la Dnum réceptionne les données et les met en forme.

Schéma 4 : Alimentation de DSNP3 par les données issues du bloc 3 de la DSN

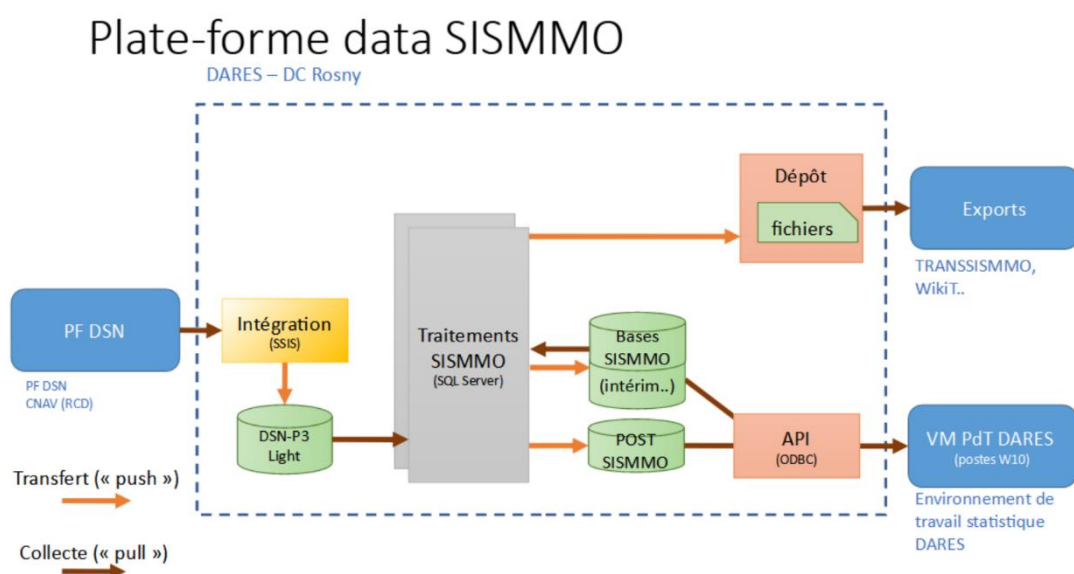


Source : Dnum

[21] La base Sismmo résulte d'un chaînage des contrats de DSNP3, dont elle conserve presque tout le contenu à quelques simplifications près (suppression des informations démultipliées comme le nom des personnes), en faisant un peu d'interprétation quand il y a deux informations différentes sur un même contrat. La Dares l'utilise pour répondre à des demandes de données individuelles opposables.

[22] La base Post-Sismmo (projet aboutissant en 2023) vise à redresser de façon automatique les anomalies de Sismmo, notamment les mauvaises déclarations de contrat, de convention collective... (voir annexe n°1 pour le détail des traitements). La Dares prévoit de l'utiliser pour des usages à des fins d'information générale (diffusion de statistiques, études) et d'évaluations.

Schéma 5 : Alimentation de Sismmo et post-Sismmo par les données issues de la DSN



Source : Dnum

[23] La base Sismmo comprend plusieurs bases dont Sismmo-intérim sur un champ réduit de la DSN, celui du travail temporaire, de façon à obtenir des informations mensuelles sur ce secteur-clé dans l'analyse conjoncturelle de l'emploi. Et, en attendant Post-Sismmo, la base Sismmo comporte la base dite MMO des mouvements par établissement, qui est le prolongement des sources existantes avant l'arrivée de la DSN.

1.2.3 L'Insee produit la base statistique BTS et ses dérivés

[24] Le système Siera de l'Insee (Système d'information sur l'emploi et les revenus d'activité) intègre la copie de l'extrait du Bloc 3 que l'institut reçoit. Il produit plusieurs bases de données redressées et enrichies (voir annexe n°1).

[25] La base Tous salariés (BTS) couvre un champ plus large que la DSN : elle rassemble les déclarations des salariés de la DSN avec celles des particuliers-employeurs (qui ne sont pas encore couverts par la DSN). Elle couvrirait déjà les salariés du secteur public avant que ceux-ci n'intègrent la DSN, en mobilisant par exemple les fichiers de paye de la fonction publique d'Etat.

[26] Les bases Flores (Fichier localisé des rémunérations et de l'emploi salarié) et Panels tous salariés sont dérivées de BTS : Flores sert à des usages territorialisés à un niveau fin en raison des redressements faits sur la localisation de l'emploi et les Panels facilitent l'étude des trajectoires professionnelles des salariés sur longue période.

1.2.4 L'Urssaf CN construit la base « Tables Tout en un » (TTE1)

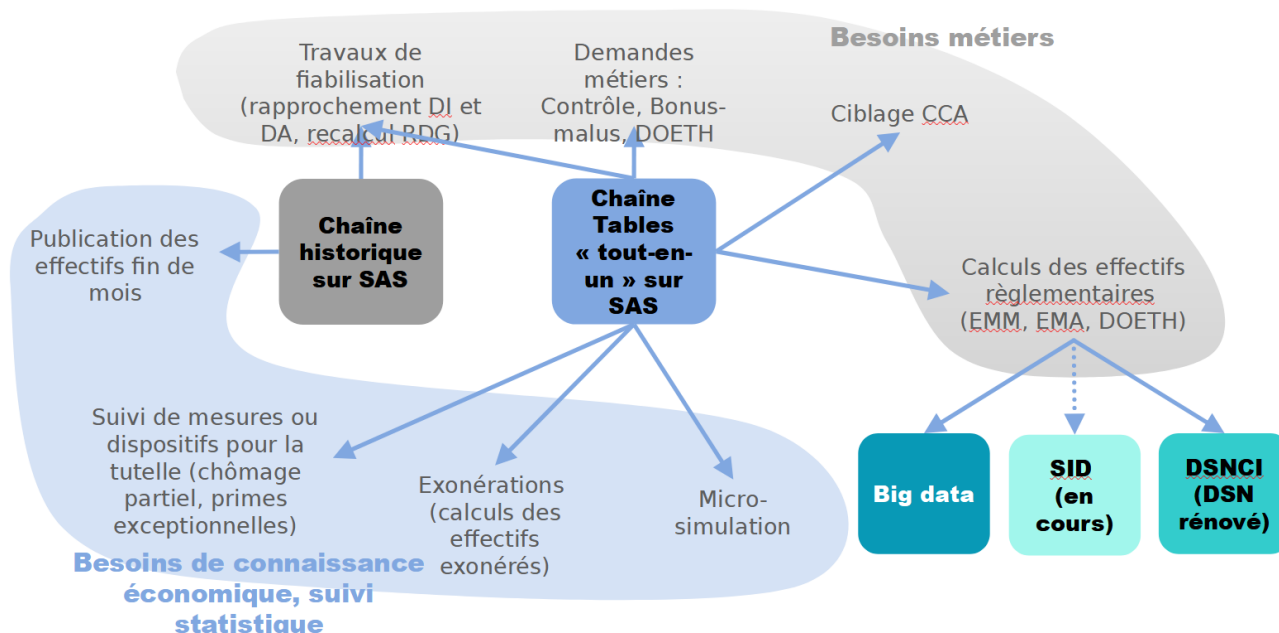
[27] La base TTE1 se prête à de multiples usages : de la fiabilisation de la DSN à la fourniture d'informations de pilotage et statistiques. Elle est alimentée par les Blocs 1 et 3 sur le champ des DSN valides du régime général couvert par les Urssaf. Elle intègre de nombreux traitements dont la mise en cohérence entre les données agrégées (DA) et les données individuelles (DI), et la prise en compte des régularisations de montants.

Schéma 6 : Chaînes statistiques utilisées par l'Urssaf CN



Les chaînes statistiques existantes répondent à des besoins différents

Au service de notre protection sociale



Source : Urssaf CN⁸⁹

1.3 Trois bases aval sur une partie des salariés de la DSN

[28] Il existe par ailleurs plusieurs bases de données qui portent sur une partie des salariés. Seules les trois plus volumineuses sont décrites ici et brièvement. Les autres sont évoquées dans la partie 3 qui dresse la cartographie des principales bases de données.

1.3.1 DSNPE : les intérimaires et demandeurs d'emploi

[29] Pôle emploi reçoit la DSN sur une partie des salariés : ceux relevant des entreprises de travail temporaire (intérimaires et salariés des ETT) et les demandeurs d'emploi, en cours (lorsqu'ils exercent une activité salariée) ou sortis des listes depuis moins de six mois (si la reprise d'emploi s'est effectuée via le salariat). L'opérateur reçoit également toutes les déclarations de fins de contrat, au titre notamment de ses activités de gestion.

[30] La base DSNPE réceptionne ces données, les traite et alimente des applicatifs métiers : indemnisation et placement des demandeurs d'emploi à partir de données individuelles, statistiques et pilotage à partir de données agrégées.

⁸⁹ CCA (contrôle comptable d'assiette), DOETH (déclaration obligatoire d'emploi des travailleurs handicapés), EMA (effectifs moyens annuels), EMM (effectifs moyens mensuels), RDG (règles de gestion), SAS (logiciel de traitements statistiques)

1.3.2 GOETH : les travailleurs handicapés

[31] A la suite du transfert de la déclaration d'obligation d'emploi des travailleurs handicapés vers la DSN au 1er janvier 2021, l'Agefiph a construit la base GOETH pour réceptionner les données provenant de la DSN, non pas depuis le Bloc 3 mais depuis Dad-e, la base étant complétée par des données de la MSA et de l'Urssaf CN issues du RCD (dont effectifs et contacts des établissements).

[32] L'Agefiph reçoit des données de la DSN uniquement sur le champ des salariés ayant signalé à leur entreprise qu'ils sont bénéficiaires de la reconnaissance de travailleur handicapé (RQTH).

[33] Les rubriques de la DSN portant sur l'obligation d'emploi de travailleurs handicapés (OETH), introduites seulement à partir de 2020, sont à ce stade de qualité encore très insuffisante. L'Agefiph considère qu'elle ne peut s'acquitter de ses missions que dans des conditions dégradées : elle doit aller chercher de l'information complémentaire dans d'autres sources de données ou directement auprès des entreprises pour mener des analyses opérationnelles territoriales et sectorielles, pour connaître la situation d'un établissement ou d'une entreprise au regard de son obligation d'emploi et ainsi l'accompagner dans sa politique d'emploi. Par exemple, l'Agefiph perdrait 20 % des cotisations si elle utilisait uniquement les données de la DSN.

1.3.3 Agora : les personnes ayant bénéficié d'une formation professionnelle et leur suivi

[34] Agora est le système d'information de la Caisse des dépôts qui a été construit afin de regrouper les informations sur les formations ayant bénéficié d'une aide financière publique, quel que soit leur financeur public, puis de les partager entre les différents financeurs. Le SI comprend notamment des données issues de la DSN, afin de suivre les parcours en emploi des bénéficiaires d'une formation. Il est alimenté par des données du Bloc 3 sur les seules personnes qui sont en train de suivre une formation financée ou qui en ont suivi une depuis moins de 24 mois.

1.4 L'existence de bases multiples interroge sur un renforcement de la coordination et sur l'information de leurs utilisateurs

[35] La présence de sept bases en aval de la DSN sur l'ensemble des salariés (ou presque) s'explique par le fait que ces bases servent des usages propres à l'organisme qui les produit. Cependant certains traitements seraient similaires ou viseraient des objectifs similaires. Cette affirmation est au conditionnel car sa validation nécessiterait de rentrer profondément dans la mécanique des traitements de chaque base : un tel rapprochement ne peut donc être mené que par les producteurs des bases de données eux-mêmes, et en tout état de cause il ne relève pas de la présente mission.

[36] Une coordination existe déjà entre certains acteurs. Au-delà des échanges informels, il existe trois espaces principaux de coordination structurés :

- les ateliers sur la qualité de la DSN, animés par le Gip-MDS, qui n'ont certes pas pour objet la coordination des bases en aval de la DSN, réunissent néanmoins entre autres leurs

producteurs et explicitent les variables de la DSN qui leur servent ensuite à la fabrication de leurs bases ;

- l'Insee, l'Urssaf CN et la Dares se coordonnent notamment pour harmoniser la définition de concepts utilisés dans leurs bases (l'emploi au sens du Bureau international du travail, le secteur public) et pour coproduire les estimations trimestrielles d'emploi, avec un partage du champ (voir annexe Statistiques publiques) ;
- un projet s'est lancé entre l'Insee, la Dares, Pôle emploi et le Gip-MDS pour mieux articuler les codes PCS et Rome en vue d'améliorer la qualité de la PCS native dans la DSN et de l'articuler avec la nomenclature Rome.

[37] La DSN a connu une longue phase de montée en puissance jusqu'en 2022 avec l'intégration de l'ensemble de la fonction publique, durant laquelle la priorité des producteurs était de s'adapter pour continuer à produire. La DSN étant arrivée à une phase plus stable, le moment semble favorable pour instaurer davantage de coordination sur les méthodes et outputs entre les producteurs des bases de données issues de la DSN.

[38] Enfin, devant la palette de bases disponibles se pose la question de leurs usages privilégiés. Un utilisateur qui voudrait utiliser la DSN se trouve devant une question loin d'être évidente aujourd'hui : « quelle est la base la plus adaptée à mon besoin ? ». En effet, il manque une cartographie qui présente les objets des bases et leurs avantages comparatifs selon les usages : richesse du contenu, mode d'accès, appropriation... Plusieurs critères entrent en jeu. La partie suivante les décrit puis la partie 3 dresse une première cartographie selon ces critères.

2 Les dimensions à prendre en compte pour une cartographie des différentes bases de données aval de la DSN

2.1 Les références sur la qualité des données : des similitudes entre les domaines données et statistiques

[39] Les références sur la qualité des données, produites par les administrations au niveau européen et français, sont différentes dans le domaine de la donnée et dans celui de la statistique, car elles visent des usages différents, mais elles comportent aussi un grand nombre de points communs.

2.1.1 Références dans le domaine de la donnée

[40] Au niveau européen, l'Union européenne fournit des lignes directrices pour la qualité des données⁹⁰ qui se construit en six étapes : définition du profil de la base de données (compréhension des données et des résultats), nettoyage des données pour les apurer et les structurer, enrichissement des données, documentation des données, estimation de la qualité des données, publication des données en format accessible.

⁹⁰ Data.europa.eu, Data Quality Guidelines, 2022, data.europa.eu/doi/10.2830/333095

[41] Ces lignes directrices citent plusieurs critères de qualité dont l'exhaustivité, l'accessibilité, l'interopérabilité, la fraîcheur, la cohérence, l'exactitude, la pertinence et la lisibilité des données.

[42] Elles incitent à une documentation fournie, un rapport sur la validation des données, un document sur les règles de validation, des recommandations d'usage, et à une ouverture des données consultables par des humains et par des machines.

[43] En France, la Dinum présente sur son site⁹¹ les qualités attendues d'une base de données, particulièrement des métadonnées (informations sur les données de la base). Elle recommande un modèle de données stable au fil du temps, des données interopérables et des données pivots qui correspondent à un référentiel public. Pour elle, ces données pivots présentent plusieurs avantages permettant une meilleure formalisation, une meilleure synthèse, une meilleure compréhension, une meilleure réutilisation et une meilleure interopérabilité.

[44] La Dinum a construit un score de qualité des métadonnées qui inclut l'indication de la fréquence de mise à jour, la couverture spatiale et la couverture temporelle. Par ailleurs, la documentation doit indiquer les changements majeurs (ruptures de séries), le mode de production des données, leurs exhaustivité et limites, les points d'attention et précautions d'usage, un dictionnaire des données (définition de chaque variable et de ses valeurs possibles ainsi que des exemples) de façon à faire gagner du temps à l'utilisateur lorsqu'il s'approprie la base de données.

2.1.2 Références dans le domaine statistique

[45] Au niveau européen, le cadre qualité commun du système statistique européen est composé du code des bonnes pratiques de la statistique européenne, complété par le cadre d'assurance qualité du système statistique européen⁹², qui fournit des orientations complémentaires sur la manière d'appliquer le code de bonnes pratiques, et des principes généraux de gestion de la qualité (par exemple, l'interaction constante avec les utilisateurs et l'amélioration continue).

[46] Le code des bonnes pratiques de la statistique européenne⁹³ énonce seize principes. Certains visent particulièrement la qualité des statistiques selon plusieurs angles : engagement sur la qualité (principe 4), impartialité et objectivité (principe 6), solidité de la méthodologie (principe 7), pertinence (principe 11), exactitude et fiabilité (principe 12), actualité et ponctualité (principe 13), cohérence et comparabilité (principe 14) et accessibilité et clarté (principe 15).

[47] Le cadre d'assurance qualité indique des pratiques pour appliquer chacun des principes du code, notamment les suivantes :

- concernant l'engagement sur la qualité, une unité dédiée à la qualité au sein de l'institut de statistique, une description des différentes phases du processus de production statistique s'inscrivant dans le Generic Statistical Business Process Model (GSBPM), publication des lignes directrices en matière de qualité ;

⁹¹ <https://guides.etalab.gouv.fr/qualite/documenter-les-donnees/>

⁹² ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V2.0-final.pdf , « Quality Assurance Framework of the European Statistical System », 2019

⁹³ www.insee.fr/fr/statistiques/fichier/4140105/Code_Bonnes_Pratiques_Stat_Euro_nov2017.pdf

- concernant l'impartialité et l'objectivité, une donnée peut être publiée en cas de problème de qualité à condition de publier également les limites clairement identifiées ;
- concernant la solidité de la méthodologie, la documentation méthodologique est élaborée pour tous les processus statistiques et est diffusée au moins sous la forme d'un résumé qui suit le modèle de la statistique européenne en la matière (Single Integrated Metadata Structure = SIMS) ;
- concernant la pertinence, l'obligation de consulter les utilisateurs, la publication des principaux résultats de la satisfaction des utilisateurs, l'établissement d'accords de coopération et d'accords sur les niveaux de service avec les principaux utilisateurs afin de définir et de documenter les besoins prioritaires ;
- concernant l'exactitude et la fiabilité, la comparaison des résultats et des productions intermédiaires avec d'autres sources d'information pertinentes ;
- concernant l'actualité et la ponctualité, la définition et la publication de règles de publication, l'examen de la possibilité de diffuser des résultats provisoires en tenant compte de l'exactitude et de la fiabilité des données ;
- concernant la cohérence et la comparabilité, un répertoire commun de concepts ou un mécanisme pour promouvoir la cohérence, des lignes directrices devraient traiter de la cohérence entre les microdonnées et les données agrégées, entre les données annuelles, trimestrielles et mensuelles, entre les données nationales et régionales, entre les statistiques de domaine et les comptes nationaux ;
- concernant l'accessibilité et la clarté, les produits statistiques sont diffusés avec les métadonnées pertinentes, les utilisateurs sont consultés sur les formes de diffusion les plus appropriées, un service d'information est en place pour répondre aux demandes et expliquer les résultats statistiques, des règles ou des protocoles pour accéder aux microdonnées sont diffusés sur le site de l'autorité de la statistique.

[48] En France ce code européen s'applique pour la production de l'ensemble des statistiques issues du SSP. De plus, l'Autorité de la statistique publique (ASP) propose deux démarches aux producteurs de statistiques publiques du « péri-SSP » : de labellisation d'une part, de reconnaissance de statistique publique d'autre part : cette dernière retient un certain nombre de critères visant notamment à permettre une information transparente des utilisateurs sur les statistiques produites. Les critères à renseigner dans la documentation sont notamment : unité élémentaire, principaux concepts et définitions, univers, champ géographique, période de référence⁹⁴.

2.2 Dimensions pour appréhender la qualité des bases de données

[49] Parmi les dimensions présentes dans les différentes références citées précédemment, une dizaine a été signalée par nos interlocuteurs sur la qualité propre aux données lors de nos auditions ou dans leurs documents, particulièrement le délai de mise à disposition et la

⁹⁴ www.comite-du-label.fr/wp-content/uploads/2020/07/Plan-type-pour-labellisation.pdf : plan-type du dossier pour labellisation ASP

robustesse. L'accessibilité et la documentation des données, qui font aussi partie de la qualité d'une base de données, sont traitées dans le paragraphe 2.3.

2.2.1 Les caractéristiques globales : finalité, périmètre couvert, délai de mise à disposition et opposabilité

[50] Parmi les dimensions qui permettent de présenter globalement une base de données, l'opposabilité est le seul critère absent des référentiels mentionnés précédemment.

Finalité	Finalité du traitement à l'origine de la base : finalités de gestion, de pilotage, statistiques publiques, d'information générale...
Périmètre couvert	Exhaustivité du champ couvert par la base par rapport à la population de référence Le périmètre peut être supérieur à celui du Bloc 3 par recours à une autre base
Délai de mise à disposition	De quasi-immédiat (soit 20 jours après la fin du mois) à plusieurs mois. En général, ce critère est anticorrélé à celui sur l'expertise des données contenue dans la base de données
Opposabilité	Opposable si la donnée déclarée par l'entreprise est présente

2.2.2 Les principaux retraitements : unités élémentaires, informations expertisées et concepts construits

[51] Ces dimensions indiquent les travaux d'enrichissement et d'expertise appliqués aux données du Bloc 3, dont la qualité dépend de la norme et qui sont soumises aux aléas réglementaires.

Unités élémentaires	<i>Ex : contrat, poste, salarié, établissement ou entreprise</i> La DSN utilise des identifiants reliés à des répertoires pour les salariés (Nir), établissements (Siret) et entreprises (Siren) mais pas pour les contrats (identifiants spécifiques, ce qui peut engendrer des incertitudes sur leur qualité) Les « postes » ne sont pas nativement dans la DSN : ils correspondent à un contrat sauf quand plusieurs contrats se suivent pour un même salarié sur un même poste (par exemple transformation d'un CDD en CDI).
Principales informations expertisées (et redressées)	Informations présentes dans le Bloc 3 expertisées pour déterminer quel traitement leur appliquer, puis redressées. <i>Ex : identification du salarié, éléments du contrat de travail, informations d'ordre économique et financier (cotisations, assiettes, masses salariales), évènements survenus pendant la période déclarée, rémunérations individuelles, durées de travail, effectifs, professions codées selon la nomenclature PCS, localisation de l'emploi.</i>
Principaux concepts construits	Informations supplémentaires, construites à partir des variables initiales de la DSN et s'adaptant aux modifications de la DSN pour représenter un concept stable dans le temps. Certains concepts ne peuvent se construire qu'après un recul d'une année complète de DSN. <i>Ex : l'emploi (au sens du BIT) et sa localisation, le salaire individuel (net, brut, horaire...), les horaires de travail (annuel, hebdomadaire...), les emplois en équivalents temps plein, les effectifs réglementaires, la qualité des emplois</i>

	(emplois stables...).
--	-----------------------

2.2.3 Les autres attributs : comparabilité temporelle, recul historique, nomenclatures et interopérabilité

[52] Les caractéristiques propres aux données peuvent permettre de suivre des évolutions dans le temps, de comparer leurs résultats avec des informations externes ou d'enrichir la base de données.

Comparabilité temporelle	Données mensuelles, trimestrielles ou annuelles Cohérence des évolutions temporelles garantie ou non
Recul historique	De quelques mois à plusieurs dizaines d'années
Nomenclatures	Nomenclatures standards qui font référence présentes dans la base, expertisées ou pas Ex : professions/métiers (PCS, FAP, ROME), secteurs d'activité (NAF, public-privé), conventions collectives (IDCC), géographie (région, département, zone d'emploi, commune).
Interopérabilité	Outre l'utilisation des nomenclatures au niveau agrégé, rapprochement au niveau individuel : * Par des identifiants standards : Siren (entreprises), Siret (établissements), salariés (Nir, Nir haché, CSNS = code statistique non signifiant) * Par des traits d'identité (nom, prénom, sexe, date de naissance, âge... pour les salariés). Opération plus aléatoire qui dépend de la qualité de ces traits dans les bases appariées et qui nécessite souvent un traitement manuel en complément.

2.3 Dimensions concernant l'accessibilité aux bases de données

[53] Une base de données doit aussi pouvoir être accessible et appropriable par les utilisateurs, quel que soit son usage. Ainsi la cartographie comprend, en plus des dimensions précédentes sur la qualité propre aux données, huit dimensions relatives à l'accès aux données en tant que tel, et à la documentation sur cet accès ainsi que celle sur les données.

2.3.1 Modalité d'accès à la base

Modalités d'accès à la base	Accès direct (total ou via un outil d'interrogation) et/ou indirect (via extractions ou tabulations) Critères à satisfaire pour accéder à la base de données, procédure pour y arriver et délai de la procédure Documentation sur ces questions et accessibilité de cette documentation
Coût d'accès	Caractère gratuit ou payant - et coût dans le second cas - selon la nature du demandeur

2.3.2 Documentation sur les données, leurs traitements constitutifs, leurs usages et limites ainsi que l'investissement pour les utiliser

[54] Pour chaque document, outre son niveau de précision, son degré d'accessibilité est important pour faciliter des réutilisations correctes et cohérentes des données : idéalement il s'agira d'un document clair et détaillé, en ligne sur internet et facile à trouver.

Dictionnaire des variables	Informations contenues dans la base de données : définition, valeurs possibles, dictionnaire des modalités
Description des traitements	Traitements automatiques et manuels qui sont à l'origine de la base (sous forme littérale ou sous forme de codes)
Conseils d'usages	Conseils d'usage et limites de la base (sous forme littérale ou sous forme d'information dans la base elle-même)
Investissement pour utiliser les données	Compétences nécessaires pour utiliser les données : structuration des données, logiciels pour les utiliser, API, datavisualisation

2.3.3 Critères généraux : taux de satisfaction des utilisateurs et usages actuels

[55] Ces deux critères sont très généraux : la satisfaction donne une information subjective mais synthétique, le comptage des usages actuels si possible par type d'usage fournit une indication sur la qualité d'une base, qui peut être biaisée s'il y a des bases récentes et/ou pas encore connues, ou si l'intérêt pour un sujet conjugué à la rareté des données conduit les utilisateurs à se satisfaire d'une base de qualité même médiocre. Ces biais peuvent être liés à une surreprésentation des utilisateurs méthodologiquement peu avertis. L'annexe Pilotage fournit une typologie des usages à des fins de pilotage. La cartographie de la partie 3 de la présente annexe utilise une typologie plus synthétique mais balayant plus large : pilotage hors contrôle (et suivi des politiques publiques), contrôle (ciblage pour accompagnement ou lutte contre la fraude), information générale/statistiques publiques.

Taux de satisfaction des utilisateurs	Taux de satisfaction Mode de recueil de la satisfaction des utilisateurs
Usages actuels	Nombre d'usages par type d'usages

2.4 La qualité d'une base dépend de l'objectif poursuivi

[56] La qualité d'une base de données dépend de l'objectif poursuivi. Une même base peut être la meilleure pour un usage et la pire pour un autre. En effet, comme le dit Isabelle Boydens, Professeur à l'Université libre de Bruxelles et Data Quality Expert, " *par la qualité d'une base de données, on désigne sa relative adéquation aux objectifs qui lui sont assignés. La " qualité totale " n'existe pas, le concept étant relatif : à partir d'un arbitrage du type coûts/bénéfices, les critères de*

qualité les plus pertinents (fraîcheur de l'information, rapidité de la transmission des données, précision...) devront être retenus dans un contexte donné. ⁹⁵

[57] À chaque usage correspondent des dimensions plus importantes (parmi celles décrites précédemment voire d'autres pour certains usages) et un ordre de priorité entre ces dimensions, afin de sélectionner la base de données la plus adéquate à la situation.

3 Une tentative de cartographie des bases issues de la DSN

[58] Cette cartographie applique les 19 critères explicités dans la partie 2 au Bloc 3 et aux sept bases qui couvrent la plus grande partie du périmètre de la DSN et qui ont été présentées dans la partie 1. Elle a été construite sur la base des documents transmis à la mission et peut contenir des approximations et des manques du fait du peu de documents existants et des évolutions encore en cours.

3.1 Le « Bloc 3 » de la DSN native est très documenté

[59] Au sein de la DSN native, le Bloc 3 est alimenté par les déclarations arrivant au Bloc 1 et passant les contrôles bloquants. C'est la base qui alimente tous les producteurs des bases aval utilisées à des fins de pilotage et de statistiques publiques. Cependant le bloc lui-même n'est pas directement utilisé à ces fins. Par conséquent, la cartographie pour le Bloc 3 contient plusieurs rubriques « sans objet ».

Finalité	Assurer une gestion efficiente de la protection sociale
Périmètre couvert	DSN
Délai de mise à disposition	Sans objet
Opposabilité	Oui
Unités élémentaires	La déclaration
Principales informations expertisées	Conformité à la norme, identification du salarié, éléments du contrat de travail, information d'ordre économique et financier, événements survenus pendant la période déclarée
Principaux concepts construits	Aucun
Comparabilité temporelle	Les déclarations arrivent au fil de l'eau, il y a des déclarations initiales et rectificatives (en annule et remplace ou en complément)
Recul historique	Fichiers conservés 5 ans, pas d'archivage
Nomenclatures	PCS, IDCC, NAF, géographie
Interopérabilité	Nir
Traitements	Certification du Nir
Modalités d'accès à la base	sans objet

⁹⁵ <https://isabelle-boydens.web.ulb.be/annaes.pdf> L'océan des données et le canal des normes, in *Responsabilité et environnement*, juillet 2012

Dictionnaire des variables	Documents très détaillés de la norme NÉODeS sur www.net-entreprises.fr/declaration/norme-et-documentation-dsn/ : cahier technique (354 pages), différentiel avec la norme précédente (235 pages)
Description des traitements	Sans objet
Conseils d'usages	Sans objet
Coûts d'accès	Sans objet
Investissement pour utiliser les données	Sans objet (il serait probablement très important)
Usages actuels (pilotage, information)	Sans objet
Satisfaction des utilisateurs	Sans objet

3.2 Dad-e et PIQ : des bases mensuelles du Gip-MDS proches du Bloc 3

3.2.1 Dad-e, base très proche du Bloc 3

Finalité	Il n'existe pas d'analyse d'impact relative à la protection des données (AIPD) isolée pour Dad-e. Dad-e fait partie intégrante du traitement de la DSN native d'après une version provisoire de l'AIPD de la DSN ; la Cnav estime cependant que Dad-e ne devrait pas en faire partie. Dans ce document, la finalité générale est d'assurer une gestion efficiente de la protection sociale qui se décline en dix finalités dont la dernière est « <i>pour le suivi des politiques publiques, fiabilisation des données, permettant de recueillir au mois le mois des informations de qualité sur la vie et l'évolution des entreprises et de l'emploi</i> »
Périmètre couvert	DSN
Délai de mise à disposition	20 jours
Opposabilité	Oui
Unités élémentaires	Le contrat
Principales informations expertisées	Conformité à la norme, identification du salarié, éléments du contrat de travail, Information d'ordre économique et financier, événements survenus pendant la période déclarée
Principaux concepts construits	Données au contrat
Comparabilité temporelle	Fichiers mensuels
Recul historique	Fichiers conservés 5 ans, pas d'archivage
Nomenclatures	PCS, IDCC, NAF, géographie
Interopérabilité	Nir
Traitements	Changement implicite de contrats Vérification et exclusion des doublons. Suppression de la moitié des assiettes (de plus d'une vingtaine comme celle du forfait social à 8 % ou du régime spécial IEG).

Modalités d'accès à la base	Accès indirect uniquement. Processus informel. Demande à la DSS ou au Gip-MDS, dans l'expression des besoins annuelle vis-à-vis de la DSN (EB) ou en dehors. Instruction technique Gip et juridique DSS (conformité DCE / Arrêté Filtre). Délais incertains (entre 2 mois et plus d'un an)
Dictionnaire des variables	Renvoi au cahier technique de la norme
Description des traitements	Pas de document transmis à la mission
Conseils d'usages	Pas de document transmis à la mission, DAD-e est utilisée exclusivement par le Gip (ou via des Webservices pour les salariés)
Coûts d'accès	Usage interne
Investissement pour utiliser les données	Forte sauf pour les Webservices
Usages actuels (pilotage, information générale, statistiques publiques)	Transmission BDD Travailleurs handicapés : pilotage (Agefiph) Transmission BDD Contrats aidés, insertion par l'activité économique et entreprises adaptées : pilotage (ASP)
Satisfaction des utilisateurs	Pas de mesure organisée. Peu d'utilisateurs externes à des fins de pilotage ou d'information générale. Les demandeurs trouvent les délais d'instruction trop longs.

3.2.2 PIQ, base anonymisée et enrichie

Finalité	L'AIPD provisoire indique comme finalité celle de disposer d'un outil décisionnel permettant directement ou indirectement de couvrir les besoins d'analyse nécessaires dans le cadre des dispositifs DSN et PASRAU. Les trois 1 ^{ères} sous-finalités ont trait à la qualité et au suivi de montée en charge, la 4 ^e à la production d'indicateurs de pilotage.
Périmètre couvert	DSN
Délai de mise à disposition	30 jours
Opposabilité	Non
Unités élémentaires	Idem Dad-e
Principales informations expertisées	Idem Dad-e
Principaux concepts construits	Idem Dad-e
Comparabilité temporelle	Fichiers mensuels
Recul historique	Fichiers conservés 5 ans, pas d'archivage
Nomenclatures	Idem Dad-e
Interopérabilité	Données pseudonymisées
Traitements	Alimentation mensuelle par des données pseudonymisées issues de Dad-e, de PASRAU et du Répertoire Commun des Déclarants (RCD).
Modalités d'accès à la base	Accès indirect uniquement. Processus informel. Demande à la DSS ou au Gip-MDS, dans l'expression des besoins annuelle vis-à-vis de la DSN (EB) ou en dehors. Instruction technique Gip et juridique DSS (conformité DCE / Arrêté Filtre). Délais incertains (entre 2 mois et plus d'un an)
Dictionnaire des variables	Liste des variables dans l'AIPD provisoire. Renvoi au cahier des charges de la norme pour le dictionnaire.

Description des traitements	Décrits sommairement dans l'AIPD provisoire.
Conseils d'usages	Pas de document transmis à la mission, PIQ est à usage direct du seul Gip-MDS
Coûts d'accès	Payant
Investissement pour utiliser les données	Faible pour les accès indirects : indicateurs, tableaux de bords et outil de requêtage (mais investissement en amont pour définir le livrable).
Usages actuels (pilotage, information générale, statistiques publiques)	<ul style="list-style-type: none"> • Suivi de la montée en charge de la Prime exceptionnelle de pouvoir d'achat (Pepa) et de l'indemnité inflation (DSS) • Projet de suivi de l'accès à l'emploi de bénéficiaires du CEJ (DGEFP) • Projet de suivi du dispositif Bonus-Malus (DGEFP/DSS/Urssaf) • Changements de règles relatives aux cotisations : pilotage (Unédic) • Quelques réponses à des demandes ponctuelles
Satisfaction des utilisateurs	Pas de mesure organisée. Les entretiens n'ont pas fait remonter d'insatisfaction de la part des utilisateurs. Les demandeurs trouvent les délais d'instruction trop longs.

3.3 Tables tout en un (TTE1) de l'Urssaf CN, une base mensuelle expertisée et enrichie

Finalité	La finalité indiquée dans la fiche de registre relative à la base source (dite transitoire) est d'aider au pilotage stratégique. Parmi les quatre sous-finalités, permettre les analyses précoces sur données réelles pour engager un processus de fiabilisation continue des données et mener des analyses exploratoires sur les données à des fins de statistiques et d'analyses économiques
Périmètre couvert	DSN régime général (pas régime agricole)
Délai de mise à disposition	20 jours
Opposabilité	Non
Unités élémentaires	Contrat, individu, établissement et entreprise
Principales informations expertisées	Cotisations, assiettes, masses salariales, effectifs
Principaux concepts construits	<p>Effectifs réglementaires : effectifs moyens mensuels (EMM) et annuels (EMA) notifiés pour assujettissement aux cotisations et contributions, effectifs OETH, effectifs alternants</p> <p>Effectifs économiques : effectifs bénéficiant des exonérations</p> <p>Distribution de la masse salariale et des effectifs selon les taux horaires</p> <p>Prime et prime de partage de la valeur (montant et nombre de salariés)</p> <p>Heures supplémentaires</p>
Comparabilité temporelle	Fichiers mensuels
Recul historique	Fichiers conservés 8 ans depuis 2018
Nomenclatures	PCS, IDCC, NAF, géographie
Interopérabilité	Nir anonymisé

Traitements	Rapprochement des données agrégées et individuelles Prise en compte des régularisations
Modalités d'accès à la base	Accès indirect sauf extrait sur le CASD Procédure informelle
Dictionnaire des variables	Renvoi au cahier des charges de la norme NÉODeS
Description des traitements	Support de formation à destination des statisticiens du réseau des Urssaf
Conseils d'usages	Formation aux statisticiens du réseau des Urssaf
Coûts d'accès	Gratuit pour les tutelles, non-facturé dans le cadre de conventions relatives au recouvrement, lequel comporte des frais de gestion
Investissement pour utiliser les données	Important pour utiliser la base
Usages actuels (pilotage, information générale, statistiques publiques)	<ul style="list-style-type: none"> Fichiers annuels : pilotage / Assurance-chômage (Unédic) Fichiers annuels : pilotage / branche professionnelle (AGFPN) Fichiers mensuels : pilotage / formation (France Compétences) * Réponses à des demandes ponctuelles des tutelles et autres directions de l'Urssaf CN Usages des statistiques publiques (voir Annexe Statistiques publiques) Open data : information générale et pilotage sur internet. Des données sur le CASD pour évaluation.
Satisfaction des utilisateurs	Pas de mesure organisée. Les entretiens n'ont pas fait remonter d'insatisfaction de la part des utilisateurs.

3.4 DSNP3-Sismmo des bases mensuelles plus ou moins enrichies

3.4.1 DSNP3 de la Dnum, une base mensuelle très proche du Bloc 3

Finalité	L'AIPD non validée indique une finalité technique, à savoir mutualiser les opérations de collecte et de stockage nécessaires pour mettre à disposition les données aux directions métiers du ministère des affaires sociales
Périmètre couvert	DSN
Délai de mise à disposition	20 jours
Opposabilité	Oui
Unités élémentaires	La déclaration
Principales informations expertisées	Aucune
Principaux concepts construits	Aucun
Comparabilité temporelle	Fichiers mensuels
Recul historique	Fichiers conservés 10 ans
Nomenclatures	PCS, IDCC, NAF, géographie
Interopérabilité	Nir transcodé
Traitements	Uniquement des traitements sur la structuration des données

Modalités d'accès à la base	Accès indirect uniquement. DSNP3 a vocation à servir les directions métiers du ministère mais sert aussi le Gip inclusion Les modalités ne sont pas connues précisément et ne font pas l'objet d'un document
Dictionnaire des variables	Renvoi au cahier des charges de la norme, liste des variables transmises à la mission (près de 400 variables)
Description des traitements	Pas de document détaillé relatif aux traitements
Conseils d'usages	Pas de document relatif aux limites et aux conseils d'usage, base utilisée directement par la Dnum et la Dares uniquement
Coûts d'accès	Gratuit
Investissement pour utiliser les données	Très important, base utilisée directement par la Dnum et la Dares uniquement. Investissement en amont, dans l'instruction du livrable, dans les accès indirects.
Usages actuels (pilotage, information générale, statistiques publiques)	<ul style="list-style-type: none"> • Égapro : pilotage / égalité professionnelle (DGT) • Projet Dat'IA : contrôle / accidents du travail (DGT) • Projet Champollion : contrôle / contrats courts (DGT) • Projet OETH : pilotage / travailleurs handicapés (DGEFP, Dreets) • Base de Connaissance des Entreprises : contrôle / diverses dimensions (Dreets) • Ramsès : pilotage secteur médico-social (DGCS) • Expérimentation Pilotage / insertion allocataires du RSA (Gip Plateforme de l'inclusion) • Quelques réponses à des demandes ponctuelles
Satisfaction des utilisateurs	Pas de mesure organisée. Les entretiens n'ont pas fait remonter d'insatisfaction de la part des utilisateurs. Cependant un projet a été arrêté et l'expérimentation sur l'insertion d'allocataires du RSA n'apporte pas une valeur suffisante d'après le Gip Plateforme de l'inclusion.

3.4.2 Sismmo de la Dares, une base mensuelle proche du Bloc 3 chaînant les contrats

Finalité	L'AIPD distingue deux finalités : 1. Une finalité de statistique : réalisation de statistiques et d'études sur le domaine de l'emploi. 2. Une finalité de maîtrise d'œuvre : réalisation de traitements pour le compte des directions des ministères sociaux en qualité de sous-traitante de ces directions
Périmètre couvert	DSN
Délai de mise à disposition	1 mois
Opposabilité	Oui
Unités élémentaires	Contrat
Principales informations expertisées	Le résultat des expertises se trouve dans la base Post-Sismmo
Principaux concepts construits	Contrat
Comparabilité temporelle	Fichiers mensuels
Recul historique	Fichiers conservés depuis le début de la DSN, la cible est a priori une conservation de 8-10 ans en attendant une doctrine d'archivage.

Nomenclatures	PCS, IDCC, NAF, géographie
Interopérabilité	Nir
Traitements	Empilement par contrat avec programme de détection et de correction automatique des doublons de contrats (croisements entre plusieurs déclarations mensuelles successives)
Modalités d'accès à la base	Accès indirect principalement. Accès réservé aux directions métiers du ministère
Dictionnaire des variables	Renvoi sur le cahier des charges de la norme
Description des traitements	Description sommaire des traitements transmise à la mission
Conseils d'usages	Pas de document relatif aux limites et aux conseils d'usage, base utilisée directement par la Dares uniquement
Coûts d'accès	Gratuit pour DGEFP/DGT. Pour les projets longs, le demandeur contribue au développement directement (ex : InserJeunes) ou sous forme de financement (ex : Diamant).
Investissement pour utiliser les données	Important pour l'accès direct. Investissement en amont, dans l'instruction du livrable, dans les accès indirects.
Usages actuels (pilotage, information générale, statistiques publiques)	<ul style="list-style-type: none"> • Liste électorale TPE : pilotage (DGT) • InserJeunes : évaluation et information générale (Ministère de l'Education) • Intérim : statistiques (Prism'emploi) • Projet Diamant : pilotage / métiers médicaux-sociaux (ARS, DGOS) • SI précarité : contrôle / contrats précaires (DGT) • WikiT : contrôle / MMO (DGT) • Projet InserSup : information générale et évaluation (Min Ens Sup Recherche) • Expérimentation : simulateur bonus-malus (DGEFP) • Projet Champollion : contrôle / recours abusif aux contrats courts (DGT) • Réponses à des demandes ponctuelles
Satisfaction des utilisateurs	Pas de mesure organisée. Les entretiens ont fait remonter des appréciations positives de la part d'utilisateurs en termes de SI, connaissance du domaine emploi-formation, construction de concepts et du contenu qualitatif de la DSN. Il est aussi remonté que la Dares avait répondu négativement à des demandes pour des raisons techniques (robustesse insuffisante de la DSN) ou par manque de moyens suffisants (priorité à la continuité de la production).

3.4.3 Post-SISMMO, une base mensuelle expertisée et enrichie par la Dares

Finalité	Statistique publique
Périmètre couvert	DSN
Délai de mise à disposition	1 mois et 1 semaine
Opposabilité	Non
Unités élémentaires	Contrats, mouvements, établissements
Principales informations expertisées	Mouvements de main d'œuvre, intérim en attendant la fin du projet
Principaux concepts construits	A lister en fin de projet : en cible, emploi, qualité de l'emploi, indication sur le niveau de salaire individuel

Comparabilité temporelle	Fichiers mensuels
Recul historique	Fichiers conservés 8 à 10 ans
Nomenclatures	PCS, FAP, IDCC, NAF, géographie
Interopérabilité	code statistique non signifiant ?
Traitements	Détection d'anomalies, d'incohérence et redressements statistiques automatiques et par contrôles de gestionnaires (retour éventuel au déclarant) Redressement des informations sur les contrats, des heures rémunérées au contrat, des heures supplémentaires, des heures d'absence et des heures d'activité partielle Correction du mode d'exercice (temps plein ou partiel) Corrections des conventions collectives Définition d'un indicateur de salaire pertinent à un rythme mensuel, le salaire mensuel de base (SMB)
Modalités d'accès à la base	Sans objet, base en projet
Dictionnaire des variables	Pour la partie établissements MMO : sur www.casd.eu/generate-csv/?language=fr&id_src=176
Description des traitements	Base en projet, documents détaillés transmis à la mission sur MMO et intérim
Conseils d'usages	Base en projet
Coûts d'accès	Identiques à Sismmo pour les demandes hors CASD Tarif CASD (pour l'accès à l'ensemble des données et logiciels du CASD) : 3 500 € par an. Tarif affiché sur https://www.casd.eu/tarifs-2/
Investissement pour utiliser les données	A estimer à la fin du projet
Usages actuels (pilotage, information générale, statistiques publiques)	En projet (fin en 2023), destinée aux usages suivants branchés sur Sismmo : <ul style="list-style-type: none"> • Usages des statistiques publiques – cf. annexe Statistiques publiques • Évaluations via la base établissements sur les MMO au CASD (administrations et chercheurs) : 53 projets dont Analyse de l'impact des coûts de licenciement, Contrats de travail et pression sociale, Analyse du coût des ruptures de contrat de travail • Évaluations via la base appariée Force (Formation, Chômage et Emploi) : 34 projets dont évaluation du PIC, effets de la formation professionnelle sur le parcours sur le marché du travail des demandeurs d'emploi, évaluation d'impact de La Bonne Boîte... • Évaluations via l'appariement avec données CNAF et Pôle emploi au sein de la base Midas (Minima sociaux, droits d'assurance chômage et parcours salariés) • Open Data : statistiques, information générale et pilotage
Satisfaction des utilisateurs	Sans objet (encore en projet)

3.5 Base Tous Salariés de l'Insee, une base annuelle expertisée très riche

Finalité	Statistique (source = Dossier d'analyse de la conformité relative à la Protection des données à caractère personnel) : les bases « Tous salariés » renseignent sur le type d'emploi et de rémunérations perçues pour chaque salarié et permettent de produire des statistiques structurelles sur les salaires et l'emploi
Périmètre couvert	DSN + particuliers-employeurs
Délai de mise à disposition	Cible 12 mois (24 mois pour 2020, probablement 16 mois pour 2021)
Opposabilité	Non
Unités élémentaires	Contrat, poste, salarié, établissement
Principales informations expertisées	Salaires individuels (brut et net), durées de travail, ETP, salaires horaires, PCS, localisation de l'emploi, comparaison A/A-1
Principaux concepts construits	Contrat, poste, emploi, rémunération individuelle, durée du travail
Comparabilité temporelle	Fichiers annuels, comparaison A/A-1 embarquée dans le fichier de l'année A
Recul historique	Fichiers accessibles depuis 1993
Nomenclatures	PCS, NAF, IDCC, catégorie juridique, géographie
Interopérabilité	Dans le système d'information en amont de la base
Traitements	<ul style="list-style-type: none"> * validation des Siret * codage de la catégorie socioprofessionnelle (PCS) sur 4 positions selon la nomenclature PCS à partir des libellés en clair par codification automatique puis reprise par un gestionnaire * codage de la commune de résidence * localisation de l'emploi au lieu de travail * mise en cohérence interne à la DSN entre les variables relatives aux éléments de rémunérations une fois les 12 mois de déclarations agrégés. * calcul des salaires brut et net reconstruits à partir des différents éléments de rémunérations disponibles dans les déclarations sociales.
Modalités d'accès à la base	<p>Accès direct et indirect.</p> <p>Procédure comité du secret puis accès via le CASD pour des fins de recherche, d'études, d'évaluation. Pour les chercheurs, les administrations et les organismes qui opèrent dans le cadre d'une mission de service public.</p> <p>1 à 4 mois de délai selon nature de la demande et du demandeur.</p> <p>Procédure FPR pour les chercheurs via le comité Quetelet-Progedo (de quelques jours à 1 mois de délai ; le délai est affiché en ligne https://data.progedo.fr/fag) et pour les administrations et organismes qui opèrent dans le cadre d'une mission de service public via l'Insee.</p> <p>Procédure PSM (produits sur mesure = tableaux à façon) pour tous (1 à 2 mois de délai en général selon la nature de la demande et du demandeur)</p> <p>Accès en ligne immédiat à un fichier statistique gratuit et documenté pour les dimensions postes et salariés</p> <p>www.insee.fr/fr/statistiques/6524154</p>

	Procédures indiquées de façon complète en ligne : www.comite-du-secret.fr/ pour CASD et FPR, www.insee.fr/fr/information/1303438 pour PSM
Dictionnaire des variables	En ligne variables et modalités www.casd.eu/generate-csv/?language=fr&id_src=91 et dictionnaire détaillé www.insee.fr/fr/information/2407785
Description des traitements	En ligne de façon détaillée sur www.insee.fr/fr/metadonnees/source/operation/s2082/presentation
Conseils d'usages	Définitions, approches possibles pour le fichier statistique en ligne sur https://www.insee.fr/fr/statistiques/6524154#documentation
Coûts d'accès	Tarif CASD (pour l'accès à l'ensemble des données et logiciels du CASD) : à partir de 3 500 € par an. Tarif affiché sur https://www.casd.eu/tarifs-2/ Tarif FPR (Fichier production recherche) : gratuit Tarif PSM : gratuit pour les chercheurs ; pour les autres, prise en charge du coût de la constitution des tableaux. À partir de 214 €. Tarifs affichés sur www.legifrance.gouv.fr/loda/id/JORFTEXT000029337174
Investissement pour utiliser les données	Oui CASD, FPR et fichier statistique en ligne, non produits sur mesure (mais investissement en amont, dans l'instruction du PSM)
Usages actuels (pilotage, information générale, statistiques publiques)	<ul style="list-style-type: none"> • Usages des statistiques publiques – cf. annexe Statistiques publiques • Pilotage et évaluations par 10 entités hors recherche (France Stratégie, CAE, Cour des comptes, DG Trésor, DGE, IGF, RCO, Opco Akto, Unédic) : par exemple pour analyser les coûts du travail ou pour mettre à disposition des Carif-Oref un outil pour piloter les offres de formation et l'orientation des demandeurs d'emploi vers des métiers qui recrutent. • Évaluations par plus de 1 000 chercheurs via des centaines de projets de recherche en cours. Liste des projets en ligne https://www.casd.eu/source/base-tous-salaries-fichier-postes/. Par exemple, projets sur l'effet des chocs de productivité sur les réallocations d'emplois, sur les écarts de salaire entre parents et non-parents, sur une évaluation de l'incidence du CICE et de ses effets sur l'emploi, l'innovation et la compétitivité... • Pilotage par 90 administrations via tableaux à façon (Opco, Carif-Oref, observatoires, Dreets, Conseils régionaux...) • 18 chercheurs via FPR (universités, écoles françaises et étrangères) • Réponses à des demandes ponctuelles • Open data : statistiques, information générale et pilotage
Satisfaction des utilisateurs	Pas de mesure organisée. Entretiens : données de très grande qualité (champ, richesse, précision) à part les délais trop longs et encore plus longs pendant la montée en charge de la DSN. Idéal = cette base tous les mois et disposer en plus de variables brutes de la DSN.

[60] Flores qui est dérivée de BTS présente une cartographie similaire avec deux différences : des informations plus robustes pour des usages à un niveau géographique fin, un délai de mise à disposition allongé du fait de traitements supplémentaires, qui a 9 projets en cours au CASD.

[61] Les autres bases dérivées de BTS (Panel tous salariés et Panel tous salariés-Echantillon démographique permanent) couvrent un échantillon représentatif de salariés et permettent l'étude de trajectoires salariées sur longue période. La version enrichie par croisement avec l'échantillon démographique permanent inclut des informations supplémentaires sur les salariés (formation initiale, ménage du salarié...).

3.6 Des points à améliorer relatifs à l'accès aux bases de données

[62] Le caractère évolutif de la DSN nécessite une documentation millésimée. C'est le cas de la déclaration DSN et de sa norme NÉODES qui fait l'objet d'une documentation en ligne (cahier technique de 350 pages), très utilisée par tous les producteurs de bases en aval mais qui n'est pas simple à manier par les utilisateurs des bases aval : la documentation de la norme est très technique, elle n'indique pas quelles sont les variables présentes dans les autres bases ni les traitements éventuels à l'origine de ces bases.

3.6.1 Les informations sur les traitements des bases de données aval

[63] A part la base BTS, les traitements appliqués aux bases sont à ce stade peu documentés ou leur documentation n'est pas transmise aux utilisateurs ; ce qui nuit à leur compréhension par leurs utilisateurs et empêche des rapprochements entre les producteurs pour des traitements moins coûteux et plus cohérents.

[64] En particulier aucune base ne partage tout ou partie des codes utilisés pour la constituer. Le faire faciliterait le travail des utilisateurs qui souhaiteraient reconstituer certains concepts et accroîtrait la cohérence des résultats de leurs travaux, notamment les codes qui construisent des groupes (les contrats pérennes, les salariés à temps partiel...).

3.6.2 Les informations sur la qualité des données et leurs usages

[65] Les informations sur la qualité des données, sur leurs usages (conseils d'utilisation, bonnes et mauvaises pratiques) et leurs limites (indicateurs/indications sur la qualité des variables) sont réduites pour la plupart des bases. Par exemple, utiliser une donnée en temps réel n'est possible que si le bruit qu'elle contient (décalages temporels, incohérences...) n'empêche pas de voir le phénomène qu'elle représente.

[66] Augmenter les informations de ce type augmenterait les usages appropriés, diminuerait le risque de mésusage qui est grand avec cette source complexe, riche et évolutive, et ferait ainsi gagner du temps aux utilisateurs.

3.6.3 Les informations sur la procédure pour obtenir les données

[67] L'accès aux bases de données est très bien documenté pour les bases statistiques : le comité du secret est le point d'entrée, le CASD le point d'accès. La documentation se trouve en ligne.

[68] En revanche, l'accès est beaucoup moins clair pour l'accès aux autres bases de données. Le fait que leur objet principal n'est pas à des fins externes est un élément d'explication. Cependant cela nuit à l'utilisation des données de la DSN notamment à des fins de pilotage.

3.7 Une cartographie à affiner et faire connaître dans un cadre plus large

3.7.1 Une cartographie à affiner et à actualiser régulièrement

[69] La cartographie ayant été faite en un temps limité, elle est incomplète et contient des approximations. De plus certaines bases ne sont pas achevées et la DSN évolue chaque année. Des échanges complémentaires avec les producteurs des bases de données pourraient affiner la cartographie et permettre de la mettre à jour régulièrement en fonction des évolutions.

[70] De plus la cartographie bénéficierait d'échanges avec les utilisateurs pour la clarifier et la compléter.

[71] Les bases de données citées dans le paragraphe 1.3 ne sont pas cartographiées parce qu'elles ne couvrent qu'une partie de la DSN (par exemple les personnes ayant reçu récemment une formation financée). Il pourrait être intéressant d'étudier si leur ajout dans la cartographie apporterait un intérêt pour les usages dans les thématiques couvertes par ces trois bases.

3.7.2 Une cartographie à faire connaître dans un cadre plus large

[72] La cartographie porte sur les bases de données issues de la DSN. Or la DSN n'est pas la seule source de données pour répondre à des demandes de données à des fins de pilotage et de statistiques publiques dans le champ de l'emploi et de la formation professionnelle, de l'insertion et du suivi des trajectoires. Certaines autres sources (qui incluent néanmoins pour certaines des éléments de la DSN) sont ainsi privilégiées par nos interlocuteurs : DRM, DGFIP, Enquêtes Insee (enquête emploi, ERFS).

- Le dispositif de ressources mensuelles (DRM) de la DSS, alimenté par la DSN et Pasrau (Passage des revenus autres) regroupe les données de salaires et de revenus de remplacement ; ce qui élargit les usages au champ des revenus de remplacement ;
- Les données fiscales (DGFIP) : elles couvrent l'ensemble des salariés plus toutes les autres situations (les non-salariés, les personnes au chômage, en formation, à la retraite...). Elles permettent de suivre les trajectoires de personnes entre ces situations en intégrant à l'analyse les données présentes dans les déclarations fiscales (rémunérations, composition familiale) ;
- L'Enquête emploi (Insee) : cette enquête couvre un grand nombre de personnes âgées de plus de 15 ans, quelle que soit leur situation sur le marché du travail. C'est très utile pour suivre les trajectoires de personnes. Les différences avec les données fiscales sont que les données arrivent plus vite et contiennent beaucoup plus d'informations pour comprendre la situation professionnelle et personnelle des personnes, en revanche les revenus sont déclaratifs et les usages vont beaucoup moins loin au niveau géographique puisqu'il s'agit d'une enquête (échantillon de 80 000 personnes tous les trimestres) et non d'une base administrative exhaustive.

- L'Enquête revenus fiscaux et sociaux (Insee) consiste en une fusion, sur l'échantillon de l'enquête emploi, de données fiscales, de données de l'enquête emploi et de données relatives aux prestations sociales. Elle conduit à des analyses très riches mais à un niveau géographique très agrégé.

[73] Ainsi, pour une bonne orientation des usages, le panorama ne peut pas se limiter aux sources cartographiées. Il est préférable de faire une instruction thématique plutôt qu'une instruction par source de données et de la faire faire par des personnes qui ont la compétence thématique en plus de la compétence « DSN ».

ANNEXE 2 : L'utilisation des données de la DSN à des fins de statistiques publiques

ANNEXE 2 : L'UTILISATION DES DONNEES DE LA DSN A DES FINS DE STATISTIQUES PUBLIQUES	88
1 LA DSN A ENGENDRE UNE REFONTE D'AMPLEUR DES SYSTEMES D'INFORMATION DU SERVICE STATISTIQUE PUBLIC, VISANT A ASSURER DAVANTAGE DE COHERENCE ET A RACCOURCIR LES DELAIS	91
1.1 LA DSN S'EST SUBSTITUEE A QUATRE DECLARATIONS ADMINISTRATIVES UTILISEES PRECEDEMMENT PAR LES DIFFERENTS ACTEURS DU SERVICE STATISTIQUE PUBLIC	91
1.2 DE NOMBREUX TRAVAUX D'EXPERTISE ONT ETE MENES PAR LE SSP ELARGI POUR QUALIFIER LES DONNEES AU FIL DES ETAPES SUCCESSIVES DE DEPLOIEMENT DE LA DSN	92
1.3 DES SYSTEMES D'INFORMATION STATISTIQUES QUI RESTENT DISTINCTS, POUR REpondre A DIFFERENTS OBJECTIFS	93
1.3.1 Le système d'information de l'Insee (Siera).....	93
1.3.2 Le système d'information de la Dares (Sismmo).....	95
1.3.3 Le système d'information de l'Urssaf Caisse nationale.....	96
1.4 LA DSN A ETE UN LEVIER D'AMELIORATION DE LA COORDINATION ENTRE LES ACTEURS DU SERVICE STATISTIQUE PUBLIC ELARGI A L'URSSAF CAISSE NATIONALE	97
1.5 L'EMERGENCE DE NOUVEAUX SYSTEMES D'INFORMATION IMPLIQUE DE CLARIFIER LE ROLE ET LES RELATIONS ENTRE LES ACTEURS	98
1.6 DIFFERENTS LIVRABLES STATISTIQUES SONT AUJOURD'HUI PRODUITS A PARTIR DE LA DSN DANS DES DELAIS QUI SE RACCOURCISSENT	99
2 LES ACTEURS DE LA STATISTIQUE PUBLIQUE REALISENT DE NOMBREUSES EXPERTISES ET RETRAIENT LES DONNEES BRUTES ISSUES DE LA DECLARATION DSN	106
2.1 TRAITEMENTS POUR CORRIGER ET DOCUMENTER LES ERREURS DECLARATIVES, AVEC DES CONTROLES DE COHERENCE INTERNE OU EXTERNE.....	106
2.2 TRAITEMENTS POUR PASSER DE LA DATE DE LA DECLARATION A LA DATE DE REFERENCE AFIN DE GERER LES RETARDATAIRES ET LES DECLARATIONS RECTIFICATIVES.....	107
2.3 TRAITEMENTS CORRECTIFS DES DOUBLONS LIES A DES DEFAUTS D'IDENTIFIANT DES CONTRATS.....	107
2.4 TRAITEMENTS ADRESSANT LES INCOHERENCES ENTRE LES DONNEES AGREGES (BLOC 1) ET LES DONNEES INDIVIDUELLES (BLOC 3)	108
2.5 DES TRAITEMENTS POUR CONSTRUIRE DES CONCEPTS ET UNITES D'INTERET ET SE CONFORMER A DES NOMENCLATURES DE REFERENCE	110
2.6 DES TRAITEMENTS POUR CONTROLER LA LOCALISATION AFIN DE FOURNIR DES INDICATEURS A DES NIVEAUX GEOGRAPHIQUES FINS.....	111
2.7 UNE BOUCLE DE RETROACTION AUPRES DU GIP-MDS POUR FIABILISER EN AMONT LA DECLARATION .	111
3 LA DSN A DEMULTIPLIE LES POSSIBILITES D'APPARIEMENTS POUR ENRICHIR LES ANALYSES ET ASSURER UN SUIVI DES DYNAMIQUES D'INSERTION PROFESSIONNELLE	111

3.1	LE PANEL FORCE, POUR SUIVRE LES TRAJECTOIRES DES PERSONNES AYANT EU UN CONTACT AVEC LE SERVICE PUBLIC DE L'EMPLOI	112
3.2	LE PANEL TRAJAM, POUR SUIVRE LES TRAJECTOIRES DES JEUNES PASSES PAR UNE MESURE DE POLITIQUE D'EMPLOI	113
3.3	LE PANEL MIDAS, POUR SUIVRE LES TRAJECTOIRES DES INSCRITS A POLE EMPLOI ET DES BENEFICIAIRES DE MINIMA SOCIAUX	113
3.4	LE DISPOSITIF INSERJEUNES, POUR SUIVRE L'INSERTION PROFESSIONNELLE DES SORTANTS DES FILIERES PROFESSIONNELLES	113
3.5	LE DISPOSITIF INSER SUP, POUR SUIVRE L'INSERTION PROFESSIONNELLE DES SORTANTS DE L'ENSEIGNEMENT SUPERIEUR	114
3.6	AUTRES PROJETS	114
4	LE SERVICE STATISTIQUE PUBLIC MET A DISPOSITION DE TRES NOMBREUX UTILISATEURS LES BASES DE DONNEES STATISTIQUES ISSUES DE LA DSN DANS LE CADRE SECURISE DU CASD	115
4.1	LES BASES DE DONNEES STATISTIQUES ISSUES DE LA DSN SONT MISES A DISPOSITION PAR LES DIFFERENTS ORGANISMES PRODUCTEURS	115
4.2	LE CASD OFFRE A PRIX COUTANT UN ENVIRONNEMENT D'ACCES AUX DONNEES ET UNE INFRASTRUCTURE DE CALCUL, AVEC UN DISPOSITIF DE SECURISATION VIA UN BOITIER.....	116
4.3	UN ACCES IMPORTANT ET CROISSANT AUX BASES DE DONNEES DU CASD SOUS LE COUVERT D'UNE PROCEDURE D'HABILITATION	116
4.3.1	... assurée par le Comité du secret	116
4.3.2	... ou assurée directement par le producteur	117
4.4	DE NOMBREUX PROJETS D'ETUDES MOBILISENT LES SOURCES ISSUES DE LA DSN	118

[1] Cette annexe établit un état des lieux des **systèmes d'information et des principaux produits de la statistique publique** en matière d'emploi et de rémunérations salariales issus de la DSN. Le périmètre retenu est celui des statistiques produites à des fins d'information générale, en se centrant sur les produits et statistiques diffusés avec une périodicité régulière. Il exclut les productions à des fins de pilotage traitées dans l'annexe n°3.

[2] **Les statistiques publiques ont été définies dans le cadre de la loi fondatrice « sur l'obligation, la coordination et le secret en matière de statistiques »** (loi française n°51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques⁹⁶, actualisée par la loi n°2016-1321 du 7 octobre 2016 pour une République numérique). Elles proviennent soit d'enquêtes statistiques au sens de la loi de 1951, labellisées par le Conseil national de l'information statistique (Cnis) et inscrites au programme visé par le ministre en charge de l'Economie et des finances, soit de collectes administratives ou de l'utilisation à des fins statistiques de sources administratives. Pour des raisons d'allègement de charge sur les répondants et de réduction des coûts, **le recours à l'exploitation des bases de données administratives est privilégié dès que cela est possible**, les enquêtes étant conçues de façon complémentaire pour éclairer les thématiques ne pouvant être approchées à partir de sources administratives. Les statistiques publiques couvrent ainsi **« l'exploitation, à des fins d'information générale, de données collectées par des administrations, des organismes publics ou des organismes privés chargés d'une mission de service public »**.

[3] **La production des statistiques publiques donne systématiquement lieu à diffusion de résultats.** À la différence de certains travaux d'étude ou d'évaluation, qui peuvent garder un certain degré de confidentialité, les statistiques publiques doivent faire l'objet d'une diffusion. Cette diffusion est aujourd'hui assurée notamment via Internet. Pour autant, elle doit notamment respecter **les règles du secret statistique**.

[4] **La production de statistiques publiques, à des fins d'information générale, relève principalement du Service Statistique Public (SSP), mais pas exclusivement.** Elle relève aussi d'autres acteurs qui ont des missions de service public comme, par exemple, les caisses de sécurité sociale ou d'autres établissements publics, au titre de leur activité de publication régulière de statistiques à partir de leurs propres exploitations des sources administratives dont ils ont la responsabilité. En matière d'emploi et de rémunérations, les principaux organismes engagés dans la production de telles statistiques publiques sont l'Urssaf Caisse nationale (ex-Acoss) et la Caisse centrale de Mutualité sociale agricole (CCMSA), qui produisent et publient sur leur champ respectif des séries d'effectifs, de salaires ou de masses salariales. Plus récemment, en matière de formation professionnelle, la Caisse des dépôts et consignations (CDC) est devenue productrice de statistiques publiques, du fait de ses compétences sur le suivi du Compte personnel de formation et de son dispositif Agora, plateforme numérique centralisant les données de la formation professionnelle, qui utilise la DSN à la fois pour alimenter les comptes de formation individuels, mais aussi pour en suivre l'impact en termes d'emploi. Depuis 2011, **l'Autorité de la statistique publique (ASP)** a souhaité encadrer cette production de statistiques publiques produites hors du SSP, en les soumettant à des règles d'assurance qualité inspirées du **Code des bonnes pratiques de la statistique européenne**⁹⁷. L'ASP a ainsi labellisé plusieurs séries statistiques

⁹⁶ <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573>

⁹⁷ <https://www.autorite-statistique-publique.fr/code-des-bonnes-pratiques-de-la-statistique-europeenne/>

produites par l'Urssaf caisse nationale et la CCMSA, celles-ci s'inscrivant en complémentarité de celles produites par le SSP. Des démarches ont été entamées par l'Autorité avec la CDC dans un objectif de cohérence d'ensemble avec les statistiques publiées par la Dares, sans aboutir à ce stade compte tenu d'autres priorités⁹⁸.

[5] Inversement, toute la production de l'Insee ou des SSM ne relève pas de la sphère des **statistiques publiques**. **Les SSM jouent un rôle d'aide à la décision dans leur domaine de compétence et réalisent des travaux d'aide à la décision et au pilotage, qu'il s'agisse par exemple d'analyser l'impact de mesures, ou de réaliser des travaux de prospective**. La **Charte des services statistiques ministériels**⁹⁹ pose ainsi que les SSM participent à « ... la production de notes d'analyses, d'indicateurs statistiques et de tableaux de bord à destination des directions d'administration centrale et des services locaux des administrations concernées, ainsi que du cabinet du ministre. En effet, grâce aux compétences de leurs agents dans le domaine de la statistique et de l'économie et à leurs bonnes connaissances des domaines propres aux ministères, liées à leur implantation dans les administrations centrales de ceux-ci, les SSM peuvent réaliser des travaux destinés à fournir des indicateurs et des expertises utiles pour l'aide au pilotage des politiques publiques ministérielles. Les informations produites dans ce cadre ne font pas systématiquement l'objet d'une diffusion publique. »

1 La DSN a engendré une refonte d'ampleur des systèmes d'information du service statistique public, visant à assurer davantage de cohérence et à raccourcir les délais

[6] Les **statistiques publiques d'emploi et de rémunérations salariales** constituent un enjeu majeur pour les statisticiens publics, du fait de leur importance dans le diagnostic économique de court ou de moyen terme. Elles permettent d'alimenter les **comptes nationaux** et de répondre aux demandes de **production d'indicateurs harmonisés aux niveaux européen et international**. Elles constituent le socle des indicateurs et des **prévisions conjoncturelles**. Elles alimentent les indicateurs relatifs au **système productif et à la statistique d'entreprises, au niveau national et au niveau localisé**. A un niveau microéconomique, il s'agit également de répondre aux **enjeux plus sociaux relatifs au fonctionnement du marché du travail, de connaître plus finement la diversité des formes d'emploi, les inégalités salariales et leurs déterminants**.

1.1 La DSN s'est substituée à quatre déclarations administratives utilisées précédemment par les différents acteurs du service statistique public

[7] L'usage des données et déclarations administratives pour produire des statistiques publiques sur l'emploi et les salaires est ancien. La DSN a ainsi remplacé quatre sources administratives majeures qui étaient mobilisées antérieurement :

⁹⁸ Plus globalement, il s'agit aussi d'organiser les rôles respectifs de la Dares et de la CDC concernant les usages statistiques du système d'information sur la formation professionnelle Agora.

⁹⁹ https://intranet.insee.fr/jcms/68195_DBFileDocument/charte-des-services-statistiques-ministeriels

- les **déclarations et enquêtes sur les mouvements de main-d'œuvre** (DMMO/EMMO) étaient historiquement collectées par la Dares. Elles permettaient de suivre les flux de main-d'œuvre (entrées et sorties de salariés) chaque trimestre ;
- les **relevés mensuels de missions d'intérim** (RMM) étaient collectés par Pôle Emploi et mobilisés par Pôle Emploi et la Dares pour établir des estimations de l'emploi intérimaire venant alimenter les statistiques d'emploi trimestrielles coproduites par l'Insee et la Dares ;
- les **bordereaux récapitulatifs de cotisations** (BRC) traités tant par l'Urssaf que par l'Insee alimentaient les séries trimestrielles d'effectifs salariés produites par l'Urssaf et les estimations trimestrielles d'emploi établies par l'Insee et la Dares. Ils permettaient aussi de suivre l'évolution trimestrielle de la masse salariale et des exonérations de cotisations sociales (Urssaf) et alimentaient la production des indices de coût du travail (Insee). La DSN a été l'occasion d'intégrer les travaux des trois producteurs (Dares, Insee, Urssaf) pour produire conjointement les estimations trimestrielles d'emploi salarié sur l'ensemble du champ à compter de 2017 ;
- les **déclarations annuelles de données sociales** (DADS) traitées par l'Insee constituaient le pilier de l'information structurelle sur les salaires depuis les années 1950. L'utilisation d'imprimés fiscaux à l'origine, puis des déclarations annuelles sociales (DAS) à partir de 1970, remplacées par les déclarations annuelles de données sociales (DADS) en 1983, puis par les déclarations sociales nominatives (DSN) à partir de 2015, a été privilégiée pour connaître précisément les rémunérations salariales. Depuis 2007, les déclarations sociales relatives aux salariés et aux non-salariés sont devenues le socle pour établir le niveau d'emploi en France¹⁰⁰, aussi bien au niveau national qu'au niveau localisé (le concept d'emploi diffère du concept de poste occupé : un traitement de la multi-activité est nécessaire au niveau individuel pour ne compter un individu en emploi à une date donnée qu'une seule fois dans le cas où il exerce plusieurs activités, le cas échéant sous différents statuts).

1.2 De nombreux travaux d'expertise ont été menés par le SSP élargi pour qualifier les données au fil des étapes successives de déploiement de la DSN

[8] **La DSN a constitué un choc déclaratif majeur impactant l'ensemble des producteurs de statistiques publiques et a conduit à une rénovation profonde de leurs systèmes d'information.** Le déploiement de la DSN s'est d'abord traduit par une phase d'apprentissage de la nouvelle déclaration qu'il a fallu gérer au fil de ses évolutions, avec un enjeu premier de continuité des séries répondant à la nécessité de produire des évolutions interprétables sur les concepts économiques d'intérêt. Chacun des acteurs s'est inscrit dans une logique conservatoire, de prolongation du « patrimoine statistique » existant, répondant à ses besoins stratégiques : il n'était en effet pas envisageable de prendre des risques trop importants en termes de livrables.

¹⁰⁰ Cela a fait suite à l'abandon en 1999 du recensement exhaustif de la population, remplacé par le recensement en continu (les données du nouveau recensement établies en moyenne sur cinq ans ne permettant plus de fournir un niveau d'emploi annuel à une date de référence précise).

Toutefois, en unifiant le processus déclaratif, la DSN a aussi apporté une opportunité de rapprochement des différents processus statistiques conjoncturels et structurels, avec en perspective une meilleure cohérence des livrables, ainsi que des mises à disposition dans des délais raccourcis. La déclaration unifiée constitue ainsi un gage de simplification à terme notamment pour l'Insee, avec une unification des sources déclaratives concernant l'emploi salarié. Le déploiement de la DSN entre 2017 et 2022 a conduit à gérer deux enjeux principaux.

[9] **Un premier enjeu relevait de l'adaptation des activités statistiques à l'évolution du périmètre d'application de la DSN : son déploiement s'est fait par étapes successives, avec différents jalons de montée en charge.** Le millésime 2022 constitue une étape majeure avec l'intégration de l'ensemble de la fonction publique, avant la dernière étape visant à incorporer les salariés des particuliers-employeurs prévue d'ici 2024-2025, avec un dispositif de type tierce déclaration (pour mémoire, à date, les déclarations de particuliers-employeurs remontent dans des dispositifs distincts de type TESE, TESA, CESU, et leur déclaration peut être différée jusqu'à 6 mois après la fin du trimestre de référence).

[10] **Le second enjeu majeur porte sur les questions de qualité des déclarations DSN pour leur usage statistique.** Cet enjeu de qualité statistique fait écho aux enjeux de fiabilisation des déclarations pour leur usage premier qui est de permettre aux métiers d'exercer leurs missions de gestion du recouvrement et des droits sociaux des salariés. Ces enjeux de fiabilisation ont été répertoriés dans le rapport du HCFIPS de juillet 2022 et dans le rapport du l'Igas sur la gouvernance du GIP-MDS. Il est important de souligner que l'indicateur de conformité à la norme de la déclaration mis en avant par le GIP-MDS (de l'ordre de 94 %) ne signifie pas que les données sont exactes : cet indicateur mesure le taux d'acceptation des DSN déposées au regard des contrôles de la norme NÉODES. Cette conformité est analysée au regard des contrôles embarqués dans le cahier technique, validant la bonne structuration des données et leur cohérence au sein de la déclaration. Mais la mission s'est vue fournir de nombreux exemples d'inexactitudes déclaratives des employeurs qui ne peuvent pas être identifiées par des contrôles internes à la déclaration DSN, et nécessitent donc des contrôles de cohérence externe avec d'autres sources d'information (administratives ou enquêtes).

1.3 Des systèmes d'information statistiques qui restent distincts, pour répondre à différents objectifs

[11] A l'occasion de la DSN, la coopération a été accrue notablement entre les acteurs du service statistique public dans le cadre de la coproduction des Estimations trimestrielles d'emploi (cf. § 1.4). Cependant, les systèmes d'information n'ont pas été mutualisés, les environnements de production et les objectifs visés étant propres à chaque organisation. Trois systèmes d'information continuent ainsi de coexister, suivant des finalités différentes.

1.3.1 Le système d'information de l'Insee (Siera)

[12] Le système d'information sur l'emploi et les revenus d'activité (Siera) est piloté et mis en œuvre par le Département de l'emploi et des revenus d'activité (Dera) de l'Insee. Aux côtés de l'enquête Emploi, le Siera constitue le pilier pour répondre aux besoins de connaissance sur l'emploi et les revenus d'activité.

[13] Le **Siera se compose d'une part de processus conjoncturels**, dont l'objectif principal est de produire chaque trimestre des Estimations trimestrielles d'emploi salarié, en coproduction avec la Dares et l'Urssaf caisse nationale. L'Insee diffuse également chaque trimestre l'indice du coût du travail (ICT), « salaires seuls » et « salaires et charges ».

[14] Le **Siera se compose d'autre part d'un processus annuel**, visant à établir des statistiques structurelles permettant la production des indicateurs de référence sur les évolutions salariales et sur les caractéristiques fines des emplois occupés. Les principaux produits du processus structurel sont le **fichier annuel portant sur l'ensemble des postes de travail et des salariés (base Tous salariés - BTS)** et le Fichier localisé sur les rémunérations et l'emploi salarié (Flores), qui décrit l'emploi salarié et les rémunérations au niveau des établissements.

[15] Le **fichier pivot, la BTS, sera constitué à terme des données de la DSN à l'issue de sa généralisation complète**. Auparavant, cette base était construite par synthèse de différentes déclarations et sources : les DADS, les fichiers de paye de la fonction publique et les déclarations des particuliers-employeurs. Les principaux traitements effectués dans BTS concernent :

- l'identification au répertoire des établissements (dans sa version statistique Sirus) ;
- le codage des variables de profession et catégorie sociale (PCS) et de commune (au lieu de résidence et au lieu de travail), par codification automatique et reprise par un gestionnaire des échecs de codification automatique ;
- l'analyse et la correction des incohérences internes aux DSN d'une même année, une fois les 12 mois de déclarations agrégés ;
- la définition du poste¹⁰¹ de travail comme unité d'analyse, chaque poste correspondant au cumul des périodes d'emploi contiguës d'un même salarié au sein d'un même établissement.
- pour le calcul des agrégats annuels sur les salaires, chacun des postes est pris en considération au prorata de son volume, qui doit être calculé à partir des rubriques de la DSN, en heures ou en équivalent temps plein ;
- les salaires, bruts et nets, doivent aussi être recalculés à partir de la DSN. à partir de laquelle celles-ci sont devenues imposables et ont donc été intégrées dans le salaire net fiscal). Le **salaire brut** correspond à l'intégralité des sommes perçues par le salarié avant déduction des cotisations salariales de sécurité sociale, de la contribution sociale généralisée (CSG) et de la contribution au remboursement de la dette sociale (CRDS). Il couvre les montants d'épargne salariale et les cotisations patronales pour complémentaire santé obligatoire. Le **salaire net** (de prélèvements sociaux) est le salaire que perçoit effectivement le salarié (en espèces ou en nature). Il est net de toutes cotisations sociales, de CSG et de CRDS mais pas du prélèvement à la source de l'impôt sur le revenu. Le salaire net couvre, en tant qu'avantage en nature, les cotisations pour complémentaire santé obligatoire payées par l'employeur, et ce depuis 2013 (année à partir de laquelle celles-ci sont devenues imposables et ont donc été intégrées dans le salaire net fiscal). Le salaire net ne couvre pas

¹⁰¹ Un poste correspond au cumul des périodes d'emploi d'un salarié dans un même établissement. Les variables quantitatives (salaire, durée de travail, nombre d'heures de travail, etc.) de chacune des périodes sont additionnées, alors que pour les variables qualitatives, ce sont celles de la période principale qui sont retenues. La période principale est la période qui possède la rémunération la plus élevée.

les montants d'épargne salariale placés sur un plan d'épargne d'entreprise qui ne sont pas soumis à l'impôt sur le revenu.

[16] Les indicateurs sur les rémunérations salariales sont établis sur un cycle annuel, après agrégation des 12 déclarations mensuelles d'une année civile, afin de lisser sur l'ensemble de l'année les éléments variables de rémunération (primes, 13^{ème} mois, etc.) ou les irrégularités déclaratives (corrections en fin d'année civile). Une approche sur 12 mois glissants aurait été trop lourde à implémenter. Les indicateurs infra-annuels sont quant à eux calculés soit sur les salaires de base, soit sur les masses salariales, rapportés par tête ou selon le volume d'emploi et corrigées des variations saisonnières).

1.3.2 Le système d'information de la Dares (Sismmo)

[17] Le dispositif Sismmo mis en place par la sous-direction Emploi de la Dares vise à produire des indicateurs sur les mouvements de main d'œuvre et sur l'intérim et à offrir une vision fine des contrats de travail.

[18] La base **Sismmo-intérim** a été constituée dès le début de la DSN pour répondre aux contraintes de délais de publication sur l'intérim (échéance mensuelle et trimestrielle). La Dares mobilise Sismmo-intérim dans le cadre de la coproduction des estimations trimestrielles d'emploi, en partenariat avec l'Insee et la Caisse nationale des Urssaf.

[19] La base **MMO** produit les données sur l'ensemble des embauches et des fins de contrats de travail au niveau des établissements. Elle permet de mesurer les entrées et les sorties selon le type de contrat (contrat à durée déterminée/contrat à durée indéterminée), ou selon certaines caractéristiques individuelles des salariés (catégorie socio-professionnelle, âge...) ainsi que la durée des contrats et les motifs de rupture (fin de contrat, démission, licenciement...). La DSN a constitué un progrès : elle a permis de compléter le champ des MMO couvert par les anciennes déclarations au sein du secteur privé, et pourra conduire à élargir les analyses à la fonction publique à partir de 2022.

[20] En plus des finalités intérim et MMO, la Dares constitue une **base de données mensuelle sur l'ensemble des contrats en fin de mois, en intégrant des corrections en cas d'incohérence sur plusieurs DSN mensuelles successives**. La granularité retenue (contrat) est ainsi plus fine que celle retenue par l'Insee (période et poste de travail). La Dares envisage de construire deux versions de ce SI, une version dite Sismmo à vocation administrative, sans redressements statistiques et une **version Post-Sismmo**, plus redressée, sur laquelle des imputations et redressements statistiques sont apportés. Sismmo vise à répondre aux besoins des directions d'administrations centrales, qui souhaitent accéder aux déclarations brutes/opposables des entreprises. Post-Sismmo s'éloigne des déclarations d'origine et vise notamment à répondre à des besoins statistiques infra-annuels de la Dares et de ses partenaires, notamment pour la réalisation d'appariements. Post-Sismmo intègre différents types de traitements statistiques : redressement et harmonisation des heures rémunérées au contrat, des heures supplémentaires, des heures d'absence et des heures d'activité partielle, en lien avec les travaux réalisés par l'Insee sur la BTS ; correction du mode d'exercice (temps plein ou temps partiel) à l'aide des heures rémunérées ; corrections des conventions collectives déclarées ; définition d'un indicateur de salaire pertinent à un rythme mensuel, le salaire mensuel de base (SMB).

1.3.3 Le système d'information de l'Urssaf Caisse nationale

[21] Au sein de l'Urssaf Caisse nationale, la direction des statistiques, des études et des prévisions (Disep) est responsable de la production des statistiques d'effectifs salariés et de masse salariale dans le secteur privé. L'Urssaf Caisse nationale ne fait pas partie du SSP, mais est « rattachée » à la sphère statistique publique : une partie de ses productions a en effet été labellisée par l'Autorité de la statistique publique¹⁰² et depuis 2017, elle participe à la coproduction des Estimations trimestrielles d'emploi.

[22] Le système d'information de l'Urssaf Caisse nationale comprend plusieurs chaînes statistiques qui répondent à des besoins aussi bien statistiques que métier, permettant :

- le calcul des effectifs réglementaires (effectif moyen mensuel, effectif moyen annuel, etc.¹⁰³) ;
- le ciblage des contrôles comptables d'assiette ;
- les travaux de fiabilisation (rapprochement données individuelles et agrégées) ;
- le suivi de mesures ou dispositifs pour la tutelle (chômage partiel, primes exceptionnelles) ;
- le suivi des exonérations ;
- des travaux de simulations pour le compte de la caisse nationale ou de la tutelle.

¹⁰² Lors de sa séance du 29 septembre 2011, l'Autorité de la statistique publique a notifié la labellisation des séries trimestrielles d'emploi produites au niveau national par l'Urssaf Caisse nationale. En 2013, la labellisation a été notifiée pour les séries de masse salariale et de Déclarations Préalables à l'Embauche (DPAE), toujours au niveau national. L'Urssaf Caisse nationale a depuis 2015 fait évoluer la dénomination de ces séries en séries trimestrielles d'effectifs salariés afin de la différencier de la production simultanée des séries d'emploi par l'Insee. En 2016, un renouvellement de labellisation a été obtenu pour ces séries, suivi en 2020 d'une labellisation étendue à des niveaux géographiques infra-nationaux. En 2021, l'Urssaf Caisse nationale a présenté de nouvelles séries dans le cadre de la démarche de reconnaissance par l'ASP de « statistiques publiques à visée d'information générale ».

¹⁰³ Cf. <https://boss.gouv.fr/portail/accueil/regles-dassujettissement/effectif.html>

L'effectif moyen annuel (EMA), est dorénavant calculé par l'Urssaf Caisse nationale à partir des informations contenues dans les DSN déposées par les établissements de chaque entreprise. Les effectifs moyens mensuels et annuels sont mis à disposition des déclarants sur le compte Urssaf en ligne.

Depuis 2021, dans le cadre du transfert de la déclaration obligatoire d'emploi des travailleurs handicapés (DOETH) au réseau des Urssaf, 3 effectifs sont calculés par l'Urssaf : l'EMA OETH correspond à l'EMA de l'obligation d'emploi des travailleurs handicapés ; l'EMA BOETH correspond à l'EMA des bénéficiaires de l'obligation d'emplois des travailleurs handicapés ; l'EMA ECAP correspond à l'EMA des emplois exigeant des conditions d'aptitudes particulières.

À compter de 2023, dans le cadre du transfert de la contribution supplémentaire à l'apprentissage (CSA) au réseau des Urssaf, 2 nouveaux effectifs seront calculés par l'Urssaf : l'EMA CSA correspond à l'effectif moyen annuel spécifique participant à la détermination du taux d'alternants (dénominateur de ce taux) et de la redevabilité à la contribution supplémentaire à l'apprentissage (CSA) ; l'EMA CFIP correspond à l'effectif moyen annuel des contrats favorisant l'insertion professionnelle (CFIP) participant à la détermination du taux d'alternants (numérateur de ce taux) et de la redevabilité à la contribution supplémentaire à l'apprentissage (CSA). Ces deux effectifs permettent de calculer le ratio d'alternants (EMA CFIP/EMA CSA) de l'entreprise qui détermine le taux à appliquer à l'assiette pour le calcul du montant de la CSA.

[23] Actuellement, pour ses travaux statistiques et métier, l'Urssaf Caisse nationale met en œuvre deux systèmes complémentaires : d'une part une « chaîne historique » : pour le calcul des effectifs en fin de mois et les publications ; d'autre part des tables « tout en un » : au niveau trimestriel, plus faciles à exploiter par les statisticiens. Ces chaînes sont en cours de refonte, avec une intégration au système d'information décisionnel de l'Urssaf dans une technologie « *big data* », avec la possibilité de disposer de données aux deux échéances déclaratives au 5 et au 15 du mois pour des usages de pilotage (l'extraction au 5 du mois permet de disposer d'indicateurs avancés sur le champ des établissements de 50 salariés ou plus), les extractions restant mensuelles pour les usages statistiques.

1.4 La DSN a été un levier d'amélioration de la coordination entre les acteurs du service statistique public élargi à l'Urssaf Caisse nationale

[24] La DSN, en tant que déclaration unifiée remplaçant des déclarations distinctes, a été l'occasion de renforcer la coopération et de procéder à une rationalisation majeure entre les différents producteurs historiques de statistiques publiques.

[25] **L'Insee et l'Urssaf Caisse nationale ont convenu en date du 13 mars 2017 d'un accord de partenariat pour assurer la coproduction (avec la Dares) des estimations trimestrielles d'emploi (ETE) salarié.** Ce rapprochement, concrétisé à l'occasion de la mise en place de la DSN, visait à mettre fin à la situation qui existait auparavant, à savoir une production « en doublon » du suivi des évolutions trimestrielles d'emploi sur le secteur privé, réalisé distinctement par le SSP (Insee et Dares) d'une part et l'Acos (aujourd'hui Urssaf Caisse nationale) d'autre part. L'existence de ces doubles traitements à partir de la source des BRC était source d'inefficience collective et de difficultés de communication voire de remise en cause des indicateurs produits, et ce d'autant plus que les séries d'emploi de l'Acos avaient fait l'objet d'une labellisation par l'Autorité de la statistique publique, reconnaissant leur pertinence.

[26] **Des travaux de fond ont été menés entre 2012 et 2017 pour rapprocher les processus de traitements** (définition précise des périmètres secteur privé/fonction publique, des méthodes de redressement et de calcul des corrections de variations saisonnières, etc.) afin de converger vers des méthodes harmonisées permettant à chacun des acteurs de la coproduction de réaliser les expertises sur son champ de production (Acos sur le secteur privé hors intérim ; Dares sur l'intérim ; Insee sur le champ fonction publique, particuliers-employeurs et agricole) de façon à éviter une redondance des travaux sur le secteur privé et de dégager des moyens à l'extension de la couverture à l'ensemble du champ salarié. Concrètement, avec l'accord de coproduction, l'organisation est la suivante :

- la production des **évolutions trimestrielles de postes dans le secteur privé** (hors intérim, agriculture et particuliers-employeurs) relève de l'Urssaf Caisse nationale ;
- celle des **évolutions trimestrielles de l'emploi intérimaire** relève de la Dares ;
- l'Insee est responsable de la production des **évolutions de postes sur le champ de la fonction publique, des salariés de particuliers employeurs et des affiliés au régime agricole**. L'Insee est également responsable de l'intégration de l'ensemble des composantes pour calculer des évolutions d'emploi salarié : correction de la multi-activité et modélisation sur des données d'emploi annuelles.

[27] **Cet accord de coproduction sur les estimations trimestrielles d'emploi a permis aux utilisateurs de disposer d'un diagnostic consolidé et cohérent, sur l'ensemble du champ salarié.** Il a aussi conduit les trois organismes à partager leur expérience sur l'impact de la montée en charge de la DSN et à s'accorder sur une définition opérationnelle harmonisée de l'emploi, défini au sens du Bureau international du travail¹⁰⁴. Enfin, cet accord a permis des gains globaux en efficacité.

[28] L'accord a également entériné le rôle du SSP et notamment de l'Insee sur la diffusion des statistiques structurelles sur l'emploi total et les rémunérations des salariés, l'Urssaf Caisse nationale abordant ces sujets sous l'angle du recouvrement.

1.5 L'émergence de nouveaux systèmes d'information implique de clarifier le rôle et les relations entre les acteurs

[29] **L'instauration de la DSN a posé de nouvelles questions aux acteurs du SSP, quant au rôle des différents acteurs impliqués dans la production de bases de données et d'indicateurs statistiques.**

[30] **En premier lieu, ces questions ont été posées par le Gip-MDS** qui en tant que producteur du processus déclaratif souhaitait aller plus loin dans la diffusion. Le Gip-MDS avait élaboré en 2019 (note du 26 mars 2019) une proposition de service visant à la construction d'un dispositif visant à mettre à disposition des acteurs de la sphère travail-emploi des indicateurs à des fins à la fois décisionnelles et statistiques. L'objectif était de mettre à disposition des données les plus actuelles possibles à un certain nombre de partenaires, ces données n'étant pas qualifiées de statistiques publiques, car non diffusées. Un second objectif était de co-construire un système mutualisé avec les acteurs de la statistique publique, devant s'appuyer sur la base Dad-e (cf. annexe 1), assortie d'outils de requête. Les directeurs de l'Insee, de la Dares et de l'Urssaf Caisse nationale se sont opposés à ce projet (lettre du 11 avril 2019), pour deux motifs principaux : d'une part, un tel projet empiétait sur les missions dévolues au service statistique public, qui était par ailleurs engagé dans un mouvement de rénovation et de coordination de ses productions ; d'autre part, sans l'ensemble des expertises, validations et retraitements nécessaires pour fiabiliser les données brutes issues de la DSN, les chiffres qui risquaient d'être diffusés par différents canaux pouvaient générer de la confusion et de la perte de confiance globale dans les indicateurs diffusés. Le Gip-MDS a néanmoins mis en place une équipe dédiée à « l'offre data » au sein de sa direction de l'offre et de la transformation digitale, composée actuellement de 5 personnes, et cet axe de

¹⁰⁴ La mesure de l'emploi au sens du BIT nécessite une « traduction » de la déclaration DSN pour établir précisément le fait d'occuper un emploi à une date t ou non. Dans le cas général, toute personne dont le contrat de travail tel que déclaré par son employeur est ouvert le dernier jour du mois relève de l'emploi salarié. La définition de l'emploi ne retient pas a priori de critère sur la rémunération, permettant d'inclure dans l'emploi des personnes en congés maladie ou en chômage partiel, même si leur rémunération est nulle (cas de 0,9 million au chômage partiel et 0,3 million en congé maladie ou maternité en avril 2020 lors de la crise sanitaire).

Il existe des cas particuliers qui doivent être précisés, par exemple concernant les apprentis, les stagiaires de la formation professionnelle, les élèves fonctionnaires, les élus, les conscrits, les VRP, les pigistes, etc. Par ailleurs, certaines populations particulières dûment identifiées sont considérées comme à risque de bénévolat et ne sont pas comptées dans l'emploi au sens du BIT lorsque leur rémunération est « insuffisante » (conventionnellement moins d',/3 du Smic mensuel)

développement continue de figurer parmi les 5 axes de la nouvelle feuille de route du Gip-MDS couvrant la période 2022-2025, validée par son conseil d'administration.

[31] **En second lieu, et de façon plus collatérale s'agissant de l'usage de la DSN, est apparu un nouvel acteur dans le champ de la formation professionnelle, la Caisse des dépôts et consignations** qui met en œuvre le nouveau système d'information Agora sur la formation professionnelle et le Compte personnel de formation. La DSN est utilisée à la fois pour alimenter les comptes personnels et pour assurer un suivi de la situation en emploi des personnes ayant bénéficié d'une formation. L'émergence de ce nouveau SI pose des questions de coordination des productions avec celles de la Dares, de labellisation éventuelle par l'ASP des séries produites, et le cas échéant d'expertises conjointes.

1.6 Différents livrables statistiques sont aujourd'hui produits à partir de la DSN dans des délais qui se raccourcissent

[32] **Si les SI produisent distinctement, la coordination se fait en aval, au niveau des « livrables » ou des produits statistiques mis à disposition des utilisateurs.** Le tableau ci-dessous décrit plus précisément les principaux produits, séries et bases, issus de chacun des acteurs. Comme on le voit dans le tableau, les délais tendent globalement à se raccourcir aussi bien pour les indicateurs conjoncturels, principalement pour satisfaire les demandes européennes, et surtout pour la production des bases individuelles. Pour la BTS, les délais seront ainsi ramenés à un peu plus d'un an après la fin de l'exercice annuel. L'objectif visé pour la base Sismmo est une production à 30 jours après la fin du mois, 1 semaine de plus pour Post-Sismmo. Une partie de ces progrès est encore en devenir, avec la dernière phase d'expertises nécessaires sur le champ de la fonction publique qui vient d'être intégrée à la DSN en 2022. Cette réduction des délais est donc encore peu visible pour les usagers.

[33] **Outre la réduction des délais de production des indicateurs trimestriels et annuels, la DSN offre, du fait de son caractère mensuel, de nouvelles perspectives encore peu exploitées, à l'exception de l'emploi intérimaire.** Des exploitations à usage interne ou de pilotage à des horizons mensuels ont cependant été portées à la connaissance de la mission. L'Insee a ainsi fait état d'une production utilisée seulement en interne, portant sur les effectifs en fin de mois, la masse salariale et le volume d'heures rémunérées, avec un historique de ces variables constitué au niveau des établissements. Des usages à fréquence mensuelle de la DSN ont en particulier été mis en œuvre à l'occasion de la crise sanitaire, afin de disposer d'indicateurs réactifs, par exemple sur le volume d'heures rémunérées pour fournir des indications sur le chômage partiel massif au cours de cette période. L'Urssaf Caisse nationale produit également des données mensuelles ou à deux échéances infra-mensuelles (après le 5 ou le 15 du mois, la première échéance concernant les établissements de 50 salariés ou plus qui fournissent une première tendance), à des fins de pilotage ou d'aide à la décision, sur sollicitation de sa tutelle ou pour ses besoins internes.

Tableau 1 : Principaux livrables du SSP fondés sur la DSN

Producteurs	Indicateurs produits	Objet	SI	Partenaires	Fréquence de diffusion	Principaux concepts et traitements	Délais de production
Insee	Estimations d'emploi salarié	Les Estimations d'emploi fournissent des niveaux et évolutions d'emploi par secteur d'activité (jusqu'au niveau A88 de la Naf rév. 2) et par zone géographique (région, département, jusqu'à la zone d'emploi) ¹⁰⁵ .	Siera	Dares, Urssaf pour les estimations trimestrielles	Trimestrielle pour l'emploi salarié Annuelle pour l'emploi total	Emploi au lieu de travail en nombre de personnes, en fin de période (année ou trimestre) Définition de l'emploi au sens du BIT (document commun Dares/Insee/Urssaf) Séries corrigées des variations saisonnières. Traitement de la multi-activité : les personnes qui occupent plusieurs emplois à une même date sont comptées une seule fois, au titre de leur emploi principal (plus forte rémunération sur l'année : traitement fait en annuel, évolutions	Estimations trimestrielles nationales : objectif à T+60 jours (actuellement T+70 jours) Estimation flash : objectif à T+30 jours (actuellement T+45 jours) Estimations trimestrielles localisées : objectif à T+80 jours (actuellement T+90 jours)

¹⁰⁵ Pour les **estimations annuelles**, le niveau de détail disponible est le suivant : France métropolitaine : niveau A88 de la Naf rév.2 pour l'emploi salarié ; niveau A5 pour l'emploi non salarié ; Régions et départements : niveau A38 pour l'emploi salarié et niveau A5 pour l'emploi non salarié ; Zones d'emploi : niveau A5 pour l'emploi salarié et sans décomposition sectorielle pour l'emploi non salarié.

Pour les **estimations trimestrielles**, le niveau de détail disponible est le suivant : France : niveau A38 de la Naf rév.2 et quelques niveaux A88 (matériel de transport, construction, commerce, hébergement-restauration) ; Régions et départements : niveau A17.

Producteurs	Indicateurs produits	Objet	SI	Partenaires	Fréquence de diffusion	Principaux concepts et traitements	Délais de production
						trimestrielles estimées à partir des évolutions de postes de travail).	
	Base Tous Salariés – BTS (fichier annuel avec différents niveaux de granularité : postes, salariés, établissements, entreprises)	<p>Synthèse annuelle sur les postes salariés, issus de la DSN (sur le champ privé depuis 2017 (auparavant DADS)) et des déclarations des particuliers employeurs (chèque emploi service universel (Cesu), prestation d'accueil du jeune enfants (Paje) et déclaration nominative simplifiée (DNS). Jusqu'en 2021 : utilisation des fichiers de paie des agents de l'État, DSN à partir de 2022.</p> <p>La synthèse de l'année N porte sur les deux années N et N-1, afin d'appliquer</p>	Siera	Dares et SSM (fonction publique d'Etat, collectivités locales, Drees) qui participent à la validation des bases	Annuelle	<p>Principaux concepts :</p> <ul style="list-style-type: none"> - les salaires nets perçus : salaires/rémunérations horaires ou en équivalent temps plein, par secteur, lieu de travail, etc. ; - les caractéristiques des salariés : âge, sexe, localisation de travail ou de résidence, type de contrat de travail, PCS, etc. - les caractéristiques des établissements <p>Principaux traitements :</p> <ul style="list-style-type: none"> - sur la localisation fine des emplois (lieu de travail et de résidence) ; - sur la qualification des professions (reprise des codes PCS à partir des libellés en clair) ; - sur le calcul des rémunérations horaires ou en équivalent temps plein sur l'ensemble de 	Objectif à N+12 mois (actuellement, délais de l'ordre de N+24 mois, après avoir atteint N+36 mois au moment de l'intégration de la DSN)

Producteurs	Indicateurs produits	Objet	SI	Partenaires	Fréquence de diffusion	Principaux concepts et traitements	Délais de production
		des traitements homogènes pour produire des évolutions annuelles fiables.				l'année, pour lisser les éléments variables de rémunération sur la période correspondant au poste occupé (primes, etc.).	
	Fichier localisé sur les rémunérations et l'emploi salarié (Flores)	Synthèse produite à partir de BTS, avec enrichissements au niveau des établissements. 3 fichiers principaux : entreprises ; établissements ; comptage de postes, selon différentes caractéristiques (sexe, CSP, etc.).	Siera		Annuelle		Objectif à N+13 mois
	Indice du coût du travail ICT (« salaires seuls » et « salaires et charges »)	L'ICT « salaires et charges » est la somme de la masse salariale, des charges sociales et des autres coûts (nets des exonérations), rapportée au volume horaire de travail. L'ICT « salaires	Siera		Trimestrielle	Indice de Laspeyres chaîné ayant pour base 100 en 2012, révisable. Prise en compte du volume horaire d'activité (contrairement au salaire moyen par tête (SMPT), qui se rapporte aux effectifs en emploi) et des évolutions	Objectif à T+ 65 jours (actuellement T+ 75 jours) Estimation flash à T+ 45 jours (projet)

Producteurs	Indicateurs produits	Objet	SI	Partenaires	Fréquence de diffusion	Principaux concepts et traitements	Délais de production
		seuls » est un indice de coût du travail hors charges sociales.				structurelles (contrairement au salaire mensuel de base (SMB), qui mesure une évolution du prix du travail à structure constante).	
Dares	Mouvements de main d'œuvre (MMO)	Suivi des contrats de travail dans les établissements employeurs sur le champ privé hors agriculture, intérim et particuliers employeurs ¹⁰⁶	Sismmo-MMO		Trimestrielle	Redressements statistiques par contrôles de gestionnaires (retour éventuel au déclarant) et avec programme de détection et de correction automatique des doublons de contrats (croisements entre plusieurs déclarations mensuelles successives). Modélisation du dernier mois pour anticiper les révisions	T+ 90 jours
	Emploi intérimaire		Sismmo-Intérim	Insee, Urssaf (coproduction Estimations trimestrielles	Mensuelle et trimestrielle	Redressements spécifiques aux Entreprises de travail temporaire (par exemple,	Estimation mensuelle à T+40 jours Trimestriel : en même temps que Estimations

¹⁰⁶ Extension en cours au secteur public.

Producteurs	Indicateurs produits	Objet	SI	Partenaires	Fréquence de diffusion	Principaux concepts et traitements	Délais de production
				d'emploi)		déclaration de CET en tant que missions d'intérim)	trimestrielles d'emploi
Urssaf CN	Effectifs salariés du secteur privé hors agriculture par secteur d'activité/ zones géographiques	Niveaux et évolutions selon nomenclatures NACE17 et NACE38 et par zones géographiques (régions, départements, zones d'emploi)	Pléiade	Dares, Insee (coproduction Estimations trimestrielles d'emploi)	Trimestrielle	Les effectifs salariés (postes) sont estimés sans traitement de la multi-activité. Traitements harmonisés avec Insee et Dares sur champ public/privé et méthodes.	Estimations trimestrielles nationales : objectif à T+60 jours (actuellement T+70 jours) Estimation flash : objectif à T+30 jours (actuellement T+45 jours) Estimations trimestrielles localisées : objectif à T+80 jours (actuellement T+90 jours)
	Masse salariale (assiette déplafonnée soumise à cotisations et contributions) dans le secteur privé hors agriculture	En cohérence avec effectifs	Pléiade		Trimestrielle		Idem

Producteurs	Indicateurs produits	Objet	SI	Partenaires	Fréquence de diffusion	Principaux concepts et traitements	Délais de production
	Exonérations de cotisations						

2 Les acteurs de la statistique publique réalisent de nombreuses expertises et retraitent les données brutes issues de la déclaration DSN

[34] La DSN n'échappe pas aux écueils habituels de l'usage des déclarations administratives à des fins statistiques, malgré les progrès importants enregistrés grâce au travail de normalisation et aux contrôles réalisés en amont lors du processus déclaratif, au travers du cahier des charges de la norme NÉODeS. Différentes expertises sont ainsi réalisées par les métiers statistiques, à la fois pour traiter les anomalies constatées, afin d'assurer la fiabilité statistique au regard des objectifs poursuivis, et pour poser les concepts utiles à la production de statistiques homogènes dans le temps et dans l'espace¹⁰⁷. Ces expertises sont nécessaires parce que la DSN est encore un processus déclaratif « jeune », mais son caractère évolutif et l'impossibilité de « tout contrôler » au moment de la déclaration rendront nécessaire un traitement de validation des données, même en régime pérenne. En outre, plus la granularité des données utilisées est fine (par région ou par département...), plus le travail de mise au point est important. Les principaux traitements réalisés par les statisticiens (Dares, Insee, Urssaf Caisse nationale) sont décrits ci-dessous.

2.1 Traitements pour corriger et documenter les erreurs déclaratives, avec des contrôles de cohérence interne ou externe

[35] Même si la qualité des données progresse au fil des années grâce au déploiement d'actions de fiabilisation, de contrôles automatiques et de réglages des logiciels de paie, les traitements en aval restent nécessaires parce que tout n'est pas contrôlé et parce que l'acte déclaratif est en constante évolution : chaque année de nouvelles entreprises se créent et déclarent une DSN pour la première fois, avec un potentiel accru d'erreurs ; par ailleurs les normes déclaratives se modifient au gré des changements législatifs ou réglementaires. Du fait du nombre important d'informations à déclarer, les employeurs et les éditeurs peuvent rencontrer des difficultés à respecter les évolutions de la norme à la maille nominative ou agrégée, générant des retards dans la prise en compte des évolutions. Les anomalies déclaratives sont particulièrement importantes quand de nouvelles formalités sont intégrées en DSN, comme cela a été le cas en 2020 pour la DOETH.

[36] **De façon classique, les informations les plus fiables sont celles qui font l'objet de contrôles bloquants et celles ont une incidence directe sur les employeurs avec donc une incitation pour eux à « bien » déclarer en DSN.** A ce stade, le secteur de l'Insertion par l'activité économique (IAE) et les contrats aidés dans le secteur marchand sont largement sous-identifiés dans les déclarations DSN (moins de la moitié).

[37] Certaines erreurs déclaratives peuvent être détectées sur des critères de cohérence internes à la DSN (ou le cas échéant en chaînant plusieurs DSN successives). D'autres ne sont

¹⁰⁷ Cf. notamment P. Rivière, Courrier des statistiques n° N1, « Utiliser les déclarations administratives à des fins statistiques »
<https://www.insee.fr/fr/information/3647035>

déTECTABLES que par confrontation avec d'autres sources externes. A titre d'exemple, les licenciements économiques de la DSN peuvent être confrontés à d'autres statistiques externes (inscription à Pôle Emploi suite à un licenciement économique, déclarations préalables à l'embauche, ou encore ruptures conventionnelles homologuées) et il s'avère qu'ils seraient nettement sous-estimés s'ils étaient calculés sur les seules données brutes de la DSN.

2.2 Traitements pour passer de la date de la déclaration à la date de référence afin de gérer les retardataires et les déclarations rectificatives

[38] Dans tous les processus déclaratifs, la question de la datation précise est un enjeu important : **il convient ainsi de distinguer la date d'obtention des informations (les déclarations peuvent arriver en retard ou être corrigées lors de déclarations mensuelles ultérieures) et la date de référence à laquelle se réfère cette information.** Il est ainsi nécessaire de chaîner plusieurs déclarations mensuelles successives pour s'assurer de l'exhaustivité et de la précision des informations se rapportant à une période ou à une date déterminée.

[39] Ces retards ont un impact plus ou moins important sur l'information contenue dans la DSN d'un mois donné en fonction de la nature de l'information à laquelle on s'intéresse. Avec un recul à M+1 mois, on observe 4 à 5 % de salariés déclarés en retard (contre 1 % à M+3 mois), mais de l'ordre de 30 % des flux de mouvements de main-d'œuvre sont mal déclarés à M+1 mois. **Il y a donc une tension entre le recul nécessaire pour obtenir des estimations à une date donnée et la demande d'une information la plus contemporaine possible.** Les défauts de contemporanéité peuvent varier selon les sujets : par exemple, une statistique qui s'appuie sur le numéro de contrat et le « bloc changement » de la DSN nécessite davantage de contrôles qu'une statistique qui s'appuie uniquement sur l'identité du salarié et de l'établissement. Même si on se donne la possibilité de réviser des estimations précoces, il convient de s'assurer de leur capacité à délivrer un signal utile, non remis en cause systématiquement en raison de bruits déclaratifs. A titre d'exemple dans un domaine différent, on peut ainsi citer le cas des indicateurs d'effectifs de demandeurs d'emploi inscrits à Pôle emploi : la diffusion trimestrielle de ces indicateurs est maintenant privilégiée dans les analyses, les indicateurs mensuels ayant été jugés trop volatils.

2.3 Traitements correctifs des doublons liés à des défauts d'identifiant des contrats

[40] **Des doubles déclarations de contrats peuvent être constatées en cas de changement non déclaré des valeurs identifiantes d'un contrat (voire d'un individu en cas de changement/erreur de Nir), qui cause à la fois la disparition d'un contrat (on ne retrouve plus ses identifiants) et l'apparition d'un nouveau contrat (qui n'est en réalité pas nouveau).** Dans ce cas, une embauche et une fin de contrat sont comptés à tort.

[41] La Dares a ainsi identifié des erreurs en travaillant sur plusieurs déclarations DSN successives d'un même employeur. Pour la production des statistiques sur les mouvements de main-d'œuvre, la Dares corrige les contrats considérés comme « disparus à tort », soit en chaînant le contrat avec un autre dans les DSN suivantes pour ne pas le compter deux fois, soit en imputant un motif de fin de contrat (démission, licenciement pour motif personnel, rupture conventionnelle, fin de

période d’essai, licenciement économique etc.), en tenant compte notamment de la durée du contrat ou encore de la catégorie socio-professionnelle du salarié.

2.4 Traitements adressant les incohérences entre les données agrégées (bloc 1) et les données individuelles (bloc 3)

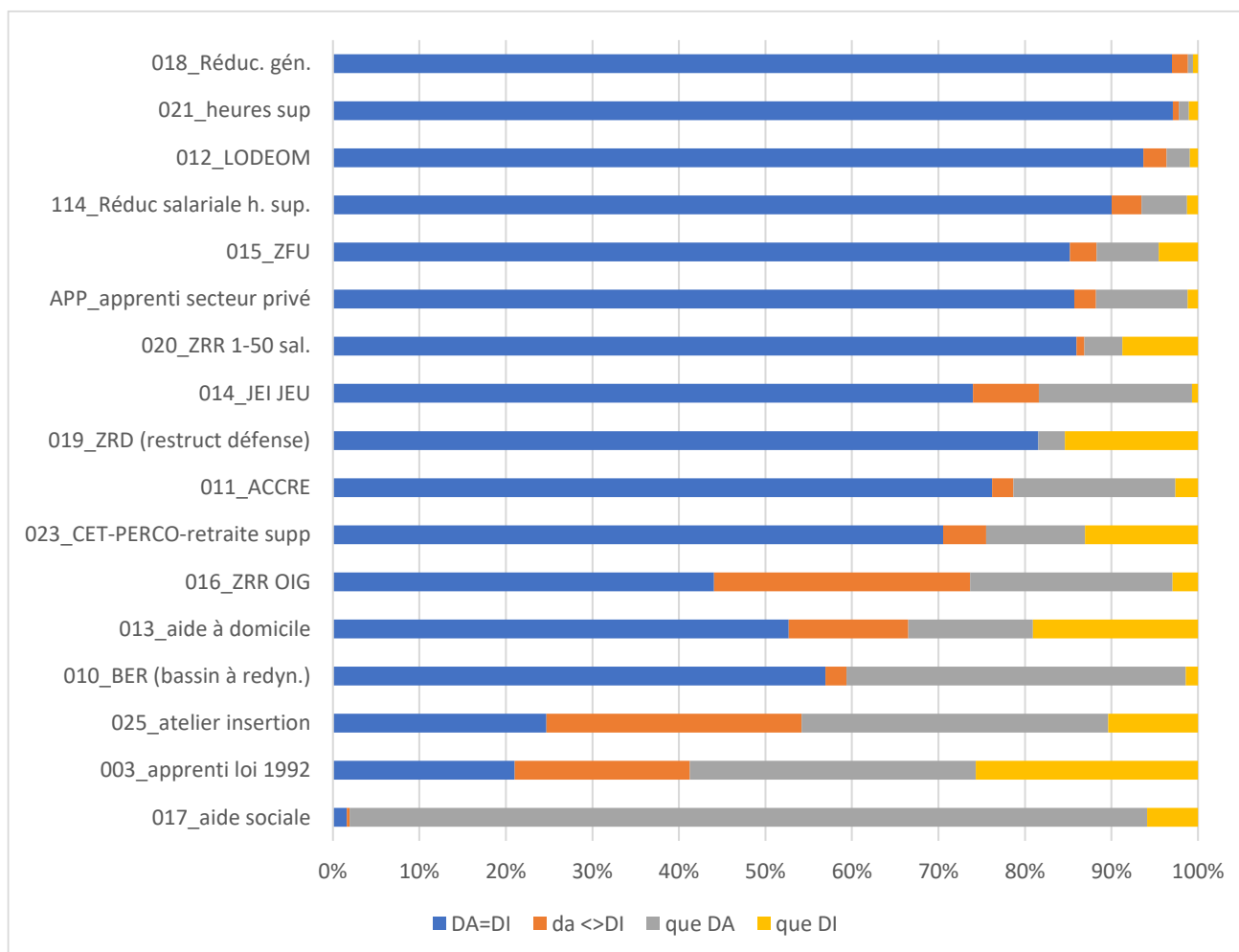
[42] Les flux DSN semblent comprendre des données redondantes : les données agrégées (DA) pour le calcul des cotisations/contributions et les données individuelles (DI). Ces informations sont issues du même flux et de la même source (les éléments de la paie) et devraient être cohérentes. Seules l’Urssaf Caisse nationale et la CCMSA sont destinataires des deux flux, au titre de leur activité de recouvrement.

[43] Du point de vue de l’Urssaf Caisse nationale, il existe à ce jour des écarts entre les deux flux, qui ne permettent pas encore d’envisager la suppression des flux agrégés.

[44] L’Urssaf a identifié 3 causes d’écarts entre DI et DA :

- La qualité insuffisante des DI pour assurer le calcul des cotisations et contributions : les travaux de fiabilisation menés en 2020 ont permis de mesurer que les régularisations sont 9 fois plus importantes relativement à l’assiette déplafonnée des données individuelles que celles concernant les données agrégées. Il en va de même pour le calcul de certains allègements généraux de charge.
- L’information agrégée est plus réactive : du fait du nombre important d’informations à déclarer, les cotisants et les éditeurs ont en effet des difficultés à suivre les évolutions de la norme à maille nominative, générant une inertie dans la prise en charge des évolutions. Cette dernière est moins importante pour la donnée agrégée car la création d’un nouveau code type est relativement facile et le cotisant s’adapte plus vite, l’information à communiquer étant souvent moins complexe.
- Pour certains dispositifs, seule la donnée agrégée permet le calcul des cotisations et contributions, la reconstitution de certaines assiettes étant impossible à partir de la norme DSN actuelle. C’est le cas par exemple de l’indemnité versée par l’employeur à son salarié en cas d’activité partielle et de la réduction de cotisations salariales assises sur les heures supplémentaires.

L'Urssaf Caisse nationale a ainsi produit une analyse détaillée sur le taux de cohérence entre données agrégées et données individuelles pour les dispositifs d'exonérations et réduction à vocation générale en septembre 2021 :



Note de lecture : Pour le dispositif d'exonération concernant les apprentis loi de 1992, seuls 21 % des établissements (en nombre) ont des montants d'exonérations issus de données individuelles et des données agrégées cohérents (écart inférieur à 1 % du montant de DA) pour la période d'emploi de septembre 2021. Un quart des DSN ont seulement des informations individuelles et un tiers des informations agrégées seules. 20 % des DSN ont un écart de montant d'exonération supérieur à 1 % de celui de la DA.

[45] S'agissant de la CCMSA, de tels écarts n'ont pas été mentionnés à la mission, car le dispositif de fiabilisation des données individuelles y est ancien et a perduré avec la mise en place de la DSN. Hors MSA, les enjeux de fiabilisation de la donnée individuelle se posent : ils doivent faire l'objet de nouveaux investissements et devraient être coordonnés par les Urssaf, à l'instar de ce qui est fait à la MSA.

2.5 Des traitements pour construire des concepts et unités d'intérêt et se conformer à des nomenclatures de référence

[46] **Pour le réemploi des sources administratives à des fins statistiques, les rubriques administratives doivent être traduites en concepts statistiques, de façon à permettre de mesurer les agrégats économiques d'intérêt.** Ces agrégats font par ailleurs l'objet de normalisation pour permettre leur usage à des fins de comparaisons avec d'autres domaines ou internationales.

- **S'agissant de l'emploi** : le concept d'emploi est défini au sens du Bureau International du Travail : il regroupe les personnes âgées de 15 ans ou plus ayant travaillé au moins une heure rémunérée pendant une semaine donnée, dite « de référence », ainsi que celles qui sont en emploi, mais qui n'ont pas travaillé durant cette semaine de référence pour un certain nombre de raisons répertoriées (congrés, arrêts maladie, chômage partiel, etc.). De façon opérationnelle, il convient ainsi de traiter les situations particulières comme les périodes de congés pris en charge par des caisses de congés payés (bâtiment) ou de congés maladie, maternité (en cas de non-subrogation). Sur un plan technique, il convient aussi de définir les périodes d'emploi, voir si on retient la maille des contrats ou celles des postes (même individu chez un même employeur) ;
- **S'agissant des qualifications** : des opérations de codification sont mises en œuvre pour la mesure de la qualification des emplois selon la nomenclature statistique des PCS : les déclarations contiennent des libellés décrivant la profession occupée par chaque salarié, ainsi que le code PCS déclaré par l'établissement. Une recodification des libellés en clair est effectuée pour s'assurer de la qualité du code relativement à la nomenclature officielle, par un recours à une base d'apprentissage en automatique puis en traitement par des gestionnaires le cas échéant ;
- **S'agissant des rémunérations** : le salaire net perçu doit être recalculé à partir du net fiscal et le salaire brut à partir de l'assiette CSG dé plafonnée. Le salaire net doit être mis en regard de la durée rémunérée, de façon à pondérer son poids dans le calcul de la moyenne. Enfin, La DSN représente la paie du mois. Les décalages de paie, rappels de salaires et déclarations en retard nécessitent des redressements et imposent la plupart du temps de confronter plusieurs DSN mensuelles successives pour conforter leur complétude et leur cohérence sur une période d'emploi.

[47] **Les statisticiens publics visent également à produire une information homogène dans le temps et produire des « séries longues ».** Chaque changement dans la législation et dans la DSN native oblige à s'interroger sur les concepts et à la façon dont la mesure des concepts doit être actualisée pour tenir compte des modifications. Certaines modifications réglementaires peuvent avoir un impact sur les rubriques déclarées : par exemple, lorsque que la mutuelle d'entreprise est devenue obligatoire (cf. loi de finances 2014, effective en 2016), la part patronale des cotisations employeurs de complémentaire santé est devenue imposable alors qu'auparavant elle était considérée comme un avantage en nature non imposable. Pour établir la statistique d'évolution annuelle du salaire, il a fallu retraiter cette extension de l'assiette imposable qui ne devait pas être confondue avec une hausse du salaire net perçu.

2.6 Des traitements pour contrôler la localisation afin de fournir des indicateurs à des niveaux géographiques fins

[48] La DSN est normalement déclarée par l'établissement, donc au lieu de travail. Il arrive cependant que certains employeurs, notamment dans le secteur public, déclarent leurs salariés de façon groupée et non pas dans l'établissement où ils travaillent. Des redressements de la localisation au lieu de travail sont donc nécessaires pour permettre l'usage des données à des niveaux géographiques fins, comme la région ou la commune.

2.7 Une boucle de rétroaction auprès du Gip-MDS pour fiabiliser en amont la déclaration

[49] Les résultats des contrôles effectués par les métiers statistiques, comme par les autres métiers en charge de la gestion des droits ou du recouvrement, peuvent être remontés auprès du Gip-MDS pour faire évoluer la norme et les contrôles embarqués, ou identifier les défauts déclaratifs selon les éditeurs concernés ou améliorer la communication auprès des déclarants en la ciblant sur les erreurs les plus fréquentes.

[50] Une des évolutions engagées actuellement concerne l'amélioration du codage de la PCS en amont, dans le cadre d'un groupe de travail Insee-Dares-Gip-MDS-Pôle emploi, expérimentant la possibilité de fournir une liste de 10 000 items pour aider à coder la PCS en amont dès la déclaration, puis sa transcription en code Rome (V3 puis V4). Si l'expérimentation est concluante, le Gip-MDS pourra étudier sa mise en application en modifiant le cahier technique, en lien avec les éditeurs.

3 La DSN a démultiplié les possibilités d'appariements pour enrichir les analyses et assurer un suivi des dynamiques d'insertion professionnelle

[51] Les appariements de bases de données afin d'enrichir l'analyse des liens entre différents domaines ou de liens de causalité à l'issue de certains événements ont été portés par l'ASP et par le Cnis dans son programme de moyen terme 2019-2023. **Les appariements sont ainsi devenus explicitement un mode de constitution de nouvelles sources statistiques, soumis à l'approbation du Cnis comme les enquêtes.** Au sein du SSP, les procédures d'appariement sont maintenant simplifiées avec la création d'un « code statistique non signifiant » transverse (en lieu et place du Nir), pouvant figurer dans toutes les bases statistiques, ce afin d'industrialiser les possibilités d'appariements entre bases contenant ce code. En revanche, si une base ne contient pas ce code, alors l'appariement risque d'être plus coûteux : il faut en effet pouvoir rapprocher automatiquement les deux bases sur des traits d'identité (nom, prénom et adresse par exemple) avec des taux de réussite pouvant varier selon la qualité des deux sources et consacrer du temps pour traiter manuellement les erreurs et échecs de rapprochement, ou *a minima* s'interroger sur les risques de biais éventuel, les non-appariés pouvant présenter des caractéristiques spécifiques (personnes nées à l'étranger par exemple).

[52] L'exhaustivité des DSN sur leur champ, la qualité des identifiants individuels et la possibilité d'avoir une information à différents horizons temporels ont démultiplié l'intérêt de ces enrichissements, à la fois pour des usages statistiques, mais aussi pour des usages métiers. **L'usage principal de ces appariements est de permettre un suivi de trajectoires individuelles, notamment d'insertion sur le marché du travail, à l'issue d'une formation ou après avoir bénéficié d'une mesure de politique d'emploi ou après avoir connu un épisode de chômage ou un problème de santé.** Ce type d'usage passait auparavant par le « panel DADS » de l'Insee, mais il était limité par le sondage au 1/12 de ce dernier.

[53] **La Dares en tant que producteur de la source Sismmo et du fait de ses missions d'évaluation joue ainsi un rôle pivot dans ces appariements, en collaboration avec différents partenaires du SSP, du service public de l'emploi et de la Cnaf.** Les principaux dispositifs existants sont décrits ci-dessous.

3.1 Le panel ForCE, pour suivre les trajectoires des personnes ayant eu un contact avec le service public de l'emploi

[54] Le panel Formation-chômage-emploi (ForCE) a été mis en place en 2020 par la Dares en collaboration avec Pôle emploi dans le cadre de l'évaluation du Plan d'Investissement dans les Compétences (PIC). Il s'agit d'un dispositif permanent de croisement des bases statistiques sur la FORMation, sur le Chômage et l'Emploi, dont l'enjeu est de **permettre la reconstitution des trajectoires professionnelles de toutes les personnes ayant eu un contact avec le service public de l'emploi (missions locales, Pôle emploi) ou ayant suivi une formation professionnelle prise en charge totalement ou partiellement par les pouvoirs publics.**

[55] Les données sont constituées par appariement de quatre bases de données :

- le Fichier historique des demandeurs d'emploi (FH) qui rassemble pour chaque demandeur d'emploi des informations sur ses épisodes d'inscription à Pôle emploi sur 10 années glissantes ;
- la base I-Milo (issue du système d'information des missions locales), qui rassemble pour chaque jeune suivi en mission locale des informations sur les contacts avec la mission locale et les dispositifs suivis en mission locale, depuis 2016 ;
- la base régionalisée des stagiaires de la formation professionnelle (Brest), qui rassemble pour chaque personne en recherche d'emploi, stagiaire de la formation professionnelle, les caractéristiques des formations suivies depuis 2017 ;
- la base Mouvements de main-d'œuvre (MMO), puis DSN, qui rassemble pour chaque salarié les informations sur les contrats de travail, depuis 2017.

[56] Les appariements sont réalisés par le Centre d'accès sécurisé aux données (CASD) sur la base du Nir préalablement chiffré, ou d'informations identifiantes (noms, prénoms, date de naissance). Les données du dispositif ForCE sont actualisées à un rythme trimestriel et mises à disposition d'équipes de recherche sur le CASD, après habilitation par le Comité du secret statistique. Chaque trimestre T, les données disponibles permettent de retracer les trajectoires jusqu'au trimestre T-3. Les délais d'intégration ont été ajustés afin d'intégrer les DSN retardataires

et correctives, qui peuvent avoir un impact important sur l'analyse, notamment sur les durées de contrats observées.

3.2 Le panel Trajam, pour suivre les trajectoires des jeunes passés par une mesure de politique d'emploi

[57] **Le panel Trajam permet de suivre les trajectoires d'emploi salarié des jeunes (sur la période 2010-2017 à partir du panel « tous salariés » de l'Insee) et leur participation à diverses mesures actives du marché du travail jusqu'à leurs 35 ans** (à partir des données issues des différents systèmes d'information de gestion. Le champ du panel Trajam est constitué des jeunes passés par au moins un dispositif sur la période.

[58] Le dispositif Trajam s'apparente au panel Force qui lui a succédé : Trajam couvrait moins de dispositifs et se limitait au devenir en emploi dans le secteur privé (hors particuliers employeurs et une partie de l'agriculture).

3.3 Le panel Midas, pour suivre les trajectoires des inscrits à Pôle emploi et des bénéficiaires de minima sociaux

[59] **Le dispositif Midas piloté par la Dares, en collaboration avec Pôle emploi et la Caisse nationale des allocations familiales, vise à reconstituer les trajectoires professionnelles des inscrits à Pôle emploi et des bénéficiaires de minima sociaux.**

[60] Midas croise trois sources de données relatives :

- aux contrats salariés issus de Sismmo fournis par la Dares ;
- à l'indemnisation des demandeurs d'emploi (données issues du Fichier historique statistique et du Fichier national des allocataires de Pôle emploi) ;
- aux bénéficiaires de minima sociaux (le revenu de solidarité active, la prime d'activité, l'allocation aux adultes handicapés), suivis par la Caisse nationale des allocations familiales (données Allstat-FR6 de la Cnaf).

[61] La convention signée en juin 2022 prévoit, à l'automne 2022, une première mise à disposition des données appariées sur la période 2017-2021, puis courant 2023, sur la période 2017-2022. Ensuite, l'appariement se fera selon une fréquence annuelle avec une profondeur temporelle croissante, couvrant jusqu'à 2017-2025.

3.4 Le dispositif InserJeunes, pour suivre l'insertion professionnelle des sortants des filières professionnelles

[62] Cet appariement est réalisé à la fois pour produire une **base de connaissance générale**, mais aussi pour produire des **indicateurs opérationnels** au niveau des établissements de formation. Le secret statistique n'est alors pas recherché au niveau des établissements et formations au sein des établissements, la diffusion de ces indicateurs d'aide à la décision étant prescrite par des textes réglementaires.

[63] **La Depp et la Dares ont développé le dispositif de suivi InserJeunes portant sur l'ensemble des jeunes du système éducatif en filières professionnelles, y compris ceux en apprentissage et dans les filières agricoles.** Une base de données rassemblant les sortants de ces filières est construite par comparaison des présents d'une année sur l'autre grâce à l'identifiant des élèves (INE) devenu exhaustif et valable pour toutes les académies. L'appariement de cette base de sortants se fait avec la DSN plusieurs fois pour faire un suivi à six, douze et vingt-quatre mois après leur sortie du système éducatif.

[64] InserJeunes a remplacé les enquêtes Insertion dans la vie active (IVA) et Insertion professionnelle des apprentis (IPA) menées depuis les années 1990 par la Depp. Il fournit des résultats à un niveau beaucoup plus détaillé qu'avant, jusqu'au niveau de l'établissement et de la formation.

3.5 Le dispositif InserSup, pour suivre l'insertion professionnelle des sortants de l'enseignement supérieur

[65] La sous-direction des systèmes d'information et des études statistiques du Ministère de l'enseignement supérieur (SIES) est en train de développer InserSup, l'équivalent de InserJeunes pour suivre l'insertion professionnelle des sortants de l'enseignement supérieur. Le principe est le même : constitution d'une base de sortants, appariements avec Sismmo pour mesurer l'insertion en emploi des sortants de l'enseignement supérieur. Les appariements se font en Code statistique non signifiant (CSNS) à partir des données d'état civil et l'identifiant INES utilisé dans l'enseignement supérieur.

[66] InserSup remplacera les enquêtes d'insertion professionnelle existantes (doctorants, masters, licences pro, DUT) et permettra d'étendre le suivi de l'insertion aux sortants des grandes écoles.

3.6 Autres projets

D'autres projets sont en cours de développement sur des problématiques sectorielles, comme le projet Badiane de la Drees, qui vise à rassembler des informations relatives au fonctionnement, à l'activité, au personnel et au public accueilli dans les structures médico-sociales (ESMS), les données issues de la DSN alimentant le volet personnel. Des questions d'articulation avec les produits et analyses issus des dispositifs de pilotage Ramses et Diamant (voir annexe 4) pourront se poser.

4 Le service statistique public met à disposition de très nombreux utilisateurs les bases de données statistiques issues de la DSN dans le cadre sécurisé du CASD

4.1 Les bases de données statistiques issues de la DSN sont mises à disposition par les différents organismes producteurs

[67] Les principales bases de données issues de la DSN présentes sur le CASD sont, par organisme producteur :

- Insee
 - BTS-Postes : base Tous salariés : fichier Postes
 - BTS-Salariés : base Tous salariés : fichiers Salariés
 - BTS-Etablissements : base Tous salariés : fichier Etablissements
 - BTS-Entreprises : base Tous salariés : fichier Entreprises
 - Flores Entreprises : Fichier localisé des rémunérations et de l'emploi salarié - Entreprises
 - Flores Etablissements : Fichier localisé des rémunérations et de l'emploi salarié - Etablissements
 - Flores Postes : Fichier localisé des rémunérations et de l'emploi salarié - Postes
 - Flores Flux de salariés : Fichier localisé des rémunérations et de l'emploi salarié - Flux de salariés
 - Panel tous salariés (ancien panel DADS)
 - Panel tous salariés - Echantillon Démographique Permanent
- Dares
 - MMO : Mouvements de Main d'Œuvre
 - Trajam : Trajectoires des Jeunes Appariées aux Mesures actives du marché du travail
 - Midas : Minima sociaux, droits d'assurance chômage et parcours salariés
 - Force : Formation, Chômage et Emploi
- Urssaf Caisse nationale
 - Sequoia_Assiette : Informations issues du suivi conjoncturel de la masse salariale du secteur privé
 - Sequoia_effectifs : Informations issues du suivi conjoncturel des effectifs salariés du secteur privé
- Le CASD permet également d'accéder à nombre d'autres sources comme toutes les enquêtes de la statistique publique sur le thème de l'emploi (enquête emploi, enquêtes

Acemo, ECMO, Conditions de travail, Reponse, Besoins en main d'œuvre, etc.)¹⁰⁸, ou à d'autres domaines, comme le Système national des données de santé (SNDS) produit par la Cnam. Seules les données de l'Education nationale sont à ce jour absentes du CASD, ce ministère ayant en projet la construction d'un hub dédié pour les chercheurs de sa sphère, tout en envisageant à terme d'alimenter aussi le CASD.

4.2 Le CASD offre à prix coûtant un environnement d'accès aux données et une infrastructure de calcul, avec un dispositif de sécurisation via un boîtier

[68] Le CASD offre un environnement de travail complet : données, logiciels, puissance de calculs et gestion de la sécurité. Chaque utilisateur du CASD travaille au sein d'un environnement partagé uniquement avec les autres membres du projet. Cet environnement, de type Windows, offre une palette de logiciels (bureautique, statistiques, calcul scientifique, etc.) qu'il est possible de compléter.

[69] Toutes les données sources pour lesquelles le projet dispose d'une habilitation sont accessibles. Des procédures spécifiques ont été mises en place notamment pour les exports : il s'agit de vérifier que les fichiers de résultats sont conformes aux règles de confidentialité définies par le déposant des données (secret statistique, secret fiscal...).

[70] La SD-BOX est un boîtier qui assure la sécurisation de l'accès aux données via l'infrastructure centrale du CASD et l'environnement de travail de l'utilisateur en intégrant de bout en bout les prérequis les plus exigeants en matière d'homologation du poste d'accès (ISO 27001, ISO 27701, homologation au référentiel de sécurité des données de santé, PSSI-MCAS...).

[71] Les services du CASD sont payants, à prix coûtant, le CASD étant un Gip à but non lucratif. Un projet avec plusieurs utilisateurs coûte en général de 3 à 6 k€ par an. Ce coût est intégré et jugé raisonnable par les utilisateurs rencontrés par la mission (et semble-t-il par les centaines d'utilisateurs du CASD) mais il a été mentionné que ce coût peut être jugé élevé par certains chercheurs qui ne l'ont pas prévu dans leur budget ou dans la sphère de l'Education nationale. Inversement, un tel coût est considéré par certains chercheurs comme insuffisant pour une demande de financement ANR.

4.3 Un accès important et croissant aux bases de données du CASD sous le couvert d'une procédure d'habilitation

4.3.1 ... assurée par le Comité du secret

[72] Le CASD a été créé à l'origine pour offrir des modalités d'accès aux données protégées par le secret statistique posé par la loi de 1951. Selon les termes de la loi, le Comité du secret

¹⁰⁸ On peut ainsi citer à titre d'exemple la source DPAE : Déclaration préalable à l'embauche (Urssaf Caisse nationale) ; la source ENIACRAMS : Echantillon national inter-régimes de compléments de revenus d'activité et de minima sociaux (Drees) ; D@ccord : Base statistique des accords collectifs d'entreprise (Dares) ; STMT : Statistique mensuelle du marché du travail (Dares, Pôle emploi)

statistique doit donner son avis sur les demandes de communication de données individuelles collectées par voie d'enquête statistique ou transmises au service statistique public, à des fins d'établissement des statistiques. Ce rôle du Comité du secret a été étendu pour gérer l'ouverture des données soumises au secret fiscal (cf. Article L135 D du Livre des procédures fiscales, qui confère au Comité du secret le même rôle).

[73] Le Comité du secret examine les demandes d'accès en séance (pour des projets nouveaux) ou en consultation électronique (pour des modifications de projets existants). L'avis du Comité prend de l'ordre de 2 à 4 mois à compter de la demande pour les projets nouveaux, compte tenu du rythme trimestriel de ses séances. Les demandes sont examinées à la lumière des critères suivants : (i) l'objet de l'étude qui justifie la demande présente-t-il un intérêt général ? ; (ii) le demandeur présente-t-il toutes les garanties pour la protection des données confidentielles qui pourraient lui être transmises ? ; (iii) la communication de ces informations ne risque-t-elle pas de porter atteinte aux règles de la concurrence que la loi sur le secret statistique a entendu protéger ? ; (iv) ces informations ne risquent-elles pas d'être utilisées à des fins de contrôle fiscal ou de répression économique, usages interdits par l'article 6 de la loi de 1951 ? ; (v) les informations demandées ne sont-elles pas excessives par rapport à l'étude qui les justifie ?

[74] Outre la procédure standard s'appliquant aux projets de recherche et d'études, le Comité du secret a mis en place une procédure « simplifiée »¹⁰⁹ qui permet à certains organismes d'avoir un accès a priori à un ensemble spécifié de fichiers. Cela concerne une douzaine d'institutions, comme l'Assemblée Nationale, l'IGF, la Direction Générale du Trésor, la Cour des Comptes, l'Igas, etc.

[75] Le nombre de demandes d'accès aux sources déposées au CASD passant par le Comité du secret s'élève à 549 en 2022 (dont 241 pour des nouveaux projets) et est en constante augmentation. Entre 2018 et 2022, la hausse des demandes s'est établie en moyenne à 7 % par an (9 % pour les seuls nouveaux projets).

4.3.2 ... ou assurée directement par le producteur

[76] Plus généralement, le CASD peut offrir un service d'accès sécurisé, selon d'autres procédures d'habilitation que celles relevant du Comité du secret. Ainsi un producteur souhaitant diffuser ses données confidentielles auprès d'acteurs externes peut les déposer au sein du CASD ; indiquer les personnes habilitées à accéder à ces données et pour quelle durée ; indiquer les personnes ou les entités habilitées à recevoir un boîtier d'accès à ces données ; définir les règles de confidentialité à faire respecter. Cette habilitation directe par le producteur, sans délégation au Comité du secret, est pratiquée notamment par la Banque de France et dans le domaine de la santé, par la Cnam, l'Atih, l'Inserm, l'Irdes ou les responsables de cohortes (Constances, Gazel, etc.).

¹⁰⁹ La procédure simplifiée permet à des organismes ayant des besoins récurrents et fréquents de données de bénéficier d'un accès à des sources protégées sous couvert de modalités d'instruction allégées. Ces organismes en contrepartie sont tenus de communiquer tous les ans au Comité un bilan de leurs activités justifiant des usages des données accédées.

4.4 De nombreux projets d'études mobilisent les sources issues de la DSN

[77] Le CASD héberge environ 500 sources de données. La DSN est la source du CASD la plus utilisée en comptant ses dérivés (BTS, Flores, FDT, MMO, Midas, Force, panel salariés, etc.). De nombreux chercheurs utilisent la base BTS, qui offre un suivi homogène dans le temps, avec des retraitements statistiques de qualité, ainsi que la possibilité d'appariements avec d'autres sources via le Siret. A date, il y a plusieurs centaines de projets de recherche en cours utilisant des données DSN, soit environ 1 000 utilisateurs.

[78] Beaucoup de chercheurs sont satisfaites de pouvoir mener des travaux sur les données traitées statistiquement afin de bénéficier des investissements réalisés par le SSP. Certains, minoritaires cependant, expriment un besoin d'accès aux données brutes qu'ils peuvent souhaiter retraiter selon leurs propres usages et objectifs, afin d'accéder à des données absentes des bases statistiques ou d'en disposer plus rapidement. La réponse à ce type de demande d'accès aux données brutes n'est pas organisée actuellement.

[79] Le CASD retrace sur son site l'ensemble des projets par producteur et par source, et les publications afférentes quand elles sont signalées par les chercheurs (<https://www.casd.eu/projets-developpees-sur-le-casd/>).

ANNEXE 3 : Utilisation des données de la DSN à des fins de pilotage des politiques du travail, de l'emploi et de la formation professionnelle

ANNEXE 3 : UTILISATION DES DONNEES DE LA DSN A DES FINS DE PILOTAGE DES POLITIQUES DU TRAVAIL, DE L'EMPLOI ET DE LA FORMATION PROFESSIONNELLE	119
1 LES DONNEES ISSUES DE LA DSN CONTRIBUENT A AMELIORER LE PILOTAGE DES POLITIQUES PUBLIQUES SUR LE MARCHE DU TRAVAIL	122
1.1 L'IMPACT DES INTERVENTIONS PUBLIQUES SUR LE MARCHE DU TRAVAIL ET LEUR ADAPTATION A SES EVOLUTIONS FONT L'OBJET D'UNE ATTENTION CROISSANTE	122
1.1.1 <i>Le recours aux données administratives peut permettre de mieux adapter les politiques publiques aux besoins.....</i>	122
1.1.2 <i>Le recours aux données administratives répond aussi à une attente croissante de réactivité de la part des responsables publics</i>	123
1.2 LES DONNEES ISSUES DE LA DSN, DONT LA QUALITE EST INEGALE, CONSTITUENT UN GISEMENT ENCORE JEUNE D'INFORMATIONS RICHES ET MENSUALISEES	123
1.2.1 <i>L'extension progressive du périmètre de la DSN a accru l'intérêt de cette base de données</i>	123
1.2.2 <i>La qualité des données brutes issues de la DSN est encore globalement peu documentée..</i>	124
1.2.3 <i>Les utilisateurs des données issues de la DSN ont signalé à la mission des points d'attention en termes de qualité</i>	124
2 LES USAGES IDENTIFIES DE LA DSN A DES FINS DE PILOTAGE PEUVENT ETRE REGROUPES SELON LEUR OBJECTIF PRINCIPAL.....	126
2.1 ANALYSER LES PARCOURS EN EMPLOI DES BENEFICIAIRES DE DISPOSITIFS DE POLITIQUES PUBLIQUES ..	126
2.1.1 <i>Suivi de l'accès à l'emploi d'une cohorte d'allocataires du RSA</i>	126
2.1.2 <i>Suivi de l'accès à l'emploi d'une cohorte de bénéficiaires du CEJ.....</i>	127
2.1.3 <i>Mesure du taux de retour à l'emploi durable des demandeurs d'emploi</i>	127
2.1.4 <i>Suivi de l'accès à l'emploi des bénéficiaires de la formation professionnelle (« Agora »).....</i>	129
2.2 CIBLER LE CONTROLE ET L'ACCOMPAGNEMENT DES ENTREPRISES DANS LE RESPECT DES DISPOSITIONS PREVUES PAR LE CODE DU TRAVAIL	130
2.2.1 <i>Mise à disposition de tableaux de bord sur les entreprises contrôlées (« Suit » et « BCE ») ...</i>	130
2.2.2 <i>Ciblage des entreprises ayant un recours abusif aux contrats courts (« SI précarité » et « Champollion »).....</i>	133
2.2.3 <i>Ciblage des entreprises en fonction du taux et de la nature des accidents du travail (« Dat'IA »)</i>	134
2.3 IDENTIFIER LE PLUS EN AMONT POSSIBLE LES TENSIONS DE RECRUTEMENT ET ENJEUX D'ATTRACTIVITE.....	136
2.3.1 <i>Analyse de l'évolution des emplois et de la masse salariale dans le secteur sanitaire et médico-social (« Diamant »).....</i>	136
2.3.2 <i>Analyse d'attractivité des métiers du grand âge et de l'autonomie (« Ramsès »).....</i>	136

2.4	ANTICIPER ET ANALYSER L'ÉVOLUTION DES EMPLOIS ET DES COMPÉTENCES AU NIVEAU D'UNE ENTREPRISE, D'UN SECTEUR, D'UNE BRANCHE OU D'UN TERRITOIRE	137
2.4.1	Repérage des entreprises susceptibles de procéder à des licenciements (« Signaux faibles »)	137
2.4.2	Objectivation des désajustements entre besoins en compétences et qualifications produites (« Octopilot »).....	138
2.5	DEVELOPPER DES ACTIONS SUR LE MARCHÉ DU TRAVAIL CIBLÉES SUR CERTAINS PUBLICS REPUTÉS PLUS ÉLOIGNÉS DE L'EMPLOI	139
2.5.1	Repérage et accompagnement des entreprises ne respectant pas l'obligation d'emploi des travailleurs handicapés (« OETH »)	139
2.5.2	Repérage et accompagnement des entreprises assujetties au calcul de l'index égalité professionnelle (« Égapro »).....	141
2.6	ANTICIPER L'IMPACT ET SUIVRE LA MONTÉE EN CHARGE D'UN NOUVEAU DISPOSITIF DE POLITIQUE PUBLIQUE	142
2.6.1	Simulations dans le cadre de l'étude d'impact préalable à l'évolution d'une niche sociale ou à l'attribution de compléments de rémunération.....	142
2.6.2	Suivi de la montée en charge de la Prime exceptionnelle de pouvoir d'achat et de l'indemnité inflation.....	144
2.7	ÉCLAIRER LE MARCHÉ CACHE DE L'EMPLOI EN IDENTIFIANT LES ENTREPRISES SUSCEPTIBLES DE RECRUTER	144
2.7.1	Identifier les entreprises susceptibles de recruter sur un territoire et à un moment donné (« La bonne boîte »).....	144
2.7.2	Identifier les entreprises susceptibles de recruter en alternance (« La bonne alternance »).....	146
3	EN L'ABSENCE D'UNE GOUVERNANCE DE SES USAGES, LES DONNÉES ISSUES DE LA DSN RESTENT INSUFFISAMMENT VALORISÉES	148
3.1	LA CAPACITÉ À IDENTIFIER ET MOBILISER LES DONNÉES ISSUES DE LA DSN SUPPOSE DÉJÀ UN CERTAIN NIVEAU DE CONNAISSANCE ET D'ACCULTURATION	148
3.1.1	Le niveau de maturité des acteurs du champ travail-emploi-formation par rapport à l'utilisation des données est inégal.....	148
3.1.2	Les cas d'usage identifiés de production de tableaux de bord ou d'indicateurs se développent souvent en ordre dispersé.....	149
3.2	LE CARACTÈRE INCERTAIN DU CADRE JURIDIQUE ET LA DIFFICULTÉ À IDENTIFIER LES « GUICHETS » PERTINENTS POUR ACCÉDER AUX DONNÉES CONSTITUENT DES FREINS À L'UTILISATION DE LA DSN	150
3.2.1	S'il peut être contourné, le cadre juridique relatif à la DSN tend encore à brider son utilisation à des fins de pilotage.....	150
3.2.2	Sans fonction d'aiguillage, le choix de la base de données mobilisée relève davantage d'une logique aléatoire que de l'adaptation au besoin	151
3.3	L'ABSENCE D'ANIMATION DE LA COMMUNAUTÉ DES UTILISATEURS DE LA DSN NE FAVORISE PAS L'ÉMERGENCE DE NOUVEAUX CAS D'USAGE	152
3.3.1	Il n'existe pas de vision globale des différents usages de la DSN à des fins de pilotage ni de partage d'expérience entre les utilisateurs.....	152
3.3.2	Les utilisateurs des données n'accèdent pas systématiquement à l'accompagnement dont ils pourraient avoir besoin	153

[1] **Cette annexe établit une cartographie des différents usages de la DSN à des fins de pilotage dans les domaines emploi-formation-insertion.** Les cas d'usage identifiés traduisent une acception large de cette notion de pilotage qui ne se limite pas, selon les termes de la lettre de mission, aux « *principaux indicateurs produits sur produits sur la base de la DSN par les divers acteurs y ayant recours à des fins de (...) pilotage opérationnel* ». Ces cas d'usage incluent également les outils de pilotage permettant de cibler ou de personnaliser des actions. L'utilisation des données issues DSN à des fins de stricte gestion est exclue du périmètre de la mission.

[2] Dans l'étude de ces cas d'usage des données issues de la DSN à des fins de pilotage, la mission distingue deux catégories de données :

- d'une part, les **données agrégées**, c'est-à-dire des données qui ne sont pas déclinées à l'échelle d'un individu, même si elles peuvent être par ailleurs présentées à l'échelle d'un établissement, d'une zone géographique, d'un secteur ou encore d'une branche ;
- d'autre part, les **données individuelles**, c'est-à-dire des données relatives à des individus qui peuvent être soit (i) nominatives quand elles permettent de les identifier (y compris à travers le croisement de plusieurs informations), (ii) soit anonymisées (elles ne permettent pas d'identifier un individu ni de le suivre mais autorisent des usages très larges des données), (iii) soit pseudonymisées (elles ne permettent pas d'identifier un individu sans information supplémentaire mais elles autorisent le suivi des données relatives à cet individu dans le temps ou leur appariement avec d'autres données si l'identifiant est commun).

[3] Pour cartographier les usages des données issues de la DSN à des fins de pilotage, la mission a identifié le type de données DSN utilisées : données brutes issues directement de la déclaration, ou données retraitées issues notamment de la statistique publique (cf. annexe n°1).

[4] En l'absence de recensement systématique des différents usages des données de la DSN à des fins de pilotage, voire des acteurs qui accèdent à ces données sous une forme agrégée, la mission a identifié, par le biais des entretiens menés, différents cas d'usage puis les a regroupés en fonction de l'objectif principal recherché. La mission ne peut néanmoins pas garantir le caractère exhaustif de cette cartographie.

[5] La mission met aussi en évidence des besoins non satisfaits, voire des besoins latents et encore peu exprimés, d'accès aux données de la DSN à des fins de pilotage. La mission a élargi au fil de ses travaux le périmètre des acteurs sollicités (administrations du MTPEI, opérateurs du MTPEI, représentants des branches professionnelles, chercheurs, start-up d'État...) pour identifier ces besoins. L'ampleur de la sous-utilisation de la DSN à des fins de pilotage reste néanmoins difficile à objectiver, même si la mission a recueilli suffisamment d'indices pour poser un constat sur leur typologie.

1 Les données issues de la DSN contribuent à améliorer le pilotage des politiques publiques sur le marché du travail

1.1 L'impact des interventions publiques sur le marché du travail et leur adaptation à ses évolutions font l'objet d'une attention croissante

1.1.1 Le recours aux données administratives peut permettre de mieux adapter les politiques publiques aux besoins

[6] **Dans un contexte où le marché du travail connaît des mutations importantes, le recours à la donnée constitue un moyen d'adapter les politiques publiques aux besoins¹¹⁰.** Ainsi, la feuille de route « données »¹¹¹ du ministère identifie l'exploitation des données comme « *un élément clef dans l'aide au pilotage des politiques portées (...), qu'il s'agisse par exemple de produire un diagnostic contextualisé, de concevoir des plans de contrôle ou de faciliter l'exercice des activités de régulation* ». Sans dresser ici un panorama exhaustif des différents enjeux auxquels les politiques de l'emploi, du travail et de la formation professionnelle doivent répondre, le recours à la donnée, en particulier à la DSN, peut contribuer à objectiver :

- des marqueurs et facteurs explicatifs de tensions de recrutement, et plus globalement des désajustements entre demande et offre de travail, à caractère à la fois géographique et sectoriels ;
- les spécificités, sur le marché du travail, de certaines catégories d'actifs (femmes, seniors, travailleurs en situation de handicap, en particulier) ;
- l'effet de la mobilisation de dispositifs d'aide publique (contrats aidés, minima sociaux, formation professionnelle, dont l'alternance, accompagnement intensif des demandeurs d'emploi...) sur les parcours des bénéficiaires ;
- l'évolution de la qualité des emplois (nature et durée des contrats de travail, niveau relatif de rémunération, *turn-over*, taux d'absentéisme, sinistralité...) dans une branche, un secteur ou un métier donné ;
- les passerelles entre secteurs et métiers facilitées par l'existence de compétences transférables, et le cas échéant par la formation professionnelle.

¹¹⁰ Parmi ces tendances on peut citer, en particulier, le vieillissement de la population, la numérisation de l'économie et l'impact de la transition énergétique. Voir par exemple l'exercice de prospective sur les métiers et les qualifications réalisé par la Dares et France Stratégie, « *Les métiers en 2030* », mars 2022. Ce rapport souligne que, d'ici à 2030, les créations d'emploi seront très différentes d'un métier à l'autre avec des risques d'inadéquation entre les besoins de recrutements anticipés et le vivier potentiel des jeunes qui débutent leur carrière professionnelle.

¹¹¹ Feuille de route des données, des algorithmes et des codes sources du MTPEI datée du 22 septembre 2021.

1.1.2 Le recours aux données administratives répond aussi à une attente croissante de réactivité de la part des responsables publics

[7] **Parce que l'attention portée à l'évaluation de l'action publique s'accroît, et que le rythme des mutations s'accélère, le recours à la donnée répond aussi à une attente croissante de réactivité¹¹².** Les données se multiplient car la numérisation accroît le volume d'informations disponibles et leur vitesse de circulation. La création de France Travail répond en particulier à cet objectif d'améliorer en continu l'efficacité des interventions sur le marché du travail. Or le pilotage des politiques sur le marché du travail est resté pendant longtemps dépendant de dispositifs déclaratifs, en particulier ceux auxquels la DSN s'est substituée, qui étaient à l'origine de délais pouvant aller jusqu'à une trentaine de mois. Si l'exploitation des données administratives ne peut pas se substituer entièrement aux enquêtes et aux panels réalisés en particulier par le système statistique public, la DSN offre toutefois la possibilité de s'appuyer sur des données plus fraîches pour apprécier les effets des actions menées et, si nécessaire, de les réajuster, dans la mesure où elle produit des données fiables à un rythme mensuel. Néanmoins, certaines tendances de nature plus structurelle (par exemple, les réallocations d'emplois entre secteurs) et certains thèmes qui reposent sur une logique annuelle (en particulier les rémunérations) continueront toujours d'exiger un niveau minimum de recul temporel, posant une limite à la réactivité.

1.2 Les données issues de la DSN, dont la qualité est inégale, constituent un gisement encore jeune d'informations riches et mensualisées

1.2.1 L'extension progressive du périmètre de la DSN a accru l'intérêt de cette base de données

[8] **La DSN constitue un gisement encore jeune mais de plus en plus riche de données mensuelles sur le travail et l'emploi, couvrant la quasi-totalité des salariés.** Bien que créée en 2012, la DSN, pleinement déployée sur le champ des salariés du secteur privé depuis 2017, ne l'est dans le secteur public que depuis 2022. Le volume global de déclarations s'élève aujourd'hui à environ 40 M d'actes par mois¹¹³. Les informations véhiculées par la DSN se sont enrichies au fil de l'extension de son périmètre marquée, en 2021 par l'intégration des données associées à l'obligation d'emploi des travailleurs handicapés et, en 2022, par celle des contributions relatives à la formation professionnelle et à l'apprentissage. La DSN remplace ainsi aujourd'hui près d'une quarantaine de formalités administratives, contre une vingtaine au moment de sa création. Les particuliers-employeurs en restent à ce stade exclus, tout comme bien sûr les travailleurs indépendants. Ce périmètre des seuls salariés peut limiter l'intérêt de la DSN pour analyser une partie des évolutions sur le marché du travail, même si des appariements restent possibles avec des panels et des données administratives relatifs au travail indépendant.

¹¹² Voir en particulier le rapport de la mission Bothorel « *Pour une politique publique de la donnée* », décembre 2020.

¹¹³ Voir la feuille de route 2022-2025 du Gip-MDS.

1.2.2 La qualité des données brutes issues de la DSN est encore globalement peu documentée

[9] **Si elle tend à progresser, la qualité des données brutes issues de la DSN reste encore inégale, et surtout peu documentée, ce qui peut fragiliser son utilisation à des fins de pilotage.** L'objectif premier du Gip-MDS était à l'origine de maximiser le volume de déclarations enregistrées en limitant les contrôles bloquants, au détriment parfois de la qualité des informations véhiculées. Néanmoins, il estime que le taux global de conformité de la DSN à la norme s'élèverait aujourd'hui à 94 %¹¹⁴, ce qui traduirait une bonne appropriation du cahier technique et du processus par les déclarants. Dans un contexte où les informations issues de la DSN contribuent de plus en plus à déterminer les droits des assurés sociaux, l'accent mis depuis la LFSS pour 2020 sur la fiabilisation des données individuelles pourrait contribuer à améliorer ce taux de conformité¹¹⁵, même si les évolutions de la norme supposent de poursuivre en continu cette démarche. Le Gip-MDS anime des ateliers qualité réunissant un échantillon d'utilisateurs des données DSN. Toutefois, il ne dispose pas d'une analyse globale des zones de non-qualité de la DSN. Cette notion de qualité doit être appréhendée à l'aune des usages envisagés de ces données, ce qui suppose que les utilisateurs disposent d'une connaissance suffisante des zones de fragilité. Or les différents utilisateurs de la DSN rencontrés par la mission ont indiqué :

- soit ne pas être en mesure d'apprécier la qualité des données brutes utilisées à défaut d'une documentation ou d'un accompagnement en lien avec leur utilisation ;
- soit identifier des fragilités sans pour autant être en mesure de les corriger et sans qu'elles remettent en cause à leurs yeux la pertinence globale des analyses effectuées ;
- soit, en particulier quand ils sont issus du service statistique public, procéder à des retraitements visant à améliorer la robustesse de ces données en fonction de l'usage qu'ils en font, y compris à des fins de pilotage.

1.2.3 Les utilisateurs des données issues de la DSN ont signalé à la mission des points d'attention en termes de qualité

[10] **Sans qu'il s'agisse d'une analyse exhaustive de la qualité de la DSN, qui n'était pas dans le périmètre de ses travaux, la mission a relevé au cours de ses échanges avec les utilisateurs les points d'attention concernant :**

- le faible niveau de renseignement par les entreprises de la rubrique relative aux « dispositifs de politique publique » et à leurs compléments (entreprises adaptées, insertion par l'activité économique, contrats de professionnalisation, contrats aidés, immersions professionnelles, en particulier), souligné en particulier par la Délégation générale à l'emploi et à la formation professionnelle (DGEFP) et l'Agence de services et de paiement (ASP), qui pose des

¹¹⁴ Feuille de route 2022-2025 du Gip-MDS.

¹¹⁵ La LFSS pour 2020 prévoit la mise en œuvre d'une démarche de fiabilisation des données individuelles reposant sur l'envoi par les organismes en charge du recouvrement de comptes-rendus normalisés signalant les anomalies à corriger avec, à horizon 2024, la possibilité pour ces organismes, en cas d'inaction des déclarants, de se substituer

difficultés en termes de pilotage mais aussi de gestion (l'ASP étant alors obligée de solliciter les informations nécessaires au paiement des aides auprès des entreprises éligibles) ;

- la fréquence des erreurs dans le renseignement des professions et catégories socioprofessionnelles (PCS)¹¹⁶ qui amène l'Insee à effectuer des corrections¹¹⁷ dans le processus de production de la base Tous Salariés (BTS);
- la fréquence des erreurs dans la localisation des lieux de travail, et donc des emplois, avec, par exemple, des établissements déclarant l'ensemble des salariés au niveau d'une seule ville alors qu'ils devraient être répartis entre plusieurs sites, ce qui oblige l'Insee à réaliser des regroupements en se fondant sur d'autres informations ;
- la proportion significative de défauts de déclaration des modifications intervenues dans les contrats de travail (par exemple, transfert d'un contrat de travail à un autre établissement), en particulier des fins de contrat, qui font l'objet de retraitements par la Dares¹¹⁸, pour éviter de fausser les statistiques sur la rotation de la main-d'œuvre ;
- le caractère lacunaire de la déclaration des codes de convention collective, ou la déclaration de codes erronés, également redressés par l'Insee et la Dares en utilisant le code APE qui permet d'établir le secteur de l'entreprise¹¹⁹ ;
- les défauts de cohérence entre la nature du contrat de travail des salariés, leur niveau de rémunération et la quotité de temps de travail déclarés par les employeurs, ces données étant là encore redressées par la Dares et l'Insee¹²⁰ ;
- la fréquence des erreurs relatives à la déclaration des données nécessaires au calcul de l'Obligation d'emploi des travailleurs handicapés (OETH), attribuées par l'Agefiph à la fois au transfert du recouvrement aux Urssaf et à l'évolution des modalités de calcul prévue par la loi n°2018-771 du 25 septembre 2018.
- enfin, dans le secteur public, dont les données en DSN font l'objet d'une utilisation plus récente, un faible niveau de renseignement du grade des salariés et du Finess des établissements des champs sanitaire, médico-social et social¹²¹.

¹¹⁶ La nomenclature des professions et catégories socioprofessionnelles (PCS-ESE) est une nomenclature placée sous la responsabilité de l'Insee et utilisée en DSN par les employeurs pour codifier la profession de leurs salariés. Elle se présente sous la forme d'un code à quatre chiffres, la désagrégation à 4 chiffres correspondant à environ 400 métiers. [Nomenclatures des professions et catégories socioprofessionnelles des emplois salariés des employeurs privés et publics | Insee](#)

¹¹⁷ Le responsable du Département de l'emploi et des revenus d'activité (Dera) a indiqué à la mission corriger environ 20 % des données sur la PCS.

¹¹⁸ Selon la Dares, en moyenne, 200 000 contrats disparaissent chaque mois de la DSN, dont 110 000 CDI.

¹¹⁹ La Dares indique qu'en 2018, 182 990 établissements (employant plus de 2,3 M de salariés) n'avaient aucune convention collective déclarée dans leurs contrats salariés en DSN « brute ».

¹²⁰ Par exemple, selon la Dares, 10 % des contrats à temps complet sont déclarés avec moins de 21,67 heures mensuelles.

¹²¹ Finess (Fichier national des établissements sanitaires et sociaux) est le répertoire national de référence des personnes morales intervenant dans les domaines sanitaire, médico-social et social et de l'enseignement des professions sanitaires et sociales. Selon le prestataire Invénis, en charge du projet « Ramsès », le taux de renseignement du Finess en DSN serait de l'ordre de 30 %, ce qui complexifie la possibilité d'identifier en DSN un établissement.

2 Les usages identifiés de la DSN à des fins de pilotage peuvent être regroupés selon leur objectif principal

[11] Les usages de la DSN identifiés dans la cartographie ci-dessous correspondent soit à des usages effectifs, soit à des usages prévus à court terme et pour lesquels une démarche d'accès aux données de la DSN a déjà été entreprise.

2.1 Analyser les parcours en emploi des bénéficiaires de dispositifs de politiques publiques

2.1.1 Suivi de l'accès à l'emploi d'une cohorte d'allocataires du RSA

[12] Au sein du **Gip Plateforme de l'inclusion**¹²², une équipe dite « data.RSA » a mené en 2022 une expérimentation visant à construire un **indicateur de « taux de sortie positive » des allocataires du RSA vers l'emploi dans cinq départements**¹²³. Après avoir sollicité dans un premier temps le Gip-MDS et la DSS, et en l'absence de réponse de leur part, le Gip Plateforme de l'inclusion s'est tourné vers la Dnum des ministères sociaux afin d'accéder à des **données brutes, individuelles pseudonymisées**. Dans une « lettre de mission » datée du 5 juillet 2022 adressée à la Dnum, et en dehors de tout cadre conventionnel Dnum/Gip Plateforme, le Gip Plateforme de l'inclusion a précisé la chaîne de traitement de la donnée. Durant cette expérimentation qui s'est déroulée de février à août 2022, les cinq Conseils départementaux ont transmis chaque mois à la Dnum une liste avec le Nir (numéro de sécurité sociale) des bénéficiaires du RSA soumis à droits et devoirs. À partir de ces Nir, la Dnum a croisé ces données avec les informations issues de la DSN sur : la présence dans la DSN, la nature du contrat de travail, la quotité de travail et le code APE. Des données agrégées ont ensuite été restituées au Gip Plateforme de l'inclusion et aux départements concernés sous la forme d'indicateurs (cf. visuel ci-dessous).

[13] Le responsable du Gip Plateforme de l'inclusion a indiqué à la mission qu'il n'était pas envisagé à ce stade d'étendre l'expérimentation à d'autres départements, les départements expérimentateurs ne s'étant pas encore pleinement saisis de ces données. Néanmoins, la mise en place de France Travail et le déploiement des expérimentations relatives à l'accompagnement intensif de bénéficiaires du RSA dans 19 départements ou bassins d'emploi pilotes pourraient renforcer le besoin de suivi opérationnel des parcours d'emploi à travers les données de la DSN.

¹²² Créé en avril 2022, ce Gip vise à offrir aux personnes éloignées de l'emploi et à ceux qui les accompagnent un continuum d'informations depuis l'entrée dans le parcours d'insertion jusqu'à l'accès à l'emploi. Le groupement a notamment pour objet la meilleure circulation des données de parcours et un pilotage par l'impact des services numériques déployés par les acteurs de l'insertion et de l'emploi.

¹²³ Alpes-Maritimes, Bouches-du-Rhône, Hautes-Pyrénées, Seine-Saint-Denis, Val-d'Oise.

Schéma 1 : Extrait d'une analyse réalisée par le Gip Inclusion et la Dnum des sorties vers l'emploi d'une cohorte d'allocataires du RSA dans les Alpes-Maritimes

Premier croisement de données RSA et DSN pour les département (partenariat GIP & DNUM des Ministères sociaux)



Source : Gip Plateforme de l'inclusion

2.1.2 Suivi de l'accès à l'emploi d'une cohorte de bénéficiaires du CEJ

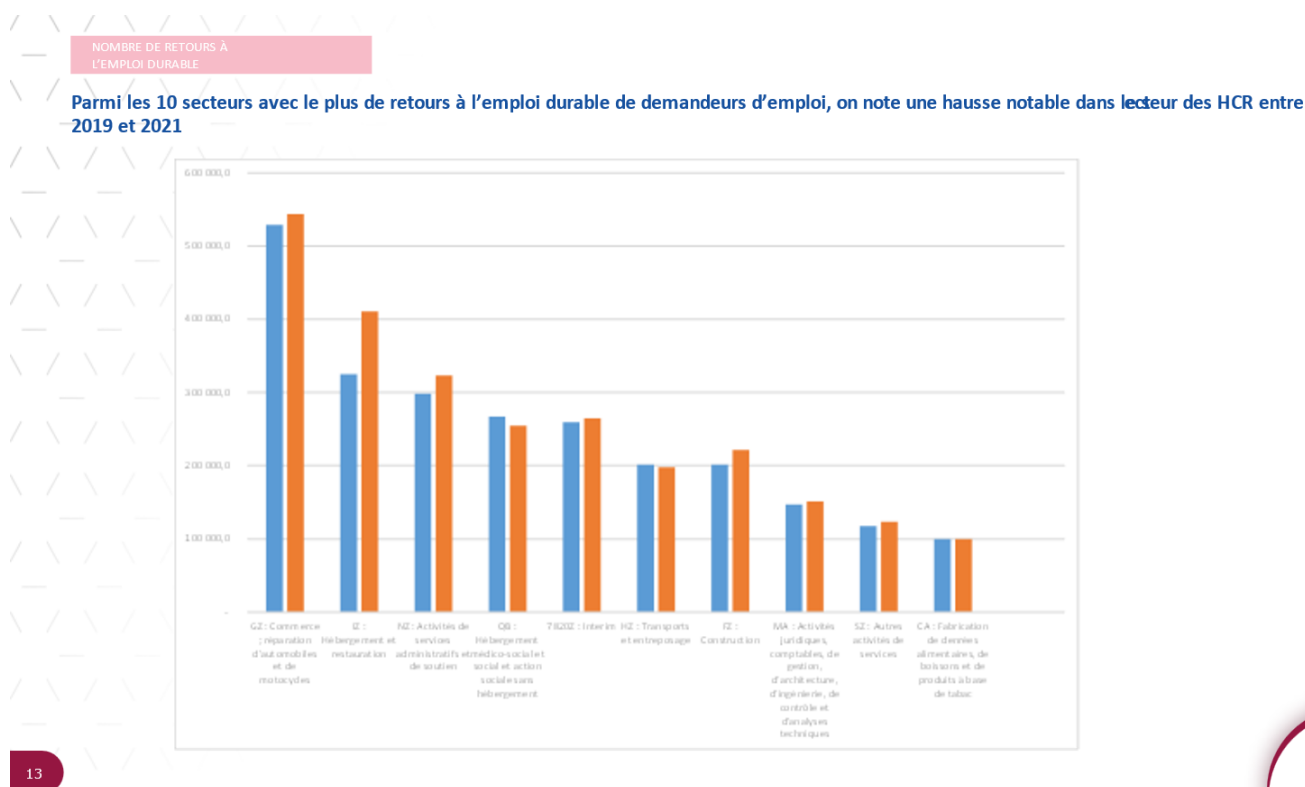
[14] La **DGEFP** a demandé au Gip-MDS d'apparier avec la **DSN des données relatives à des cohortes de jeunes en Contrat d'engagement jeunes (CEJ)**, dispositif entré en vigueur le **1^{er} mars 2022**. L'intervention du Gip-MDS a été validée en juillet 2022 par la DSS au regard des finalités prévues par le décret encadrant le traitement DSN. Depuis septembre 2022, les deux opérateurs du CEJ, à savoir le réseau des Missions locales et Pôle emploi, envoient au Gip un flux comprenant le Nir des jeunes en Cej actifs ou sortis de Cej pour lesquels la situation en emploi doit être vérifiée. En appariant ces données, dans un premier temps avec la base Dad-E puis avec celles, anonymisées cette fois, de la base PIQ, le Gip-MDS produira en retour des **données agrégées** sur la situation en DSN à 3, 6, 9 et 12 mois des bénéficiaires du CEJ, et à six mois après la date de sortie du dispositif. Pour éviter le risque de générer des agrégats trop fins et, potentiellement, d'identifier des individus, le Gip-MDS produit trois tables de sortie, sans croisement des différentes dimensions : un flux « secteur d'activité » ; un flux « type de contrat », un flux « pérennité de l'emploi ». La Dares qui développe d'autres outils d'appariements n'a pas été associée à cet exercice (pas même dans la conception des indicateurs de suivi). Cette démarche reste, à fin février 2022, en phase de test, une première livraison d'indicateurs étant prévue fin mars 2022.

2.1.3 Mesure du taux de retour à l'emploi durable des demandeurs d'emploi

[15] La **convention tripartite 2019-2022 signée entre Pôle emploi, l'État et l'Unédic** prévoyait la mesure du nombre de retours à l'emploi durable. Cette mesure devait s'appuyer, dans un premier temps, sur une enquête menée à un rythme semestriel et, à terme, sur un indicateur construit par les parties à la convention sur la base de la DSN. Le groupe de travail en charge de définir cet

indicateur a présenté ses conclusions le 30 novembre 2022. Il a proposé d'utiliser les données issues de la DSN, auxquelles Pôle emploi a accès sur une période de six mois suivant la désinscription du demandeur d'emploi, pour mesurer trois indicateurs : un indicateur calculant un taux d'accès à l'emploi durable d'une cohorte d'entrants sur les listes de demandeurs d'emploi ; un indicateur calculant un taux d'accès à l'emploi durable d'un « stock » de demandeurs d'emploi inscrits à une date donnée ; un indicateur donnant chaque mois un nombre d'accès à l'emploi durable des personnes inscrites sur les listes ou ayant quitté les listes depuis moins de six mois. L'emploi durable inclut a minima les CDD et missions d'intérim de six mois et plus, ainsi que les CDI, mais chaque indicateur est aussi décliné en incluant les contrats successifs chez le même employeur d'une durée plus courte, mais dont la durée cumulée dépasse six mois. Ces indicateurs sont mesurés au travers de la DSN, qui permet en particulier d'identifier les secteurs cumulant le plus de retours à l'emploi durable de demandeurs d'emploi (cf. graphique ci-dessous). Les créations d'entreprises sont également prises en compte via le Fichier historique (FH) des demandeurs d'emploi¹²⁴.

Graphique 1 : Extrait du support de présentation des travaux relatifs à la mesure du retour à l'emploi durable des demandeurs d'emploi



Source : Support du comité d'études tripartite du 30 novembre 2022 communiqué à la mission par Pôle emploi.

¹²⁴ Le Fichier historique de Pôle emploi est une base de données administratives recensant toutes les personnes ayant été inscrites à un moment ou à un autre à Pôle emploi sur les dix dernières années. On retrouve en particulier dans ce fichier le motif d'annulation d'une demande d'inscription et l'activité réduite.

2.1.4 Suivi de l'accès à l'emploi des bénéficiaires de la formation professionnelle (« Agora »)

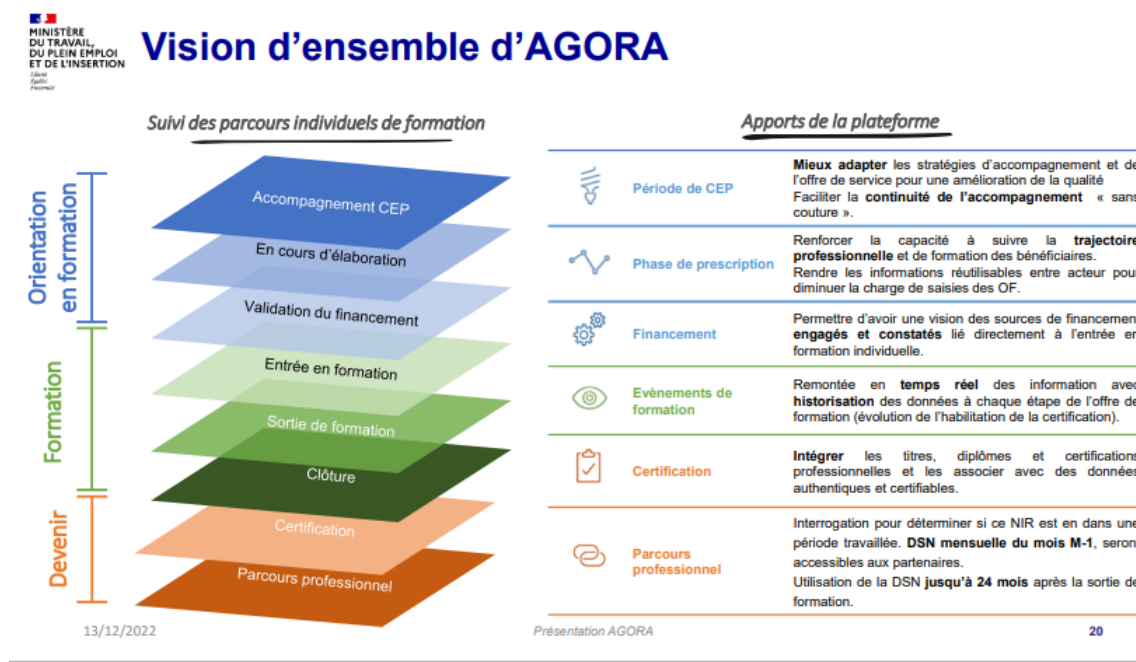
[16] La **Caisse des dépôts** est responsable, en lien avec la DGEFP, de la plateforme dite « Agora », présentée comme un « **data hub** » de la formation intégré au sein du système d'information relatif au Compte personnel de formation (CPF). Prévue par la loi n°2016-1088 du 8 août 2016, cette plateforme vise à favoriser le partage en temps réel de données entre les organismes financeurs de la formation professionnelle (Pôle emploi, Régions, Opco), la Caisse des dépôts et les institutions et organismes chargés du conseil en évolution professionnelle¹²⁵ (cf. schéma ci-dessous). Sa création s'est accompagnée de la construction d'un référentiel national permettant d'harmoniser le langage utilisé et de faciliter l'exploitation des données. Ouvert depuis le 29 mars 2021, le service « Agora » était selon la DGEFP utilisé début février 2022 par 37 acteurs ayant « accroché » leur SI avec la plateforme.

[17] L'utilisation de la DSN dans le cadre d'« Agora » se fonde sur des dispositions réglementaires qui ne sont pas spécifiques à cette plateforme, mais qui encadrent les données à caractère personnel enregistrées dans le SI CPF. Il s'agit en l'espèce des « *données relatives au parcours professionnel du titulaire* », sachant que la Caisse des dépôts n'est autorisée par l'arrêté « filtre » à utiliser les données de la DSN qu'à la seule fin d'« *alimenter le compte personnel de formation des salariés concernés* ». La DGEFP a adressé à la DSS le 21 septembre 2022 une demande d'ajustement des textes relatifs à la DSN pour autoriser un usage de ces données par la Caisse des dépôts plus large que celui prévu aujourd'hui. Elle reste en attente d'un retour de la part de la DSS.

[18] Les données de la DSN accessibles via « Agora » sont des **données brutes, individuelles et pseudonymisées**. À partir du Nir d'un stagiaire de la formation professionnelle, chaque organisme financeur peut, par un système d'API, « appeler » des données sur son parcours professionnel en sortie de formation sur une durée de 24 mois. Si le stagiaire a accédé à un emploi dans le périmètre de la DSN, la plateforme Agora renseigne en particulier sur la date de début de contrat, sa date de fin, sa nature, sa quotité horaire, l'activité principale de l'entreprise concernée ou la présence d'un éventuel dispositif de politique publique (ex : contrat aidé). Le recours à ces données a vocation, à terme, à se substituer aux enquêtes réalisées par les organismes financeurs sur le devenir des stagiaires à trois, six et douze mois.

¹²⁵ Voir l'article L. 6353-10 du code du Travail.

Schéma 2 : Extrait du support de présentation d' « Agora » daté du 13 décembre 2022



Source : DGEFP.

2.2 Cibler le contrôle et l'accompagnement des entreprises dans le respect des dispositions prévues par le code du Travail

2.2.1 Mise à disposition de tableaux de bord sur les entreprises contrôlées (« Suit » et « BCE »)

2.2.1.1 « Suit »

[19] La **DGT** a mis en service en 2015 un système d'information du système d'inspection du travail (dit « Wiki'T ») en vue de regrouper les données utiles à l'exercice de ses missions. Ce système d'information, en cours de refonte et renommé « Suit », intègre depuis 2018 les données relatives aux mouvements de main d'œuvre issues de la DSN¹²⁶. Ces données mensuelles, **agrégées au niveau d'un établissement**, sont transmises par la Dares à la DGT, après un **retraitement** effectué dans Sismmo, à M+3 par rapport aux déclarations en DSN. Elles permettent aux membres de l'inspection du travail d'identifier, dans le cadre d'un tableau de bord sur une entreprise ou un établissement, le volume total des effectifs, décomposable par genre et par nature de contrat (CDD, CDI, intérim).

¹²⁶ Ce SI intègre aussi des données sur les déclarations d'accidents de travail, les procès-verbaux d'élection en comité social et économique d'entreprise et l'ensemble des saisies réalisées par les membres de l'inspection du travail dans le cadre de leurs contrôles. Ce traitement de données à caractère personnel est encadré par le décret n°2022-771 du 2 mai 2022.

Schéma 3 : Copie écran d'une recherche effectuée sur l'outil « Suit » avec focus sur les données issues de la DSN relatives aux effectifs

SUIT | Recherche multicritères | Recherche multicritères disponibles | Nouvelle recherche

Ettablissement [REDACTED]

FICHE SYNTHÈSE | INTERVENTIONS | SUITES | DOSSIERS D'AFFAIRE | DOCUMENTS DE L'ETABLISSEMENT

Status: **ACTIF** | Dernier contrôle le: 07/06/2022 | Service: Unité de contrôle n°2 de l'Essonne | Section: 0 | Date création: 24/12/2020

IMPRIMER LA FICHE SYNTHÈSE | PARTAGER

Identité | Source: SIRENE

Ettablissement

Siret: [REDACTED] | Raison sociale: [REDACTED] | Régime: Travail

Code NAF: 78.20Z | Libellé de l'activité: Activités des agences de travail temporaire | Nom de l'enseigne: [REDACTED]

Adresse: [REDACTED]

0 SUITES | FONCTIONNALITÉ À VENIR | FONCTIONNALITÉ À VENIR

FONCTIONNALITÉ À VENIR | FONCTIONNALITÉ À VENIR | FONCTIONNALITÉ À VENIR

TOIT DÉVELOPPER

Données générales (circled in red)

Données générales | TOUT DÉVELOPPER

Effectifs | Source : DSN

Dernier mois	Effectif total
12-2022	1220
Femmes	Hommes
420	800
Total CDI	Total CDD
34	14
CDI Intérimaire	Intérimaires envoyé en mission
113	1059
	Intérimaires mis à disposition
	-

ACCÉDER À L'HISTORIQUE

Source : DGT.

2.2.1.2 Base de connaissance des entreprises (BCE)

[20] La **Dnum** a créé en mars 2022 une plateforme dite « **Base de connaissance des entreprises** » (**BCE**) visant à simplifier l'accès aux données des entreprises et leur exploitation par les services du ministère du Travail. Les données de BCE sont consultables à partir du portail FCE (Fiche commune entreprises) par les agents habilités (cf. copie écran ci-dessous). Ces données, qui incluent en particulier des **données brutes agrégées par entreprise ou établissement** sur les effectifs physiques, les ETP et le code de convention collective issus de la DSN¹²⁷ (avec leur évolution dans le temps), permettent de cibler les entreprises sur des critères comme leur taille, leur secteur d'activité, la baisse de leur chiffre d'affaires ou encore l'absence de visite par les services de la Dreets depuis un an.

Graphique 2 : Copie écran d'une recherche effectuée sur BCE

The screenshot displays the 'Fiche Commune Entreprise' (FCE) page for a company. The header includes the French flag, the text 'Fiche Commune Entreprise', and navigation buttons: '← Liste des résultats' and '+ Nouvelle recherche'. A yellow banner states 'Ce site est en beta-test. Aidez-nous à l'améliorer en donnant votre avis' with an 'Imprimer' button.

The main content area is divided into several sections:

- Left sidebar:**
 - Company description: 'Commerce de détail de livres en magasin spécialisé'.
 - Buttons: '→ Voir la fiche entreprise' and 'Aucun établissement'.
 - Address: 'Siège social', '851 955 088 00014', '93 - BAGNOLET'.
- Top right:** 'Siège Social' and 'SIRET' (partially redacted).
- Open since:** 'Ouvert depuis le 17/06/2019'.
- Activity code:** '47.81Z commerce de détail de livres en magasin spécialisé'.
- Key indicators:**
 - Effectif:** 5
 - Visites:** Pas d'intervention connue
 - Mut Eco:** Activité partielle
 - Aides:** Oui
- Bottom section: 'Activité'**
 - Lien de succession:** Source: Insee / Sirene Sirene 2023
 - Effectifs:**
 - Tranche Effectif INSEE:** 3 à 5 salariés (Source: Sirene 2020)
 - Effectif physique:** 5 (Source: DNUM / DSN novembre 2022)
 - Buttons:** 'Afficher le détail et l'historique des effectifs'
 - ETP data:**
 - Effectif ETP décembre 2022:** 3,34 (Source: Gcp-Mds / DSN décembre 2022)
 - Effectif ETP novembre 2022:** 3 (Source: Gcp-Mds / DSN novembre 2022)
 - Effectif ETP octobre 2022:** 2,37 (Source: Gcp-Mds / DSN octobre 2022)

Source : Accès accordé à la mission par la Dnum, recherche effectuée le 15 février 2023.

¹²⁷ Outre ces données issues de la DSN, les principales données regroupées dans BCE sont issues de la base Sirene de l'Insee, des visites et contrôles effectués par les Dreets (bases EOS, SORA et Wiki'T), de la base DACCORD sur les accords signés par les entreprises, des bases relatives aux aides et agréments (ex : base APART pour l'activité partielle) et aux accidents du travail.

2.2.2 Ciblage des entreprises ayant un recours abusif aux contrats courts (« SI précarité » et « Champollion »)

2.2.2.1 « Si précarité »

[21] La **DGT** a mis en service en septembre 2021 un outil dit « SI précarité » à destination de l'inspection du travail pour cibler dans leurs contrôles les entreprises présentant une part élevée de contrats précaires. Un contrat précaire est défini dans l'outil comme un CDD ou une mission d'intérim. Les données issues de la DSN correspondent à des **données agrégées à l'échelle d'une entreprise retraitées par la Dares** dans Sismmo. Elles sont transmises à M+3 par rapport à la déclaration en DSN. Comme le montre la copie écran ci-dessous, l'outil donne la possibilité de réaliser différents filtres (par région, département, niveau d'effectif souhaité...) et de calculer plusieurs indicateurs pour apprécier le taux de recours aux contrats précaires en fonction de la taille de l'entreprise, dans le temps et par rapport au secteur d'activité ou à la région d'implantation.

Graphique 3 : Copie écran du « SI précarité »


**MINISTÈRE
DU TRAVAIL,
DU PLEIN EMPLOI
ET DE L'INSERTION**
Service
Région Île-de-France

Recours aux contrats précaires par les établissements Île-de-France de juin 2021 à juillet 2022

1

Effectif moyen

0,00

16.205,93

[Afficher le tableau détaillé](#)

Pointez l'angle supérieur droit du tableau pour afficher ce bouton puis cliquez sur les 3 petits points pour exporter les données

...

SIRET	RAISON SOCIALE	NUM. VOIE	TYPE VOIE	LIB. VOIE	CODE POSTAL	COMMUNE	CODE NAF	LIBELLE NAF	TK RECOURS MIN	TK RECOURS MOY	TK RECOURS MAX	EFF. MOYEN	TK RECOURS MOY NAF	TK RECOURS MOY REG
1	MAP	NICKPACRE NERCE	91290	PREMEY	91290	PREMEY	2599A	Fabrication d'articles métalliques...	0,00	0,00	0,00	1,00	17,99	0,00
	NON RENSEIGNE	91000	ST DENIS	91000	ST DENIS	91000	8011C	Extraction de gemmes ornementales...	0,00	0,00	0,00	0,00	7,75	0,75
14	RUE	BARRET DE JOUR	75007	PARIS 07	75007	PARIS 07	8010Z	Transports maritimes et côtiers d...	0,00	0,00	0,00	0,00	4,07	0,00
25	RD	SUCHET	75016	PARIS 16	75016	PARIS 16	8010Z	Transports maritimes et côtiers d...	0,00	0,00	0,00	0,00	4,07	0,00
17	RUE	VANNEAU	75007	PARIS 07	75007	PARIS 07	8010Z	Transports maritimes et côtiers d...	0,00	0,00	0,00	1,57	4,07	0,00
16	RUE	DES GOUVERNEURS	77100	MUSY ST GEORGES	77100	MUSY ST GEORGES	8010Z	Transports maritimes et côtiers d...	0,00	0,00	0,00	0,00	4,07	0,00
140	RUE	DE BILLANCOURT	91200	BOULOGNE BILLANCOURT	91200	BOULOGNE BILLANCOURT	7010Z	Activités des sièges sociaux	0,00	0,00	0,00	0,00	14,87	4,81
	NON RENSEIGNE	77140	NANDOUX	9221Z	Services auxiliaires des transports...	0,00	1,00	2,41	79,90	0,41	0,40			
	NON RENSEIGNE	77140	BARBESON	9221Z	Services auxiliaires des transports...	0,00	0,18	22,23	11,87	0,41	0,40			
	SAINT DE VAL DE LOIRE	77140	SOLPES SUR LOIRE	9221Z	Services auxiliaires des transports...	0,00	0,00	0,00	20,41	0,41	0,40			
	DISTRICT DE MAROLLES	77100	MAROLLES SUR SEINE	9221Z	Services auxiliaires des transports...	0,00	0,00	0,00	0,26	0,41	0,40			
	DISTRICT DES BRUNES	77100	MONTEBAU SUR LE JARD	9221Z	Services auxiliaires des transports...	0,00	0,00	0,00	41,14	0,41	0,40			
11	PL	DE L'EUROPE	78140	VILLIERS VILLANCOUVER	78140	VILLIERS VILLANCOUVER	9221Z	Services auxiliaires des transports...	0,00	0,00	0,00	2,16	0,41	0,40
10	RD	ORNAHO	91000	ST DENIS	91000	ST DENIS	4441Z	Commerce de gros commerce de...	2,31	3,96	6,02	190,89	0,40	0,40
10	RUE	NEULLET	75017	PARIS 17	75017	PARIS 17	8010A	Administration d'immeubles et s...	0,00	0,00	0,00	0,00	4,04	0,16
44	RUE	LEON GARNIER	92110	CLICHY	92110	CLICHY	2811Z	Fabrication de verre creux	0,00	11,23	40,00	10,87	10,02	11,09
2	RUE	DU PARVIS	77100	MONTEBAU SUR SEINE	77100	MONTEBAU SUR SEINE	9010Z	Activités comptables	0,00	0,00	0,00	19,51	1,86	1,26
20	RUE	BEZOUT	77140	NANDOUX	9221Z	Services auxiliaires des transports...	0,00	0,00	0,00	4,16	1,86	1,26		
9	RD	BARRET VAL DE LOIRE	77100	PARIS 17	77100	PARIS 17	8010F	Activités comptables	0,00	0,00	0,00	1,00	0,00	1,10

500 917 établissement (s)

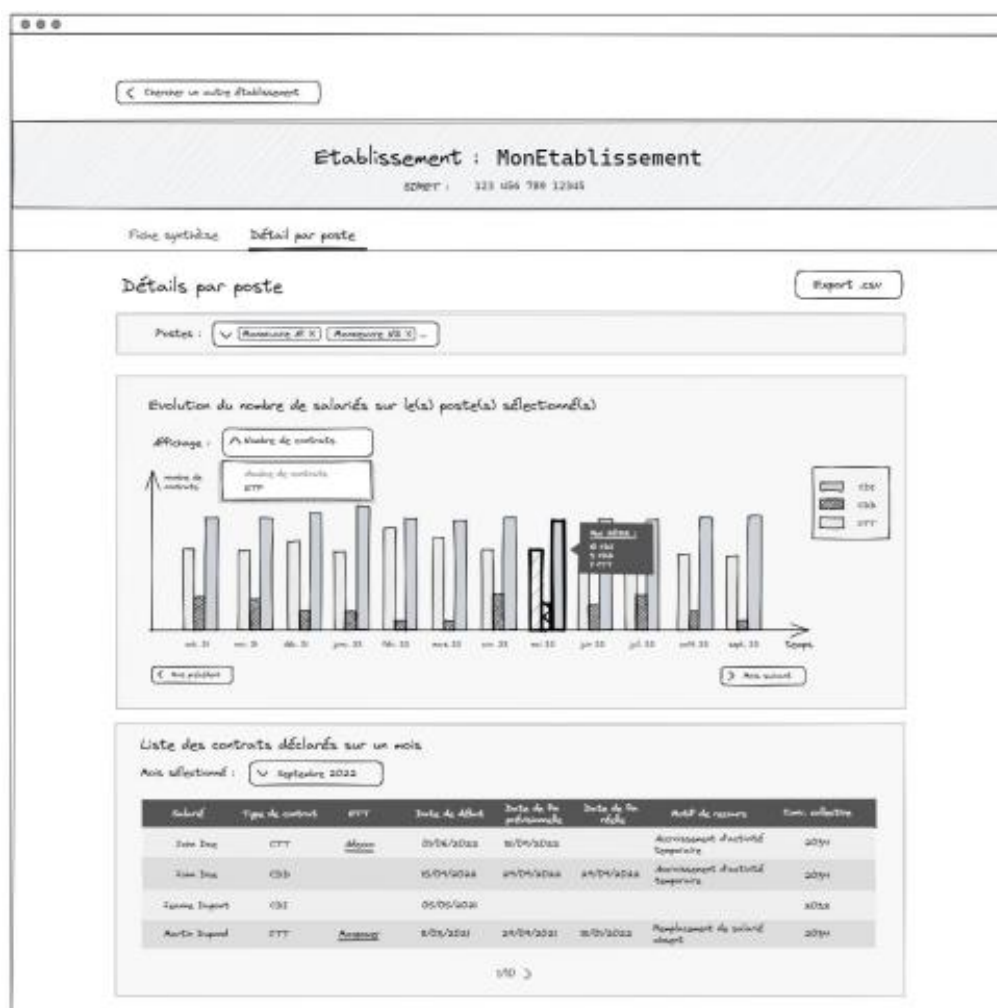
Source : DGT.

2.2.2.2 Projet « Champollion »

[22] La **Dnum**, en lien avec la DGT et la Dares, élabore un outil pour l'inspection du travail mobilisant cette fois les **données individuelles et nominatives de la DSN issues de Sismmo**. Ce projet est soutenu par Étalab qui finance partiellement trois Entrepreneurs d'intérêt général (EIG) mis à disposition de la Dnum pendant douze mois depuis septembre 2022. Selon la Dnum, l'objectif premier de ce projet, inscrit dans la feuille de route « données » du ministère en charge du Travail, est de prouver la valeur des données issues de la DSN en les rendant intelligibles par

les utilisateurs, en l'espèce les services de l'inspection du travail. Le cas d'usage identifié concerne, comme le « SI précarité », la lutte contre le recours abusif aux contrats courts. Cependant, contrairement au SI précarité, ce traitement devrait mobiliser un algorithme susceptible d'apprécier la probabilité d'un recours abusif aux contrats courts. Grâce aux données individuelles et nominatives (qui devraient être, selon la Dnum, anonymisées ou pseudonymisées), il permettrait de descendre au niveau du poste de travail pour, par exemple, apprécier le respect du délai de carence entre deux CDD (cf. maquette provisoire ci-dessous).

Graphique 4 : Vue de la maquette de l'appliquatif « Champollion » au 9 décembre 2022



Source : Dnum, support du Copil « Champollion ».

2.2.3 Ciblage des entreprises en fonction du taux et de la nature des accidents du travail (« Dat'IA »)

[23] La **DGT**, en lien avec la Dnum, a créé en 2021 un outil dit « Dat'IA » permettant de mieux visualiser les données relatives aux accidents du travail, issues des déclarations transmises par les entreprises et numérisées par la Cnam, et de calculer un indice de fréquence des accidents du travail par secteur d'activité et par établissement. Les **données d'effectifs agrégées par établissement et brutes issues de la DSN**, mises à disposition par la Dnum, sont utilisées dans l'outil pour rapporter le nombre d'accidents du travail recensés aux effectifs par année (cf. copies

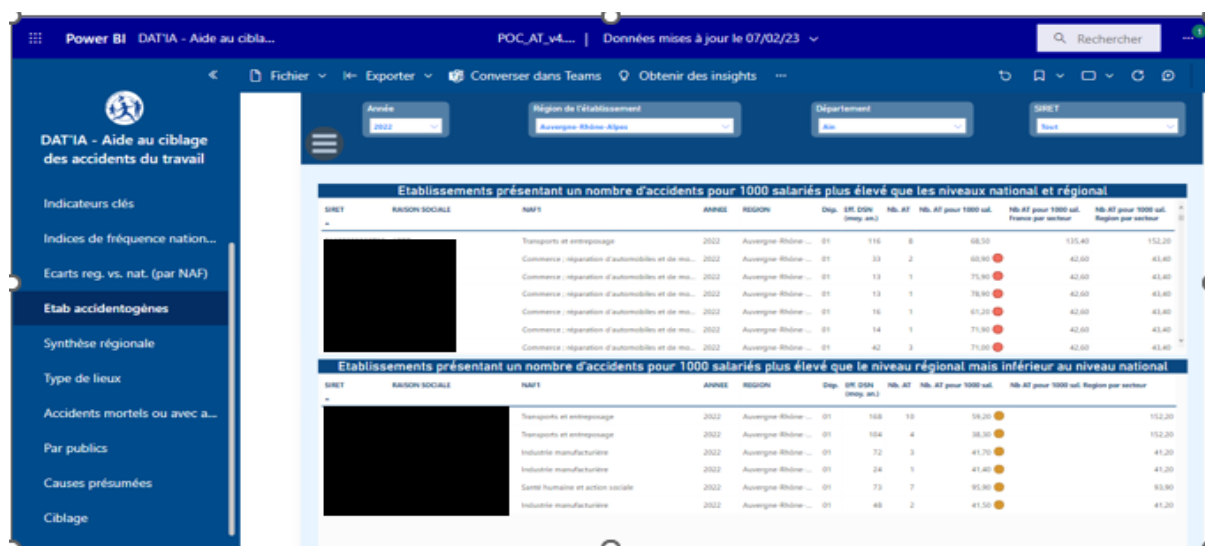
d'écran ci-dessous), et comparer ce taux à celui d'entreprises du même secteur ou de la même région. Par ailleurs, l'outil intègre un outil de reconnaissance textuelle¹²⁸. Il s'agit d'un outil permettant de reclasser automatiquement des déclarations d'accidents du travail par principale cause (travail en hauteur, chutes, équipement de travail mobile...), cette cause étant extrapolée à partir de l'analyse textuelle des déclarations. Dans leur ensemble, ces données doivent permettre à l'inspection du travail de cibler dans leurs contrôles les établissements les plus accidentogènes en repérant plus facilement les écarts à la moyenne. Cet outil devrait être déployé en 2023.

Graphique 5 : Copie écran de la vue « synthèse régionale » de l'outil « Dat'IA » au 15 février 2022



Source : Dat'IA, accès accordé à la mission par la DGT.

Graphique 6 : Copie écran de la vue « établissements accidentogènes » de l'outil « Dat'IA » au 15 février 2022



Source : Dat'IA, accès accordé à la mission par la DGT.

¹²⁸ Selon la DGT, l'algorithme, en cours de rodage et d'amélioration, serait fiable à 70 %.

2.3 Identifier le plus en amont possible les tensions de recrutement et enjeux d'attractivité

2.3.1 Analyse de l'évolution des emplois et de la masse salariale dans le secteur sanitaire et médico-social (« Diamant »)

[24] La **Dnum**, et en son sein le **service à compétence nationale en charge des SI des Agences régionales de santé (ARS)**, développe, en lien avec la Dares et la Direction générale de l'offre de soins (DGOS), un cube¹²⁹ DSN dans le SI « Diamant ». « Diamant » est un outil décisionnel de pilotage par la donnée centrée sur la régulation de l'offre de soins. Il est utilisé par les ARS, la DGOS, la Haute autorité de santé (HAS) et les 135 établissements supports de Groupements hospitaliers de territoire ». L'objectif de ce cube¹³⁰, qui devrait être mis en service en juin 2023, est d'utiliser les **données mensuelles à M+1 agrégées de la DSN, retraitées par la Dares**, pour mesurer, et si possible anticiper, des tensions sur le plan des ressources humaines au niveau d'un établissement du champ sanitaire ou médico-social ainsi que dans les maisons et centres de santé. Ces tensions pourront être appréciées sur la base de données issues de la DSN telles que l'évolution du volume d'heures supplémentaires, de l'absentéisme ou encore du taux de *turn-over*. Les facteurs d'attractivité des établissements (par exemple, l'impact de la mise en œuvre d'une nouvelle prime ou encore les différences de niveau et structure de rémunération) pourront aussi être mieux appréciés sachant que jusqu'ici, les données utilisées par les ARS étaient celles du bilan social annuel des établissements, moins fines et surtout moins fraîches. Une formation est prévue pour accompagner le déploiement, avec notamment une alerte sur la prudence à avoir vis-à-vis des données ayant un recul de moins de trois mois.

2.3.2 Analyse d'attractivité des métiers du grand âge et de l'autonomie (« Ramsès »)

[25] Si la **DGCS** utilise à ce jour principalement l'outil « Ramsès » pour simuler l'impact des accords de branche sur l'évolution de la masse salariale des services et établissements du secteur social et médico-social à but non lucratif, elle souhaiterait, du fait de l'intégration des services et établissements du secteur public en DSN, aussi utiliser ces données pour apprécier les leviers d'attractivité des métiers du grand âge et de l'autonomie. Toutefois, **elle estime avoir besoin à cette fin de données individuelles anonymisées** permettant en particulier d'identifier les structures de rémunération par secteur/type d'employeur/convention collective, ce qu'elle indique ne pas pouvoir faire avec le seul total des heures rémunérées. Par ailleurs, elle a adressé à la DSS une demande d'accès à des données supplémentaires auxquelles elle n'a pas aujourd'hui droit dans le cadre de l'arrêté « filtre ». Selon elle, ces données lui permettraient de calculer un taux de maintien dans l'emploi, un taux d'absentéisme, une durée moyenne de contrats ou encore d'apprécier le recours aux dispositifs de politique publique (contrats aidés, alternance...).

¹²⁹ Un « cube » est une vue restreinte des données dans un SI.

¹³⁰ Le développement de ce cube bénéficie d'un financement européen dans le cadre de l'appel à projets ITN5 (cycle de vie de la donnée) et d'un financement du plan de relance.

2.4 Anticiper et analyser l'évolution des emplois et des compétences au niveau d'une entreprise, d'un secteur, d'une branche ou d'un territoire

2.4.1 Repérage des entreprises susceptibles de procéder à des licenciements (« Signaux faibles »)

[26] La **Direction générale des entreprises (DGE)** a indiqué à la mission avoir obtenu en 2022, de la part de la Dares et avec le soutien de la direction de cabinet de la ministre en charge du Travail, un jeu de **données agrégées de la DSN, issues de Sismmo**, qu'elle souhaiterait tester, puis éventuellement utiliser dans le cadre de la start-up d'État « Signaux faibles ». Lancée au niveau régional en 2015, et élargie au niveau national en 2018, « Signaux faibles » repose sur un partenariat entre plusieurs administrations (DGE, DGEFP, Banque de France, Urssaf Caisse nationale, Direction générale des finances publiques – DGFIP) qui mettent en commun leurs données et leurs expertises pour identifier les entreprises fragiles et cibler l'action publique d'accompagnement (cf. copie écran ci-dessous). Ces données permettent à un algorithme de dessiner le profil statistique des entreprises fragiles et de prédire une entrée en procédure collective. Dans ce cadre, selon la DGE, les données sur les rotations de main d'œuvre issues de la DSN, si elles montrent par exemple une augmentation du volume de départs de salariés, que ce soit dans le cadre de licenciements ou de démissions, pourraient permettre d'identifier plus rapidement une éventuelle fragilité économique.

Graphique 7 : Copie écran de « Signaux faibles » avec les données relatives à une liste anonymisée d'entreprises

The screenshot shows the 'Signaux-Faibles' web application. The main table displays a list of companies with their SIREN numbers, activity codes, and financial data. The table is titled 'Détection - Septembre 2021'. The left sidebar contains navigation options: Consultation, Listes de détection, Suivi d'établissements, Contribuer, and Se déconnecter. The right sidebar contains filters for 'Secteur d'activité', 'Hors secteurs COVID-19', 'Zone géographique', 'Chiffre d'affaires min', 'Effectif min de l'entreprise', and 'Visibilité selon statut des procédures'. The main table has columns for company name, SIREN, activity code, CA 2019, EBE, and REX. The data is as follows:

Company Name	SIREN	Activity Code	CA 2019	EBE	REX
RHXXZWLMEPFZVQWSDWY	703413174 31029 (sage)	703 - Commerce de gros	6858 / 7052	1223973 k€	23264 k€ -2225 k€
THIZRMWEBGBNXXQCOIGP	70843827 17711 (sage)	703 - Commerce de gros	446 / 446	61600 k€	-3864 k€ -6111 k€
HQTFKEBHNTJMPSALEVBK	81720224 74511 (sage)	703 - Commerce de gros	697 / 697	40882 k€	3300 k€ 3704 k€
BWPYKNOMIYHRSWEQFDP	86150877 75788 (sage)	703 - Commerce de gros	704 / 704	158457 k€	4285 k€ 5753 k€
RETXLUXJTCWEKJKEOC	84502448 89023 (sage)	703 - Commerce de gros	230 / 230	37538 k€	-2818 k€ -1517 k€
YCUJWDFMLYKPGVBPNC	14010084 80334 (sage)	703 - Commerce de gros	394 / 641	53025 k€	396 k€ 692 k€
PSUSXWYGBFNYHNTKQ	50843055 17801 (sage)	703 - Commerce de gros	2540 / 8740	1270975 k€	60407 k€ 86716 k€

20/06/2022

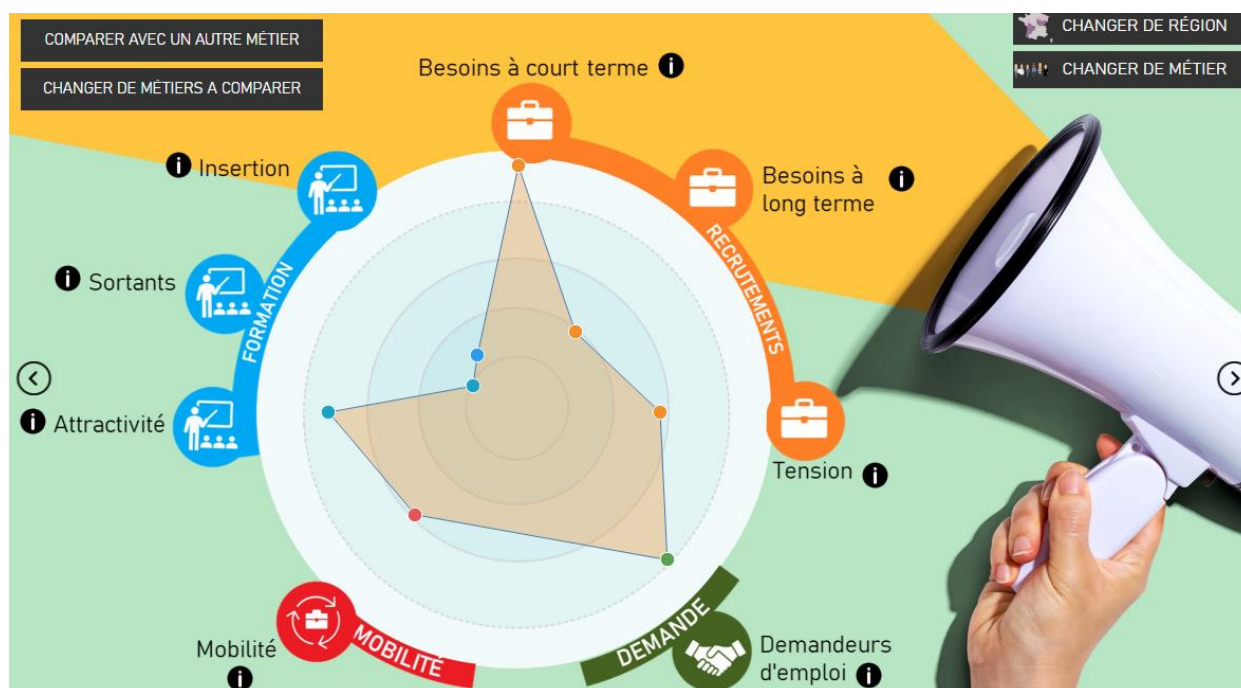
4

Source : DGE.

2.4.2 Objectivation des désajustements entre besoins en compétences et qualifications produites (« Octopilot »)

[27] La **tête de réseau des Carif-Oref (RCO)**, qui assure en particulier la mise à disposition des données nécessaires à l'exécution des missions Carif-Oref¹³¹, a accès depuis 2021 des **données individuelles pseudonymisées de la DSN issues des bases « tous salariés » retraitées (postes et salariés) hébergées au CASD (cf. annexe n°2)**. Ces données sont utilisées dans le cadre de la plateforme « Octopilot » mise à disposition des acheteurs de formation. La plateforme met en relation les données issues de la DSN avec celles relatives à la formation initiale et à la formation professionnelle, des données sur les demandeurs d'emploi mais également des données prospectives fournies par France Stratégie, pour objectiver les éventuels désajustements, par région et par métier, entre les besoins en compétences et les qualifications produites (cf. copie écran ci-dessous). Les données issues de la DSN sont aussi utilisées dans cet outil pour apprécier les passerelles entre métiers, et donc le potentiel de mobilité professionnelle, en examinant, pour un métier donné, les métiers qui étaient exercés par les salariés avant et ceux vers lesquels évoluent ces salariés après.

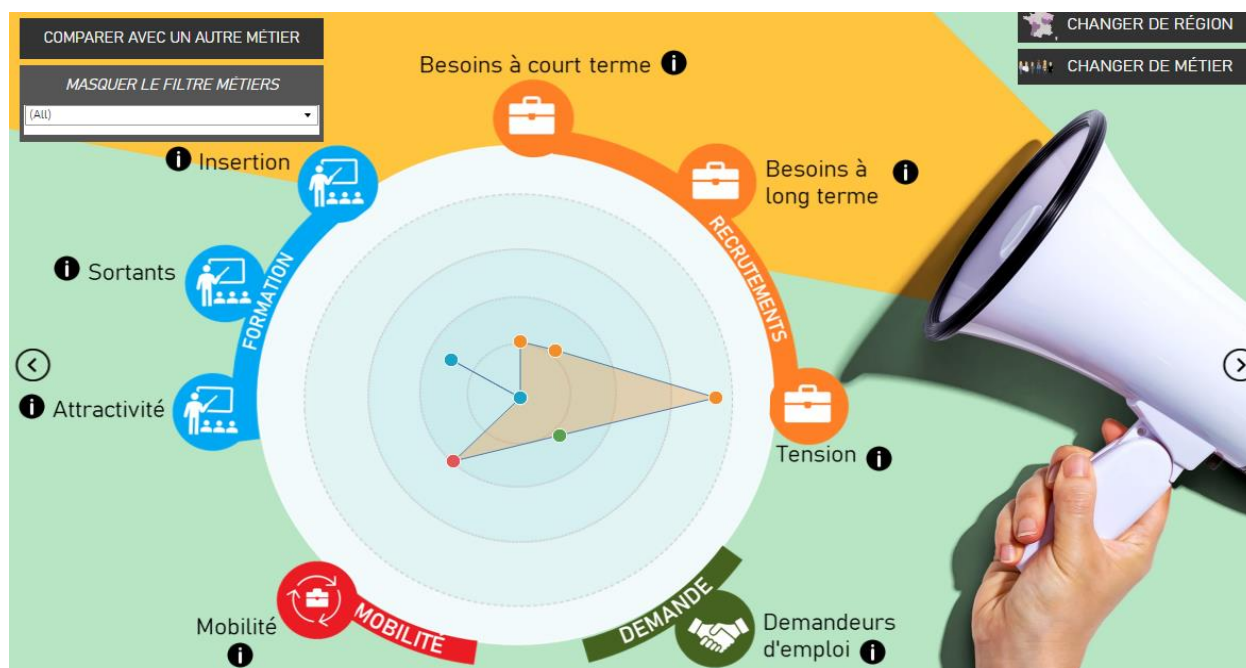
Graphique 8 : Vue des résultats dans « Octopilot » de la recherche pour le métier « agents d'entretien » dans la région Pays-de-la-Loire



Source : Accès à Octopilot accordé à la mission par RCO. Plus la toile est proche du cercle, plus la variable examinée est importante.

¹³¹ Structures partenariales portées par l'État et la région de type associatif ou Gip, les Centres animation ressources d'information sur la formation-Observatoires régionaux emploi formation (Carif-Oref) assurent une fonction de centres de ressources en charge de missions d'information, d'observation et d'analyse, et de professionnalisation des acteurs de la formation. Les missions de leur tête de réseau (RCO) ont été précisées par le décret 2021-792 du 22 juin 2021.

Graphique 9 : Vue des résultats dans « Octopilot » de la recherche pour le métier « cadres de la banque et des assurances » dans la région Pays-de-la-Loire



Source : Accès à Octopilot accordé à la mission par RCO. Plus la toile est proche du cercle, plus la variable examinée est importante.

2.5 Développer des actions sur le marché du travail ciblées sur certains publics réputés plus éloignés de l'emploi

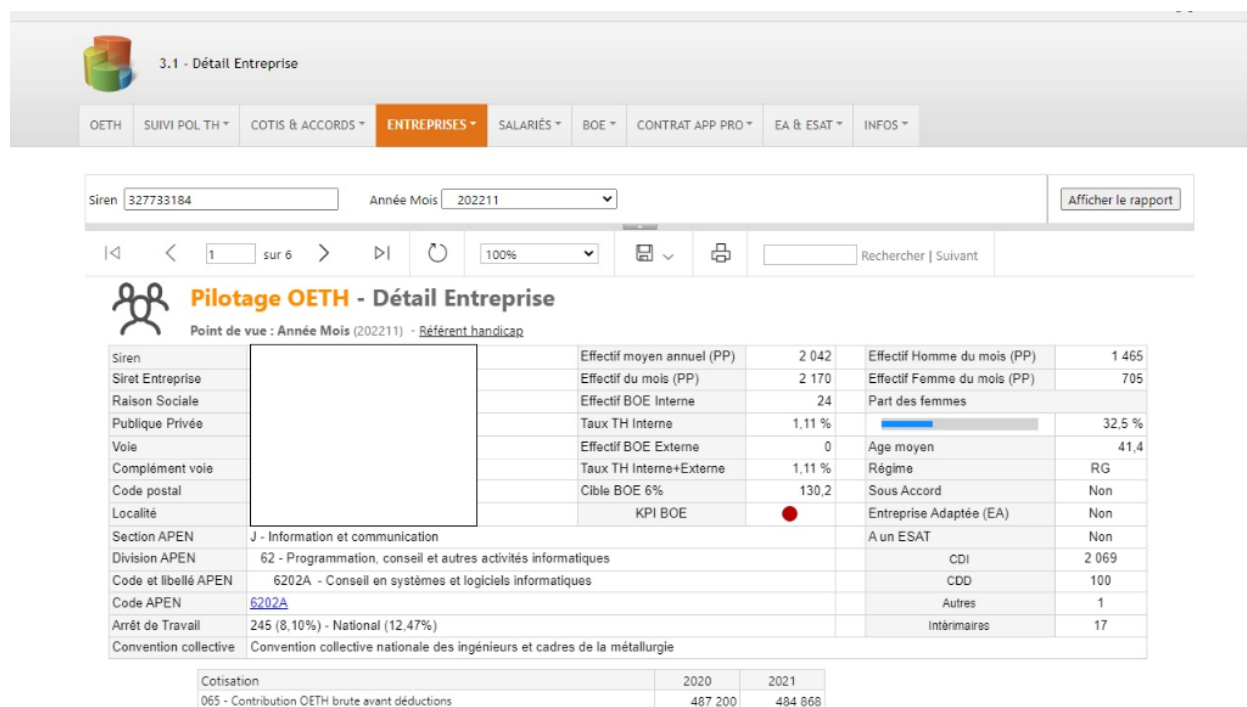
2.5.1 Repérage et accompagnement des entreprises ne respectant pas l'obligation d'emploi des travailleurs handicapés (« OETH »)

[28] La **DGEFP** a demandé en 2022 à la Dnum de réaliser un outil décisionnel destiné aux Dreets utilisant les **données brutes agrégées au niveau des entreprises et des établissements** de la DSN pour identifier les entreprises qui ne respectent pas l'Obligation d'emploi des travailleurs handicapés (OETH) et leur proposer un accompagnement adapté¹³². Cet outil, élaboré sans association de la Dares ni de l'Agefiph, devrait être déployé au 2^{ème} semestre 2023 avec des données qui seront rafraîchies tous les mois. Il permet d'obtenir le détail des effectifs, dont ceux

¹³² La loi n°2018-771 du 5 septembre 2018 pour la liberté de choisir son avenir professionnel a réformé l'OETH. Le taux d'emploi des personnes en situation de handicap est resté fixé à 6 % de l'effectif moyen annuel de l'entreprise mais les modalités de calcul ont évolué depuis le 1^{er} janvier 2020. Le décompte de l'obligation est effectué au niveau de l'entreprise (pour toutes les entreprises de 20 salariés et plus), et non plus de l'établissement. Tous les types d'emploi sont pris en compte dans le décompte des bénéficiaires de l'obligation d'emploi, y compris, par exemple, les stagiaires. Les effectifs des salariés employés par l'entreprise (ex : intérim) et relevant d'un emploi exigeant des conditions d'aptitude particulières sont également pris en compte. La première déclaration de cette obligation par le biais de la DSN, prévue également par cette loi, a été effectuée en 2021. En cas de non-respect de l'OETH, les entreprises s'exposent au versement d'une contribution financière annuelle à l'Agefiph. Son montant est calculé en fonction du nombre de travailleurs handicapés que l'entreprise aurait dû théoriquement employer.

utilisés pour le calcul de l'OETH, au niveau d'une entreprise ou d'un établissement (cf. copie écran ci-dessous), avec un focus possible sur les salariés intérimaires. Il donne aussi accès à un tableau de bord permettant d'apprécier la performance, en termes de respect de l'OETH, de chaque région, au regard du volume global d'entreprises assujetties à cette obligation (cf. copie écran suivante). Pour éviter toute confusion avec les données statistiques sur l'emploi handicapé élaborées par la Dares, la DGEFP demandera aux Dreets de ne pas diffuser les données de l'outil dans des communications externes.

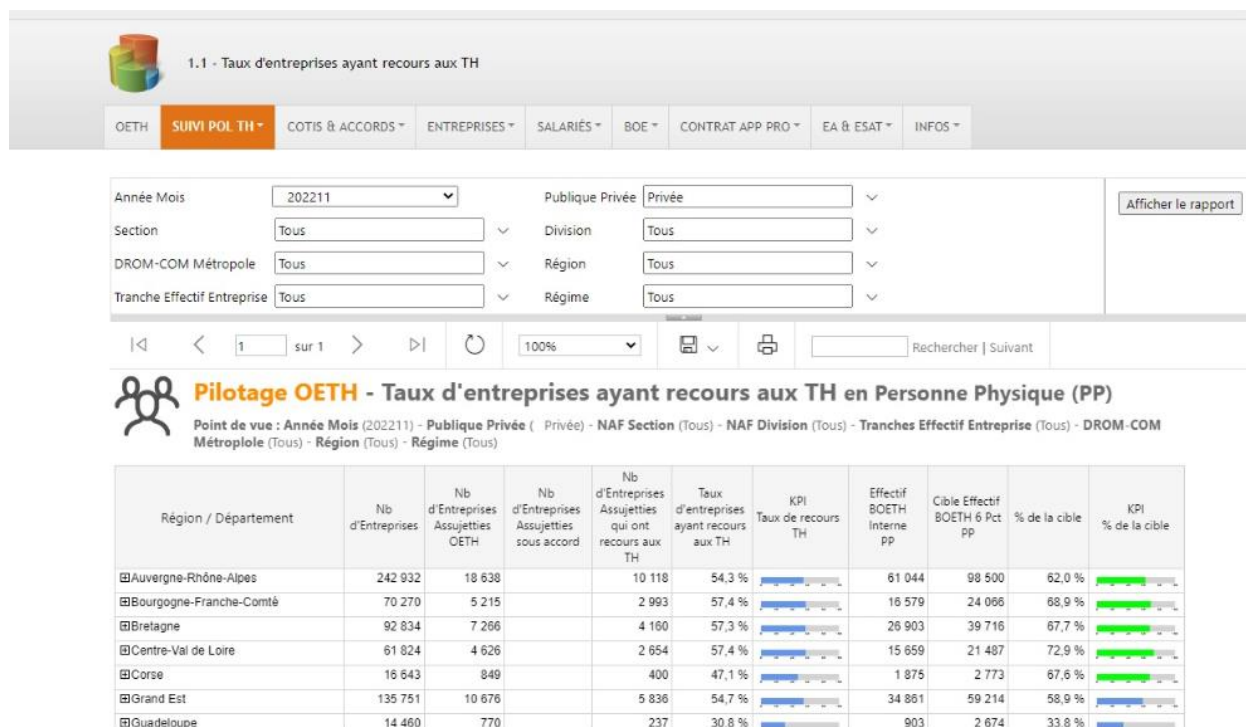
Graphique 10 : Copie écran de l'outil « OETH » (informations issues de la DSN relatives à une entreprise anonymisée)¹³³



Source : DGEFP.

¹³³ BOE : Bénéficiaires de l'obligation d'emploi.

Graphique 11 : Copie écran de l'outil « OETH » (taux d'entreprises ayant recours aux travailleurs handicapés par région)



Source : DGEFP.

2.5.2 Repérage et accompagnement des entreprises assujetties au calcul de l'index égalité professionnelle (« Égapro »)

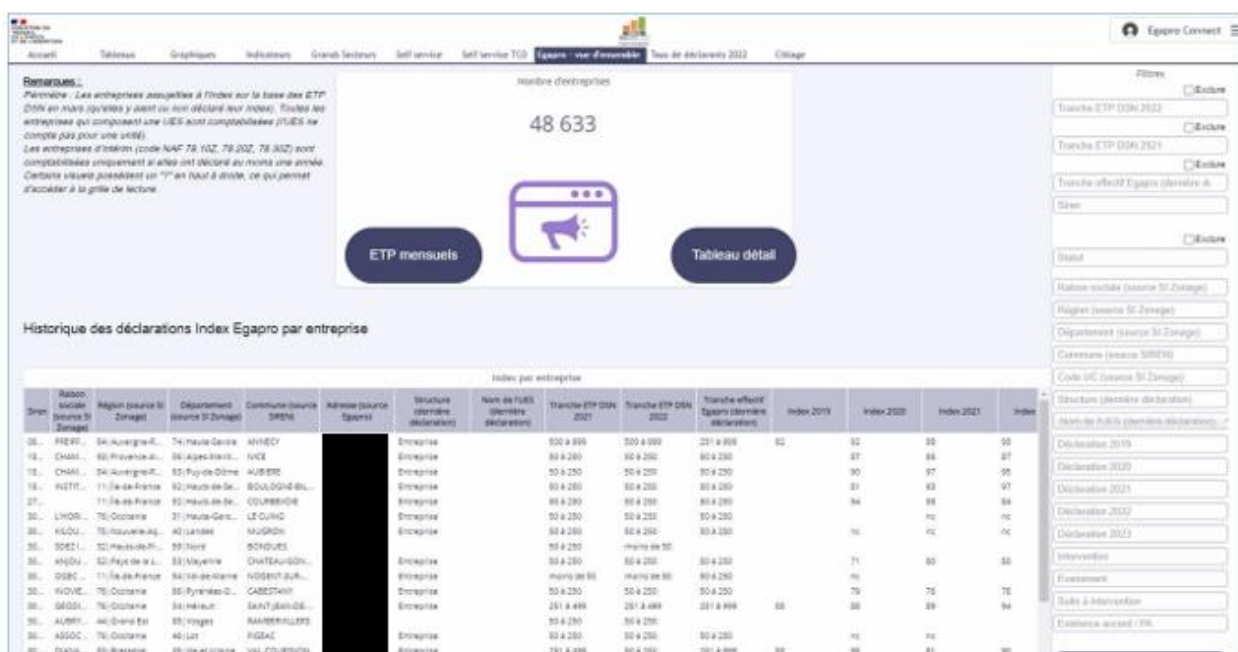
[29] La **DGT** a créé, en lien avec la Dnum, une plateforme dite « Égapro », à partir de laquelle les entreprises assujetties peuvent calculer et publier chaque année leur index de l'égalité professionnelle¹³⁴. Seules les entreprises d'au moins 50 salariés sont assujetties à cet index. Les données utilisées pour le calcul de l'index sont strictement déclaratives. Si les données issues de la DSN pourraient être utilisées pour calculer une partie des indicateurs contribuant à la note globale¹³⁵, comme l'écart de rémunération femmes/hommes, la DGT souhaite à ce stade préserver

¹³⁴ Instauré par la loi n°2018-771 du 5 septembre 2018 pour la liberté de choisir son avenir professionnel, l'index permet aux entreprises concernées d'évaluer sur 100 points le niveau d'égalité entre les femmes et les hommes en s'appuyant sur les critères suivants : écart de rémunération femmes/hommes ; écart de taux d'augmentations individuelles ; nombre de salariées augmentées à la suite de leur congé maternité ; parité parmi les dix plus hautes rémunérations ; écart de taux de promotions (seulement pour les entreprises de plus de 250 salariés). Depuis 2019, toutes les entreprises de plus de 250 salariés et, depuis 2020, toutes les entreprises d'au moins 50 salariés sont concernées par cette obligation. Depuis 2022, en cas d'index inférieur à 85 points, les entreprises doivent fixer et publier des objectifs de progression de chacun des indicateurs. Si l'index est inférieur à 75 points, les entreprises doivent adopter et publier des mesures de correction et de rattrapage. En cas de non-publication de son index par une entreprise assujettie, elle peut être redevable d'une pénalité pouvant aller jusqu'à 1 % de sa masse salariale.

¹³⁵ Le Gip-MDS a réalisé en 2021 un prototype d'automatisation du calcul de l'index à partir des données DSN. Il a conclu à un taux de conformité élevé de l'index calculé automatiquement via les données DSN à

les marges de manœuvre des entreprises dans cette démarche. En revanche, la DGT utilise les données brutes de la DSN agrégées au niveau de l'entreprise, mises à disposition par la Dnum, pour apprécier chaque année la part des entreprises assujetties à l'obligation de calcul de l'index (donc toutes les entreprises d'au moins 50 salariés) qui ont bien réalisé cette démarche (cf. copie écran ci-dessous). Elle peut ainsi calculer le taux de répondants qui s'élevait à 61 % en 2022 (85 % pour les entreprises de 1 000 salariés et plus) et identifier les entreprises n'ayant pas rempli leur obligation (ce qui peut donner lieu à des mises en demeure, voire à des pénalités financières, prononcées par l'inspection du travail). La DGT souligne que ce taux de répondants est cependant un proxy car la définition retenue par le code du travail des effectifs comptabilisés pour apprécier si une entreprise est assujettie à l'index diffère de la notion d'ETP en DSN.

Graphique 12 : Copie écran de l'outil Égapro avec comparaison de la tranche d'effectifs déclarée par l'entreprise et de la tranche d'ETP déclarés en DSN



Source: DGT

2.6 Anticiper l'impact et suivre la montée en charge d'un nouveau dispositif de politique publique

2.6.1 Simulations dans le cadre de l'étude d'impact préalable à l'évolution d'une niche sociale ou à l'attribution de compléments de rémunération

2.6.1.1 Simulations réalisées par l'Urssaf Caisse nationale sur les niches sociales

[30] **L'Urssaf Caisse nationale** a signalé à la mission plusieurs exemples de recours en 2021 et 2022 aux données brutes de la DSN à des fins de simulation de **l'impact sur les finances publiques**

celui déclaré par les entreprises. Néanmoins, la DGT a indiqué au Gip qu'elle ne souhaitait pas à ce stade automatiser le calcul de l'index.

de l'évolution de niches sociales ou de compléments de rémunération décidés par les pouvoirs publics. Réalisées par la direction des statistiques, des études et des prévisions de l'Urssaf Caisse nationale, à la demande de la Direction de la sécurité sociale ou d'un cabinet ministériel, ces simulations se fondent sur des données brutes, individuelles et/ou agrégées. Elles ont porté, à titre d'illustration, sur la répartition des nombres d'heures supplémentaires réalisées par tranche d'effectifs, sur les rémunérations brutes des métiers de la petite enfance, la participation employeur au financement des services à la personne et de la garde d'enfants ou encore sur le montant moyen des indemnités de cessation de contrat par âge. Ces données ne font pas l'objet d'une publication par l'Urssaf Caisse nationale mais peuvent être intégrées, le cas échéant, à l'étude d'impact transmise au Parlement dans le cadre des LFSS.

2.6.1.2 Simulations réalisées par la DGCS sur l'impact des accords salariaux dans le secteur social et médico-social à but non lucratif (« Ramsès »)

[31] **La DGCS utilise de son côté les données issues de la DSN pour simuler l'impact par catégorie de financeur (État ou Conseils départementaux) des accords nationaux, qui peuvent transposer à une branche des dispositifs décidés par les pouvoirs publics (ex : primes Séguret secteur des établissements et services sociaux et médico-sociaux à but non lucratif)¹³⁶, et doivent être agréés par une commission nationale¹³⁷. Une partie de ces données sont publiées dans un rapport annuel remis en février au Parlement qui indique en particulier le taux d'augmentation de la masse salariale dans les branches concernées.**

[32] Ces simulations s'appuient sur un outil d'exploitation d'agrégats dit « Ramsès » développé en 2019 par un prestataire de la Direction du numérique (Dnum) des ministères sociaux, Invénis¹³⁸. La DGCS a indiqué à la mission que, dans un premier temps, la Dares avait été sollicitée pour réaliser cet outil mais qu'elle n'avait pas été en mesure de le faire en raison principalement d'un manque de moyens. Les **données brutes pseudonymisées** de la DSN correspondant à 18 codes Naf et 19 identifiants des conventions collectives (IDCC) sont mises à disposition du prestataire par la Dnum dans la limite des données auxquelles la DGCS a droit conformément à l'arrêté « filtre » (cf. annexe n° 4). Des règles de calcul des équivalents temps plein, des primes et de la masse salariale brute sont intégrées à l'outil. Les données sont exploitables sous la forme de tableaux via « Dig Dash » (tableaux de bord) permettant l'utilisation de différents filtres (par année, code Naf, code IDCC, région ou département, code CSP, nature du contrat, tranche d'âge, genre). Les données des établissements et services sociaux et médico-sociaux du secteur public sont encore en cours d'intégration à l'outil. Selon la DGCS, le principal enjeu de fiabilisation du

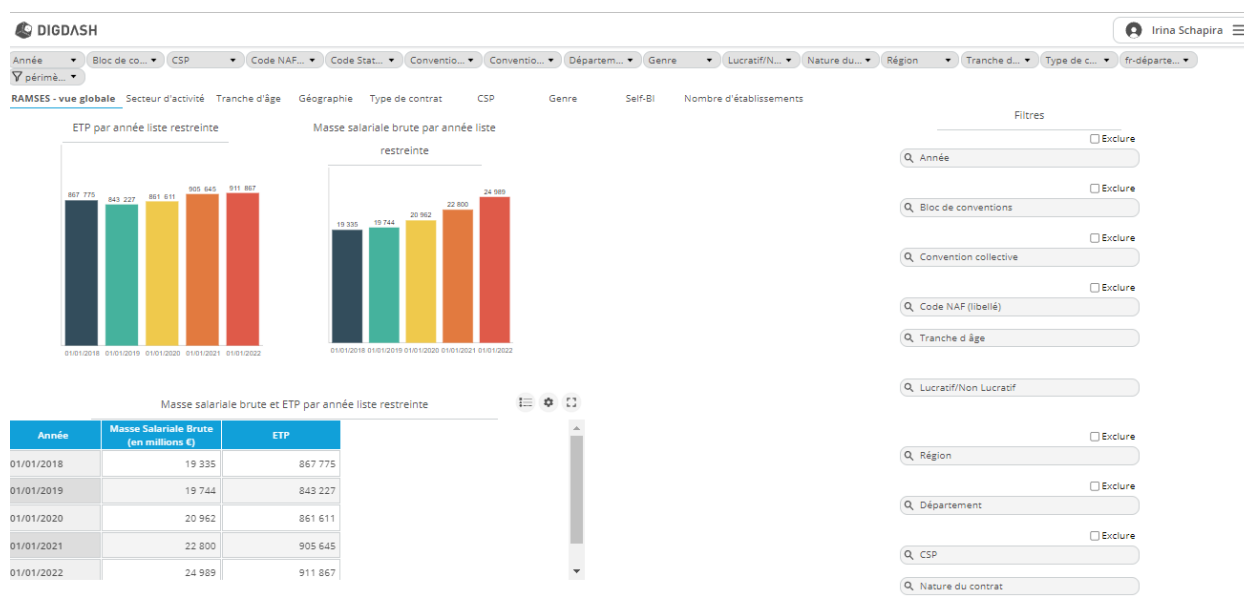
¹³⁶ En effet, ces établissements et services sociaux et médico-sociaux mobilisent, pour leur fonctionnement, des financements publics justifiant une autorisation des dépenses par l'autorité de tarification et un contrôle de l'État. Voir l'article L. 314-6 du Code de l'action sociale et des familles

¹³⁷ En 2021, cette commission nationale d'agrément a instruit 605 textes, dont 23 accords nationaux (sur lesquels 20 ont été agréés).

¹³⁸ Ce traitement automatisé de données à caractère personnel, dit « Ramsès » (pour « Réguler et accompagner la masse salariale et l'emploi du secteur social et médico-social ») est encadré par un arrêté du 6 décembre 2019 qui précise que « la finalité de ce traitement est de construire des indicateurs agrégés permettant de piloter les effectifs et la masse salariale des établissements sociaux et médico-sociaux ». Une analyse d'impact de ce traitement sur la protection des données personnelles, datée du 6 décembre 2019, a également été communiquée à la mission.

traitement « Ramsès » est lié au faible taux de renseignement du Finess des établissements, introduit en DSN en 2022, alors qu'il est important pour identifier l'autorité de tarification.

Schéma 4 : Copie d'écran du tableau de bord issu de l'outil « Ramsès »



Source : Ramsès, accès accordé à la mission par la DGCS.

2.6.2 Suivi de la montée en charge de la Prime exceptionnelle de pouvoir d'achat et de l'indemnité inflation

[33] Le **Gip-MDS** a été mobilisé par la DSS pour assurer un rôle d'ensemblier et de suivi des données relatives au paiement de la Prime exceptionnelle de pouvoir d'achat (Pepa) et de l'indemnité inflation créées respectivement en 2019 et 2021. Les données mensuelles remontées par le Gip-MDS à partir des bases Dad-E et PIQ (cf. annexe n°1) ont porté sur les versements (nombre, montants, mois de versement), les destinataires (âge, sexe, localité) et les organismes ou entreprises verseurs. S'agissant de l'indemnité inflation, le Gip-MDS a mis en place des contrôles qualité spécifiques pour identifier d'éventuels doublons.

2.7 Éclairer le marché caché de l'emploi en identifiant les entreprises susceptibles de recruter

2.7.1 Identifier les entreprises susceptibles de recruter sur un territoire et à un moment donné (« La bonne boîte »)

[34] **Pôle emploi** a créé un service, « La bonne boîte », qui à partir de l'analyse des **Déclarations préalables à l'embauche (DPAE)**, identifie, par métier et secteur géographique, les entreprises ayant une forte probabilité de recruter dans les six mois à venir, même si elles ne déposent pas toujours d'offres d'emploi auprès de Pôle emploi (cf. copie écran ci-dessous). L'objectif est d'encourager les demandeurs d'emploi à envoyer à ces entreprises des candidatures spontanées

et ainsi de faciliter les appariements sur le marché du travail. Le service met à jour les potentiels d'embauche mensuellement grâce à un algorithme prédictif qui analyse les recrutements passés.

[35] **Pôle emploi a indiqué à la mission souhaiter utiliser à terme les données issues de la DSN, plutôt que les DPAE, pour alimenter l'algorithme de « La bonne boîte ».** En effet, ces données de la DSN contiennent des informations plus fiables et plus riches que les DPAE et seraient de nature à améliorer l'algorithme. **Toutefois, Pôle emploi n'a accès aujourd'hui aux données de la DSN que pour les seuls demandeurs d'emploi et sur une durée maximale de six mois suivant leur désinscription.** À l'inverse, les DPAE sont accessibles via les Urssaf et la Caisse centrale de la mutualité sociale agricole (CCMSA) trois ans après la désinscription d'un demandeur d'emploi. Pôle emploi a fait une démarche en 2016 auprès de la Cnil pour élargir le champ des données de la DSN auxquelles elle a accès dans le cadre du « SI concernant les demandeurs d'emploi et les salariés » prévu par l'article R. 5312-38 du code du Travail. La Cnil ayant alors donné un avis négatif à une partie de ces demandes, les dispositions réglementaires encadrant les données utilisées par ce SI sont restées inchangées. Toutefois, Pôle emploi a indiqué à la mission souhaiter renouveler cette demande dans le cadre de la création de France Travail.

Graphique 13 : Copie écran d'une recherche sur « La bonne boîte » (métier recherché : boulanger, localisation : dans un rayon de 10 km autour des Lilas)

The screenshot shows the 'La bonne boîte' website interface. At the top, there are logos for the French Republic, 'La bonne boîte', and 'pôle emploi'. Below the search bar, the criteria are 'Boulangerie - viennoiserie' and 'Les Lilas 93260'. The results section shows '1246 entreprises sont susceptibles de recruter en Boulangerie - viennoiserie - Les Lilas'. A map displays the search area around Les Lilas. Three enterprise listings are visible:

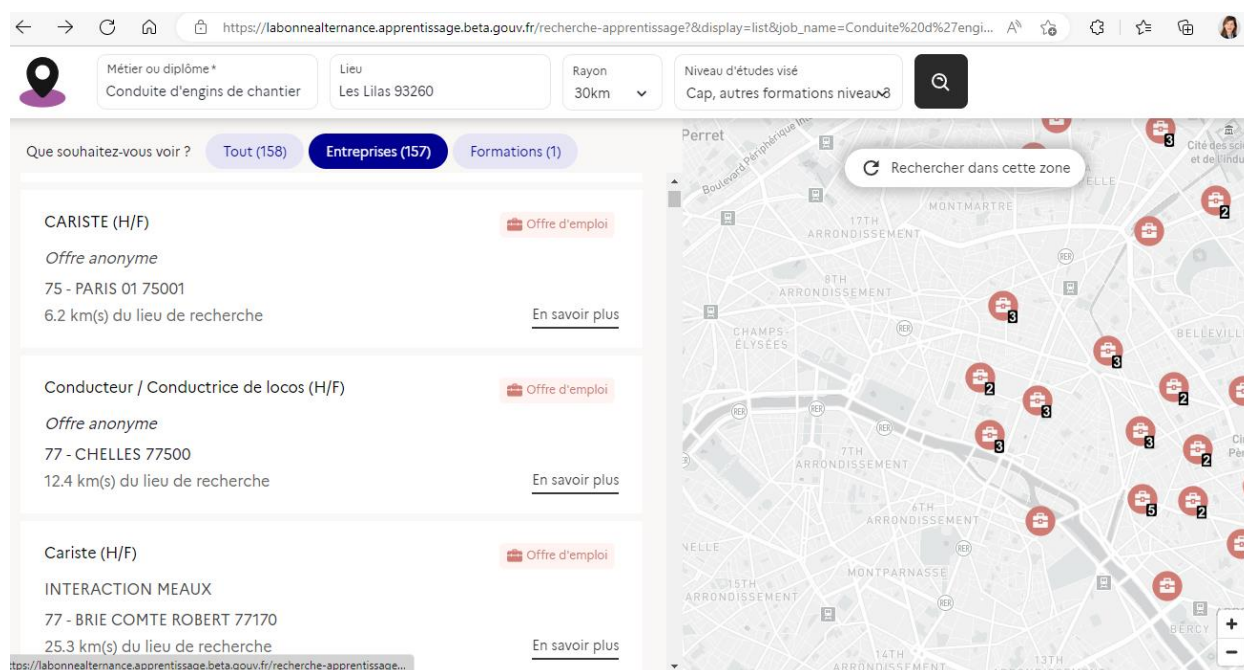
- LA ROMAINVILLE - CLICHY-SOUS-BOIS**: Fabrication industrielle de pain et de pâtisserie fraîche, 100 à 199 salariés, 8.6 km de votre lieu de recherche. Potentiel d'embauche: 5/5.
- MAISON LANDEMAINE LEVAIN D'ANTAN JA - PARIS-18**: Boulangerie et boulangerie-pâtisserie, 10 à 19 salariés, 5.8 km de votre lieu de recherche. Potentiel d'embauche: 4/5.
- LA PATISSERIE CYRIL LIGNAC - SAINT-OUEN**: Pâtisserie, 50 à 99 salariés, 8.5 km de votre lieu de recherche. Potentiel d'embauche: 5/5.

Source : Recherche effectuée par la mission le 15 février 2023.

2.7.2 Identifier les entreprises susceptibles de recruter en alternance (« La bonne alternance »)

[36] La mission interministérielle pour l'apprentissage pilote la **start-up d'État « La bonne alternance »**, service numérique dont l'objet est de diffuser des offres d'emploi en alternance et d'aider les candidats à cibler les entreprises susceptibles de recruter en alternance (cf. copie écran ci-dessous). Un **algorithme prédictif** a été conçu pour identifier le potentiel de recrutement en alternance des entreprises. Cet algorithme se fonde aujourd'hui sur les recrutements de personnes ayant été inscrites à Pôle emploi (apprécié via les DPAE) et sur les données d'une base administrative relative aux alternants (Déca). **La start-up a contacté en août 2022 le Gip-MDS pour accéder à des données brutes issues de la DSN** qui lui permettraient d'augmenter la précision de l'algorithme. En effet, ces données permettraient d'apprécier la durée des contrats pour, par exemple, identifier une entreprise qui n'a pas recruté depuis plus de deux ans en alternance (les DPAE ne fournissent pas d'informations sur la durée prévue d'un contrat). Une expérimentation a été menée avec le Gip-MDS pour tester un flux sur la structure des données DSN pertinente. La start-up attend aujourd'hui l'accord de la DSS pour mettre en place un flux mensuel à partir de la base Dad-E du Gip-MDS.

Graphique 14 : Copie écran d'une recherche sur « La bonne alternance » (métier recherché : cariste, localisation : Les Lilas)



Source : Recherche effectuée par la mission le 15 février 2023.

Tableau 2 : Synthèse des usages effectifs identifiés des données issues de la DSN à des fins de pilotage regroupés par organisme

Organisme utilisateur des données DSN	Nom du traitement de données	Finalité	Organisme fournisseur des données DSN	Données brutes ou retraitées	Données individuelles ou agrégées
DGT/inspection du travail	SI précarité	Ciblage et accompagnement	Dares	Retraitées	Agrégées
	Égapro	Ciblage et accompagnement	Dnum	Brutes	Agrégées
	Dat-IA	Ciblage et accompagnement	Dnum	Brutes	Agrégées
	Champollion	Ciblage et accompagnement	Dnum	Brutes et retraitées	Agrégées et individuelles
	WikiT/Suit	Ciblage et accompagnement	Dares	Retraitées	Agrégées
DGEFP/Dreets	OETH	Ciblage et accompagnement	Dnum	Brutes	Agrégées
Dreets	BCE	Ciblage et accompagnement	Dnum	Brutes	Agrégées
Caisse des dépôts/organismes financeurs de la formation professionnelle	Agora	Suivi du parcours en emploi	Caisse des dépôts	Brutes	Individuelles
DGCS	Ramsès	Identification des tensions RH	Dnum	Brutes	Agrégées
Réseau des Carif-Oref	Octopilot	GPEC	CASD	Retraitées	Agrégées et individuelles
DGOS/ARS/HAS/GHT	Diamant	Identification des tensions RH	Dares	Retraitées	Agrégées
Mission interministérielle à l'apprentissage	La Bonne alternance	Éclairage du marché caché de l'emploi	Gip-MDS ¹³⁹	Brutes	Agrégées

Source : Mission

¹³⁹ Il s'agit de l'organisme sollicité par « La Bonne alternance », l'arbitrage de la DSS étant encore attendu.

3 En l'absence d'une gouvernance de ses usages, les données issues de la DSN restent insuffisamment valorisées

3.1 La capacité à identifier et mobiliser les données issues de la DSN suppose déjà un certain niveau de connaissance et d'acculturation

3.1.1 Le niveau de maturité des acteurs du champ travail-emploi-formation par rapport à l'utilisation des données est inégal

[37] **L'utilisation à des fins de pilotage des données issues de la DSN suppose déjà un certain niveau d'acculturation aux démarches axées sur les données.** Si l'objet de la mission n'était pas de dresser un état des lieux des compétences en sciences de la donnée au sein des organismes pilotes des politiques du travail, de l'emploi et de la formation, force est de constater que leur niveau inégal d'acculturation à l'utilisation des données constitue un frein à la valorisation du potentiel de la DSN. Dans le cadre de la feuille de route « données » du MTPEI, la Dnum a certes mis en place des initiatives pour sensibiliser davantage les équipes de pilotage à l'intérêt du recours à la donnée, comme l'organisation d'un webinaire à l'attention des futurs inspecteurs du travail ou la refonte en cours du catalogue de formation de la Dnum. Néanmoins, ces actions restent encore peu en phase avec les priorités métiers des acteurs rencontrés.

[38] **La mission a relevé plusieurs indices traduisant une maturité des organismes pilotes généralement encore faible sur l'utilisation des données, en particulier celles issues de la DSN :**

- le recours à la donnée par les acteurs du pilotage repose généralement sur des compétences externalisées, soit celles d'un prestataire (Invénis, en l'espèce, pour « Ramsès », « Égapro » et « Dat'IA), ou celles de la Dinum(EIG¹⁴⁰)/Dnum (projets « Champollion » et « OETH »), ce qui peut expliquer le fait que les interlocuteurs rencontrés par la mission soient en difficulté pour préciser la nature ou encore la source des données utilisées ;
- une partie des cas d'usages identifiés relève d'une logique « descendante », c'est-à-dire que, même si leur conception a associé des acteurs de terrain, ils émanent d'initiatives prises au niveau central : c'est le cas, par exemple, de « Champollion » ou encore d'« Agora » (s'agissant de ce dernier cas d'usage, la mission note que les Opco rencontrés ne savaient pas que cette plateforme contenait des données issues de la DSN) ;
- les divers acteurs connaissent encore mal les informations auxquelles ils peuvent accéder via la DSN, comme par exemple la DGOS dont le pilotage des RH du secteur sanitaire repose encore beaucoup sur des enquêtes *ad hoc*, mais aussi les observatoires de branche dont la Gestion prévisionnelle des emplois et des compétences repose encore largement sur des enquêtes dont le taux de réponse est jugé insuffisant.

[39] **À cet égard, Pôle emploi, qui développe depuis 2018 des projets à vocation prédictive fait plutôt figure d'exception même si son usage des données issues de la DSN est contraint par le cadre juridique actuel.** Financé par le Fonds de transformation de l'action publique (FTAP), le

¹⁴⁰ Entrepreneurs d'intérêt général.

programme « Intelligence emploi » inclut différents projets fondés sur des algorithmes comme, pour les demandeurs d'emploi, un service d'analyse automatique du CV et, pour les entreprises, un service de prédiction sur le risque qu'une offre d'emploi ne soit pas pourvue dans un délai de 30 jours après son dépôt (« *widget* »). Pôle emploi souhaiterait aujourd'hui mobiliser davantage les données issues de la DSN, ce qui suppose une évolution du cadre juridique, pour enrichir ces algorithmes et développer de nouveaux services.

3.1.2 Les cas d'usage identifiés de production de tableaux de bord ou d'indicateurs se développent souvent en ordre dispersé

[40] En l'absence de communauté des utilisateurs des données issues de la DSN et de cartographie des usages, les passerelles entre les cas d'usage identifiés sont limitées, avec un risque de dispersion des moyens :

- alors que leurs objectifs sont proches, les projets « Ramsès » et « Diamant » sont développés de façon autonome à partir de données de la DSN différentes (données brutes pour « Ramsès » et données retraitées par la Dares pour « Diamant ») ;
- si le projet « Champollion » prend appui sur le « SI précarité », la similarité de leurs objectifs interroge le choix de retenir ce cas d'usage pour faire la preuve de la valeur des données issues de la DSN ;
- la communauté des start-up d'État intervenant dans le champ travail-emploi-formation (« La Bonne Alternance », « Diagoriente »¹⁴¹, « Immersion facilitée »¹⁴², en particulier), dont une partie exprime le souhait d'utiliser des données issues de la DSN, ne partage pas systématiquement les informations sur les moyens d'y accéder et sur la nature des données qui seraient mobilisées par chaque projet ;
- pour compenser leur faible connaissance de la DSN en particulier et de la science des données en général, certains acteurs travaillent avec la Dares pour concevoir les données en entrées et les indicateurs en sorties de leur projet et/ou pour extraire les données de la DSN du SI de la Dares (« Diamant », « Champollion »...), mais d'autres se privent de cette possibilité (« OETH », par exemple).

[41] Par ailleurs, la mission relève que la majorité des cas d'usage identifiés traduit encore une approche classique de l'utilisation de la donnée :

- la majorité des projets identifiés utilisent les données agrégées de la DSN pour produire des tableaux de bord améliorés par le recours à des techniques de datavisualisation facilitant l'utilisation de filtres et le croisement des données ;
- une partie des projets recensés mobilisent de façon longitudinale les données individuelles DSN dans une logique de suivi des individus et des parcours : c'est en particulier le cas d'« Agora » et d'« Octopilot » ;

¹⁴¹ « Diagoriente » propose un service en ligne d'orientation ciblé sur les jeunes en recherche d'emploi.

¹⁴² « Immersion facilitée » vise à favoriser la mise en relation des entreprises et des candidats à une période de mise en situation professionnelle (ou immersion).

- enfin, seul le projet « Champollion » a pour objet d'utiliser les données de la DSN à des fins prédictives dans le cadre d'un algorithme fondé sur l'intelligence artificielle (la partie IA de « Dat'IA » porte sur des données non DSN).

3.2 Le caractère incertain du cadre juridique et la difficulté à identifier les « guichets » pertinents pour accéder aux données constituent des freins à l'utilisation de la DSN

3.2.1 S'il peut être contourné, le cadre juridique relatif à la DSN tend encore à brider son utilisation à des fins de pilotage

[42] Les utilisateurs recensés de la DSN, y compris ceux qui font partie des destinataires de données prévus par l'arrêté « filtre », considèrent l'application du cadre juridique actuel comme excessivement contraignante (cf. annexe n°4) :

- parce qu'il ralentit l'accès aux données issues de la DSN en raison du processus de modification de l'arrêté « filtre » : la DGT a indiqué à la mission avoir été confrontée à un délai de plus de deux ans entre sa demande d'accéder en droit aux données sur les mouvements de main d'œuvre¹⁴³, auxquelles elle avait accès avant leur intégration à la DSN, et la modification de l'arrêté « filtre » l'y autorisant ;
- parce qu'il repose sur une maille trop fine qui oblige à chaque fois les destinataires des données, quand leurs besoins évoluent, à solliciter l'accès à des nouvelles données, avec un délai souvent important (près de six mois d'attente de la part de la DGEFP pour obtenir un accord de la DSS afin d'utiliser les données agrégées de la DSN aux fins de suivi du parcours des bénéficiaires du CEJ), en particulier quand il s'agit de données individuelles qui nécessitent au préalable de modifier l'arrêté « filtre » ;
- parce que les finalités retenues par le décret du 28 mars 2013 comme par l'arrêté « filtre » sont jugées souvent trop restrictives par rapport aux usages envisagés : ainsi, la Caisse des dépôts, qui joue un rôle global de pilotage en matière de formation professionnelle, n'est autorisée à utiliser les données de la DSN que pour « *alimenter le compte personnel de formation des salariés concernés* », la DGEFP, dont le champ d'intervention couvre l'ensemble des dispositifs favorables à l'emploi, pour « *suivre l'évolution des quotas de contrats d'apprentissage* » et la DGT, qui pilote le respect du code du Travail, jusqu'à la modification de l'arrêté « filtre » en 2022 pour « *mesurer l'audience et la représentativité syndicale et patronale* ».

[43] Pôle emploi et la DGCS ont indiqué à la mission avoir entrepris des démarches pour accéder à davantage de données issues de la DSN :

- la DGCS a demandé récemment à la DSS l'accès à des données individuelles de la DSN qui lui permettraient d'apprécier les leviers d'attractivité des métiers (cf. *supra*) ;

¹⁴³ En pratique, la DGT a cependant eu accès indirectement à ces données *via* la Dares.

- Pôle emploi a saisi en 2017 et en 2018 la Cnil pour avoir son avis, qui s'est avéré négatif, sur la possibilité d'accéder aux données de la DSN de l'ensemble des salariés, et non seulement sur les seuls demandeurs d'emploi, afin d'enrichir et de développer ses services aux demandeurs d'emploi et aux entreprises avec, par exemple, la possibilité d'utiliser, avec son accord, les données relatives au parcours professionnel antérieur du demandeur d'emploi afin d'accélérer et d'améliorer le diagnostic de sa situation au moment de l'inscription, ou encore une offre de certification du CV des demandeurs d'emploi destinée aux entreprises.

[44] **Concernant les acteurs qui ne sont pas destinataires de données de la DSN au titre de l'arrêté « filtre » :**

- **les Opco et, plus généralement, les branches professionnelles** expriment depuis 2019 le besoin d'accéder à des données plus riches et, pour certains, des données individuelles de la DSN afin de réaliser des études répondant mieux aux besoins des partenaires sociaux, mais aussi de leurs partenaires institutionnels (Régions en particulier) sur des sujets prioritaires comme l'évolution de la structure des rémunérations, l'emploi des seniors, l'égalité professionnelle entre hommes et femmes, et ce, souvent, à l'échelle d'un territoire (y compris parfois le bassin d'emploi) ou d'un métier au sein de la branche : or les données agrégées publiées en *open data* par la Dares et l'Insee ne répondent qu'imparfaitement à leurs besoins, raison pour laquelle la DGEFP, en lien avec France compétences, a soumis à la DSS une demande d'accès direct à des données agrégées et individuelles de la DSN, sans réponse à ce jour mais que la DSS n'estime a priori pas compatible avec le cadre juridique de la DSN ;
- la **Direction générale des entreprises**, et en son sein différents porteurs de projets (le « *hub* » recensant les aides d'État, « Place des entreprises » qui offre un service d'accompagnement aux TPE-PME confrontées à des difficultés, la plateforme « Impact » qui aide les entreprises à générer et à s'approprier leurs données environnementales, sociales et de bonne gouvernance) ont indiqué à la mission avoir sollicité la DSS ou le Gip-MDS pour obtenir des données agrégées de la DSN, en particulier sur les effectifs des entreprises, afin d'enrichir leur offre de services aux entreprises : ils étaient à fin janvier 2023 en attente d'un retour de la DSS ou du Gip-MDS ;
- enfin, plusieurs **start-up d'État**, dont certaines, comme « Immersion facilitée » ont identifié par l'intermédiaire de la mission l'existence de données dans la DSN utiles à leurs travaux, s'interrogent sur le meilleur moyen d'accéder à ces données, certaines, comme le Gip Plateforme de l'inclusion ayant privilégié un contournement du cadre juridique avec un accès direct via la Dnum, d'autres, comme « La bonne alternance », ayant sollicité un accord préalable de la DSS, en suspens à date.

3.2.2 Sans fonction d'aiguillage, le choix de la base de données mobilisée relève davantage d'une logique aléatoire que de l'adaptation au besoin

[45] **Pour les utilisateurs indirects des données de la DSN, il n'existe pas de doctrine claire en matière de recours aux différents organismes accédant déjà à la DSN.** En conséquence, comme l'illustrent les cas d'usage identifiés par la mission, le choix d'un interlocuteur, et de la base de données associées, repose sur une logique en grande partie aléatoire, voire opportuniste :

- c'est parce qu'il n'obtenait pas de réponse de la part de la DSS que le Gip Plateforme de l'inclusion s'est finalement tournée vers la Dnum ;
- la DGCS avait sollicité en 2019 la Dares pour répondre à son besoin d'utiliser les données de la DSN pour piloter la masse salariale du secteur médico-social mais, cette dernière ayant indiqué ne pas avoir les moyens nécessaires, elle a finalement recouru à un prestataire sur un marché-cadre de la Dnum ;
- la DGEFP a indiqué à la mission avoir privilégié un suivi du parcours en emploi des bénéficiaires du CEJ par le Gip-MDS, et non par la Dares, sans en avoir échangé au préalable avec cette dernière et essentiellement parce qu'elle faisait l'hypothèse que la Dares ne parviendrait pas à fournir les données dans le délai souhaité ;
- la responsable de la start-up d'État « Signaux faibles » à la DGE estime avoir réussi à accéder rapidement à un jeu de données DSN de la Dares grâce au soutien apporté par la directrice de cabinet de la ministre du Travail, qui était elle-même à l'origine de ce projet dans des fonctions antérieures.

[46] **Quand la DSS est saisie, elle n'est pas non plus en mesure de jouer un rôle d'aiguillage et.** De plus, l'offre de services « data » du Gip-MDS reste contrainte par les moyens actuels (de l'ordre de cinq ETP), même si elle constitue un des axes de sa feuille de route 2022-2025. Or la DSS, tout en indiquant à la mission qu'elle ne jugeait pas opportun que le Gip-MDS consacré davantage de moyens à cette offre de services, n'a pas communiqué les critères qu'elle utilisait pour prioriser les demandes d'accès aux données adressées au Gip-MDS.

3.3 L'absence d'animation de la communauté des utilisateurs de la DSN ne favorise pas l'émergence de nouveaux cas d'usage

3.3.1 Il n'existe pas de vision globale des différents usages de la DSN à des fins de pilotage ni de partage d'expérience entre les utilisateurs

[47] **La mission a relevé l'absence de tout cadre d'échange entre les différentes catégories d'acteurs sur l'intérêt d'utiliser les données de la DSN à des fins de pilotage.** L'approche des cas d'usage par la source administrative utilisée comporte certes des limites mais elle pourrait se justifier au regard (i) de son caractère récent (ii) de son potentiel encore relativement méconnu (iii) des zones de fragilités en termes de qualité détectées par les utilisateurs :

- le Gip-MDS anime certes de son côté des « ateliers qualité » mais leur objectif est moins d'échanger sur les limites de l'utilisation de la DSN à des fins de pilotage que d'élaborer des plans d'action à l'attention des éditeurs de paie pour améliorer la fiabilité des déclarations, en cohérence avec leur cœur de métier ;
- la DGEFP a mis en place une comitologie sur les données sociales mais elle s'adresse uniquement aux sous-directions en interne, son rôle d'animation vis-à-vis des acteurs externes se limitant à faire remonter et à centraliser les demandes d'accès aux données de la DSN, avec parfois un appui ponctuel pour améliorer les expressions de besoin ;
- la DSS a bien identifié la question de l'accès aux données dans l'avis de vacance du poste d'expert de haut niveau relatif à la future « *mission interministérielle en charge des projets*

de collecte et d'exploitation des données nécessaires à la gestion des prestations sociales », mais pas celle de leurs finalités d'utilisation à des fins de pilotage des politiques publiques ;

- enfin, la Dnum, qui a instruit dans le cadre du projet de « *Labor data hub* » différents scénarios pour favoriser les usages des données, joue à ce stade un rôle d'animation qui se limite à des projets déterminés, en particulier « Champollion ».

[48] Or les interlocuteurs rencontrés par la mission ont exprimé de façon unanime le souhait de disposer d'un tel cadre pour échanger sur :

- les avantages et inconvénients respectifs des différentes bases de données ;
- les retraitements utilisés pour améliorer leur qualité ;
- les moyens d'assurer la conformité des traitements réalisés au cadre de protection des données personnelles ;
- les moyens de favoriser leur appropriation par les acteurs sur le terrain (inspection du travail, Drees...) ;
- le retour d'expérience sur les outils et services développés grâce aux données de la DSN et les moyens de les enrichir, en particulier avec le recours à l'intelligence artificielle.

3.3.2 Les utilisateurs des données n'accèdent pas systématiquement à l'accompagnement dont ils pourraient avoir besoin

[49] Il n'existe aujourd'hui aucune offre formalisée d'accompagnement à l'utilisation des données de la DSN. Or, parce que le niveau de maturité « *data* » des acteurs et leurs moyens sont hétérogènes, une partie d'entre eux estime avoir besoin d'un appui pour tirer le meilleur parti des données issues de la DSN. En pratique, la Dares et la Dnum fournissent un tel appui mais (i) sans que cette offre soit proposée de façon proactive (ii) sans que les moyens à allouer à cet appui soient clairement identifiés (iii) sans critères transparents permettant de sélectionner les projets d'utilisation de la DSN qui justifieraient un soutien spécifique. Ainsi, interrogée sur les raisons pour lesquelles la Dnum avait apporté un soutien, et donc des moyens, au projet « OETH » de la DGEFP, elle a indiqué que c'était surtout parce que cette direction en avait fait la demande.

[50] S'agissant des acteurs positionnés en dehors des ministères sociaux, y compris dans le privé (branches professionnelles, Opco), l'appui qu'ils peuvent recevoir pour utiliser les données de la DSN est encore plus aléatoire. Grâce au rôle d'intermédiaire joué par France compétences, les Opco ont obtenu de la part de la Dares une adaptation de ses livrables, en particulier des estimations trimestrielles d'emploi, susceptible de faciliter leur utilisation des données de la DSN. Le rôle de France Compétences dans l'appui à l'utilisation des données issues de la DSN reste par ailleurs limité alors que les compétences et les moyens des observatoires rattachés aux Opco semblent hétérogènes. Le réseau des Carif-Oref (RCO), qui utilise des données de la DSN hébergées au CASD pour la plateforme « Octopilot », a indiqué à la mission avoir bénéficié d'un soutien de la part de l'Insee, qui tenait surtout au fait que des liens interpersonnels avaient été noués quand RCO acquérait des données sur-mesure auprès de l'Insee.

ANNEXE 4 : Encadrement de l'accès aux données DSN et de leurs usages

ANNEXE 4 : ENCADREMENT DE L'ACCES AUX DONNEES DSN ET DE LEURS USAGES154

1 LA REGULATION DE L'ACCES AUX DONNEES ISSUES DE LA DSN ET DE LEURS USAGES REPOSE SUR UNE ARCHITECTURE JURIDIQUE COMPLEXE156

- 1.1 LES DISPOSITIONS RELATIVES A LA DSN CONDITIONNENT L'ACCES A CES DONNEES A DES USAGES PRECIS PAR UN NOMBRE LIMITE D'ACTEURS156
 - 1.1.1 *La DSN constitue un traitement de données à caractère personnel se substituant à plusieurs déclarations ou formalités antérieures*156
 - 1.1.2 *Au niveau réglementaire, les usages et les destinataires des données issues de la DSN sont étroitement liés à sa fonction première*156
 - 1.1.3 *Ce cadre juridique, difficile à faire évoluer, n'a pas été pensé pour favoriser l'utilisation des données issues de la DSN*.....157
- 1.2 CE CADRE JURIDIQUE, QUI EST AUJOURD'HUI QUESTIONNE, N'EPUISE PAS LES VOIES D'ACCES ET LES USAGES ASSOCIES AUX DONNEES ISSUES DE LA DSN.....158
 - 1.2.1 *Les dispositions, postérieures à la DSN, sur l'utilisation des données administratives interrogent le cadre juridique actuel*.....158
 - 1.2.2 *Ce cadre juridique peut par ailleurs être contourné sur la base de dispositions autorisant un usage indirect des données issues de la DSN*.....160
- 1.3 DES REGLES SPECIFIQUES ENCADRENT L'USAGE DES DONNEES DE LA DSN RETRAITEES A DES FINS D'INFORMATION GENERALE163
 - 1.3.1 *L'usage des données de la DSN traitées par la statistique publique obéit à des standards propres, relevant de textes de niveaux national et européen*.....163
 - 1.3.2 *Les données statistiques sont réutilisables à des fins d'information générale ou de recherche, mais pas a priori à des fins de pilotage, même si des réflexions sont en cours pour faire évoluer ce cadre*.....164

2 LES MODALITES D'ACCES AUX DONNEES DE LA DSN NE REPONDENT PAS A TOUS LES BESOINS DES UTILISATEURS165

- 2.1 LE PROCESSUS D'INSTRUCTION DES DEMANDES D'ACCES AUX DONNEES DE LA DSN EST JUGE TROP LENT165
 - 2.1.1 *La DSS est chargée d'instruire ces demandes dans le cadre d'un processus imbriqué avec celui de l'évolution de la norme DSN*165
 - 2.1.2 *Le délai de modification de l'arrêté-filtre, indispensable pour accéder à des données individuelles, est jugé trop long par rapport aux besoins*.....166
 - 2.1.3 *L'objectif de simplification associé à la DSN tend à exclure les évolutions de la norme à des fins d'information générale ou de pilotage*166
- 2.2 LA DSS N'EST PAS EN MESURE D'INSTRUIRE ET D'AIGILLER DE FAÇON PERTINENTE LES DEMANDES D'ACCES AUX DONNEES DE LA DSN167
 - 2.2.1 *La DSS n'a pas les moyens ni l'expertise pour instruire correctement les demandes d'accès aux données de la sphère travail-emploi*.....167
 - 2.2.2 *La DSS mobilise le Gip-MDS pour répondre aux demandes d'accès aux données sans capacité à orienter vers les bases les plus adaptées aux besoins*..... 171

3	LA STRATEGIE DE VALORISATION DES DONNEES ADMINISTRATIVES QUESTIONNE LES MODALITES DE REGULATION APPLIQUEES A LA DSN	172
3.1	LE DEVELOPPEMENT DE L'USAGE DES DONNEES ADMINISTRATIVES EST DEvenu UN ENJEU CENTRAL POUR LE PILOTAGE DES POLITIQUES PUBLIQUES	172
3.1.1	<i>L'assouplissement de l'accès aux données administratives est un préalable à leur valorisation</i>	<i>172</i>
3.1.2	<i>La valorisation des données administratives contribuerait à la transformation, en particulier numérique, de l'action publique</i>	<i>173</i>
3.2	LE PROJET DE LABOR DATA HUB, QUI RESTE UN OBJET A DEFINIR, VISE A SUSCITER UN USAGE ELARGI DES DONNEES ADMINISTRATIVES TELLES QUE LA DSN	174
3.2.1	<i>Dans le cadre de la feuille de route du MTEI, la création d'un « hub » est l'un des vecteurs identifiés pour développer l'usage des données</i>	<i>174</i>
3.2.2	<i>En dépit du cadrage réalisé par la Dnum, le Labor data hub reste à ce jour un objet administratif non identifié.....</i>	<i>175</i>
3.2.3	<i>Les « hubs » de données identifiés par la mission dans d'autres ministères confirment la variété des fonctions et modèles d'organisation possibles.....</i>	<i>178</i>
3.3	EN PRATIQUE, LE CASD CONSTITUE UN « HUB » INTERSECTORIEL D'ACCES A DES DONNEES DONT UNE PARTIE SONT ISSUES DE LA DSN	179
3.4	L'ASSOUPLEMENT DES MODALITES D'ACCES AUX DONNEES ADMINISTRATIVES DOIT ETRE CONCILIE AVEC LES REGLES RELATIVES A LA PROTECTION DES DONNEES	179
3.4.1	<i>Les utilisateurs des données DSN sont chacun responsables de la conformité des usages au RGPD.....</i>	<i>179</i>
3.4.2	<i>En pratique, la mission a constaté des standards variables selon les utilisateurs des données issues de la DSN.....</i>	<i>180</i>

1 La régulation de l'accès aux données issues de la DSN et de leurs usages repose sur une architecture juridique complexe

1.1 Les dispositions relatives à la DSN conditionnent l'accès à ces données à des usages précis par un nombre limité d'acteurs

1.1.1 La DSN constitue un traitement de données à caractère personnel se substituant à plusieurs déclarations ou formalités antérieures

[1] Créée par la loi n°2012-387 du 22 mars 2012 en vue de regrouper et simplifier les démarches déclaratives des entreprises à partir des données issues de la gestion de la paie des salariés, la **Déclaration sociale nominative (DSN) constitue un traitement de données à caractère personnel**. La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés définit un traitement de données à caractère personnel comme « *toute opération ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé* », sachant que constitue une donnée à caractère personnel, toujours au sens de cette loi, « *toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres* ».

[2] La DSN se substituant à une quarantaine de formalités déclaratives préexistantes¹⁴⁴, la loi prévoit que ses données servent essentiellement dans le cadre de ces démarches, tout en ouvrant la voie à un usage plus général mais circonscrit aux missions des organismes destinataires, comme pour tous les traitements de données à caractère personnel. Ainsi, l'article L. 133-5-3 du code de la sécurité sociale dispose que « *les données de cette déclaration servent au recouvrement des cotisations, des contributions sociales et de certaines impositions, à la vérification de leur montant, à l'ouverture et au calcul des droits des salariés en matière d'assurance sociale, de formation et de prévention des effets de l'exposition à certains facteurs de risques professionnels, à la détermination du taux de certaines cotisations, au versement de certains revenus de remplacement ainsi qu'à l'accomplissement par les administrations et organismes destinataires de leurs missions* ».

1.1.2 Au niveau réglementaire, les usages et les destinataires des données issues de la DSN sont étroitement liés à sa fonction première

[3] Les finalités des données issues la DSN, telles que définies par l'article 3 du décret n°2013-266 du 28 mars 2013 qui place ce traitement sous la responsabilité du ministre en charge de la sécurité sociale, sont étroitement liées à la fonction d'origine de ce dispositif. Le terme de « pilotage » n'est pas utilisé dans la description des finalités de ce traitement. Les treize destinataires¹⁴⁵ des données directement issues de la DSN, également listés par le

¹⁴⁴ En 2017, selon le Gip-MDS, la DSN s'est substituée à 26 formalités déclaratives préexistantes. Sous l'effet de l'extension du périmètre de la DSN à de nouvelles formalités, ce nombre s'est élargi.

¹⁴⁵ Urssaf ; caisses d'allocations familiales ; caisses primaires d'assurance maladie et caisses d'assurance retraite et de la santé au travail ; Pôle emploi ; ministère chargé du travail, ministère chargé de la santé,

décret n°2013-266 du 28 mars 2013¹⁴⁶, sont aussi ceux qui utilisent ces données majoritairement à des fins de gestion, en lieu et place des informations qui étaient auparavant collectées à travers des formalités spécifiques. Le décret précise que les données de la DSN sont transmises à ces administrations et organismes « *dans la limite des informations qui sont strictement nécessaires à l'accomplissement de leurs missions* », le V de l'article IV renvoyant à un arrêté du ministre chargé de la sécurité sociale pris après avis de la Commission nationale de l'informatique et des libertés (Cnil), qui comporte la liste exacte des données communicables à chacun dans le respect en particulier des finalités du traitement.

1.1.3 Ce cadre juridique, difficile à faire évoluer, n'a pas été pensé pour favoriser l'utilisation des données issues de la DSN

1.1.3.1 Le décret du 28 mars 2013 ne favorise pas l'utilisation des données issues de la DSN à des fins de pilotage

[4] **L'utilisation des données de la DSN à des fins de pilotage n'est pas prévue par les textes réglementaires.** Si l'évolution du cadre réglementaire a conduit à étendre les finalités associées au traitement des données de la DSN, la rédaction retenue, qui exclut le terme de « pilotage », reste impropre à un développement de l'usage des données à cette fin. Les révisions du décret n°2013-266 du 28 mars 2013 ont permis d'ajouter les finalités suivantes qui peuvent se rattacher à une fonction de pilotage¹⁴⁷ :

- « *faciliter [le] placement [des salariés] en cas de situation de recherche d'emploi* », la rédaction de ce 2° de l'article 3 interpellant par l'absence de référence aux demandeurs d'emploi et par l'utilisation d'une notion, celle de placement, à l'acceptation plus étroite que celle, communément utilisée aujourd'hui, d'accompagnement (ou de conseil en évolution professionnelle) ;
- « *permettre aux services de l'Etat de contrôler l'application du droit du travail, notamment en matière de lutte contre le travail illégal, et de mesurer l'audience et la représentativité syndicale et patronale dans les entreprises* », finalité absente de la version initiale du décret.

1.1.3.2 La granularité de l'arrêté « filtre » traduit aussi un encadrement strict de l'accès aux données issues de la DSN

[5] **L'encadrement réglementaire de l'usage des données issues de la DSN apparaît, en raison du degré de granularité de l'arrêté « filtre », comme particulièrement strict.** En effet, l'arrêté

ministère chargé des affaires sociales ; organismes chargés de la gestion d'un régime de retraite complémentaire obligatoire ; organismes, institutions et entreprises chargés de la gestion de garanties de protection sociale complémentaire ; organismes chargés de la gestion d'un régime spécial ; caisses de congés payés des professions du spectacle ; Institut national de la statistique et des études économiques (Insee) ; direction générale des finances publiques (DGFiP) ; Agence de services et de paiement (ASP) ; Caisse des dépôts et consignations.

¹⁴⁶ Voir le III de l'article 4.

¹⁴⁷ Ces deux finalités ont été ajoutées lors de la révision intervenue en 2016. Au total, le décret n°2013-266 du 28 mars 2013 a été révisé quatre fois à date.

du 9 février 2022 fixant les données de la DSN adressées aux administrations et organismes compétents, dit arrêté « filtre », prévoit en annexe 1 un tableau listant, pour chaque destinataire, et par catégorie de données, les informations transmises. Outre l'Insee qui a accès à la totalité des informations mentionnées par l'arrêté, sept acteurs présents dans le champ travail-emploi-formation professionnelle figurent parmi les destinataires de données¹⁴⁸.

[6] **Par ailleurs, les données sont distribuées entre organismes attributaires de façon hétérogène, sans logique évidente.** Ainsi, pour une même catégorie d'informations, comme l'identité du salarié, la Dares a accès à son niveau de formation le plus élevé, mais pas la Délégation générale à l'emploi et à la formation professionnelle (DGEFP) ni Pôle emploi. Dans la catégorie des données relatives au contrat ou à la relation de travail, la Direction générale du travail (DGT) a accès aux informations relatives aux facteurs d'exposition à la pénibilité mais pas la DGEFP.

[7] **Si cet arrêté « filtre » a été révisé à six reprises depuis 2016, sa maille reste trop fine pour s'adapter à l'évolution des usages des données dans le cadre des traitements prévus par ce même texte.** L'annexe 2 de cet arrêté apporte en effet des « *précisions sur les traitements visant certains destinataires des données de la DSN* ». Ainsi, s'agissant de la DGEFP, les traitements annuels mis en place à partir des données de la DSN viseraient uniquement à « *suivre l'évolution des quotas de contrat d'apprentissage dans les entreprises de 250 salariés et plus* » et, pour la Caisse des dépôts, à « *alimenter le compte personnel de formation des salariés concernés* ». En pratique, comme la mission l'a constaté, le périmètre des données transmises aux différents acteurs et les finalités des traitements mis en place excèdent ceux prévus par cet arrêté (cf. annexe n°3). Cette absence d'effectivité ex post questionne d'autant plus la pertinence du cadre rigoureux posé ex ante, nonobstant la nécessité de prévoir un cadre conforme aux dispositions relatives à la réglementation sur la protection des données personnelles (cf. *infra*).

1.2 Ce cadre juridique, qui est aujourd'hui questionné, n'épuise pas les voies d'accès et les usages associés aux données issues de la DSN

1.2.1 Les dispositions, postérieures à la DSN, sur l'utilisation des données administratives interrogent le cadre juridique actuel

[8] **Le cadre juridique de la DSN est questionné à l'aune de deux catégories de dispositions introduites entre 2016 et 2018 :**

- d'une part, les dispositions issues de la loi n°2016-1321 du 7 octobre 2016 pour une République numérique dont le I de l'article 1^{er} pose un principe de communication, entre administrations, dans le respect des règles relatives à la protection des données personnelles, des documents administratifs qu'elles détiennent, les informations figurant dans ces documents pouvant être utilisées à des fins d'accomplissement de missions de

¹⁴⁸ Délégation générale à l'emploi et à la formation professionnelle (DGEFP), Direction générale du Travail (DGT), Pôle emploi, Agefiph, Caisse des dépôts et consignations, Agence de services et de paiement (ASP), Dares.

service public autres que celles pour les besoins de laquelle les documents ont été produits ou reçus¹⁴⁹ ;

- d'autre part, le Règlement général sur la protection des données (RGPD), entré en application le 25 mai 2018, qui complète la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (LIL) et prévoit en particulier un allègement des formalités administratives dans la plupart des cas et une responsabilisation des responsables de traitement de données.

[9] L'impact de ces dispositions sur la pertinence, voire la légalité, de l'arrêté « filtre », fait l'objet d'appréciations divergentes :

- d'un côté, le directeur interministériel du numérique (Dinum), en sa qualité d'administrateur général des données¹⁵⁰ (AGD), a estimé, dans un courrier du 16 novembre 2021 adressé à la Direction du numérique (Dnum) des ministères sociaux¹⁵¹, que les dispositions issues de la loi pour une République numérique, ainsi que le principe de responsabilisation des administrations créant des traitements de données, rendaient illégal l'arrêté « filtre » fondé sur une logique de liste exhaustive liée à un régime d'autorisation préalable¹⁵² ;
- de l'autre côté, la Direction des affaires juridiques (DAJ) des ministères sociaux, ainsi que la DSS, soulignent que la DSN implique la mise en œuvre d'un traitement de données à caractère personnel qui doit être conforme au RGPD et à la LIL du 6 janvier 1978, ce que ne remet pas en cause l'article 1^{er} de la loi pour une République numérique, et exige par conséquent un encadrement au niveau réglementaire afin de préciser les catégories de données traitées et la liste de leurs destinataires¹⁵³.

¹⁴⁹ Les échanges de données entre administrations à des fins de traitement de démarches des usagers sont encadrés par les articles L. 114-8 et suivants du code des relations entre le public et l'administration, tels que modifiés par l'article 62 de la loi n°2022-217 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale, qui développe les possibilités d'échange d'informations ou de données entre administrations afin de traiter les demandes ou déclarations présentées par le public ou d'informer les personnes sur leurs droits. Le décret d'application de ces dispositions est cours de finalisation à date.

¹⁵⁰ La circulaire n°6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources a confié au directeur interministériel du numérique, en sa qualité d'administrateur général des données, algorithmes et codes sources, la coordination du réseau des administrateurs ministériels des données.

¹⁵¹ Ce courrier fait suite à une saisine de l'administratrice des données du ministère en charge du Travail datée du 11 octobre 2021 qui interrogeait l'AGD sur les contraintes, d'ordre réglementaire, à l'extension de l'exploitation de la DSN prévue par la feuille de route ministérielle relative aux données.

¹⁵² Dans ce courrier, l'AGD écrit en particulier qu'« il apparaît que le législateur a clairement entendu organiser un régime de circulation large des données de la DSN puisque pouvaient y avoir accès non seulement les organismes de protection sociale mais également l'ensemble des services de l'État » et que « eu égard au cadre juridique (...), aux orientations définies par le Premier ministre, et par la simple application de la hiérarchie des normes, il m'apparaît clairement que l'arrêté « filtre » ne peut avoir pour effet de limiter le cadre de la circulation des données prévues par la loi ».

¹⁵³ Cet encadrement est nécessaire pour respecter les principes de limitation des finalités et de minimisation des traitements de données prévus à l'article 5 du RGPD qui dispose que « les données à caractère personnel (...) doivent être collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités, (...) [et] adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées ».

[10] Ainsi, la DAJ des ministères sociaux indique que, si un exercice de liste de « catégories » de données de la DSN communicables par destinataire reste nécessaire, il pourrait se limiter à un décret, sans que le renvoi à un arrêté « filtre » soit indispensable, dès lors que le RGPD est respecté :

- le décret n°2013-266 du 28 mars 2013 devrait alors être modifié afin de supprimer le renvoi à un arrêté, pour définir au niveau du décret les catégories de données communicables à chaque destinataire en lien avec les finalités du traitement de données ;
- un acte réglementaire resterait nécessaire, au sein d'une administration de l'État, pour la création d'un traitement de données utilisant les données issues de la DSN, ce texte pouvant *a priori* être un arrêté dès lors que le traitement relève de la compétence d'un chef de service¹⁵⁴ ;
- la liste exacte des données de la DSN utilisées dans le cadre de chaque traitement de données créé par une administration devrait ensuite figurer dans l'Analyse d'impact relative à la protection des données (AIPD)¹⁵⁵ réalisée, le cas échéant, par le responsable de traitement.

1.2.2 Ce cadre juridique peut par ailleurs être contourné sur la base de dispositions autorisant un usage indirect des données issues de la DSN

1.2.2.1 Les conditions d'utilisation du Nir, encadrées par un décret *ad hoc*, offrent plus de souplesse que les dispositions relatives à la DSN

[11] Le décret encadrant la mise en œuvre des traitements de données comportant l'usage du Nir autorise un usage de cette donnée à des fins plus larges que celles prévues par les textes relatifs à la DSN. Ce décret n°2019-341 du 19 avril 2019 définit les catégories d'acteurs concernés et les finalités des traitements pour lesquels l'utilisation du Nir, qui figure dans la DSN, est autorisée, y compris dans les champs du travail et de l'emploi. Or les finalités associées à l'usage du Nir dans ces champs sont à la fois plus précises et plus nombreuses que celles prévues par le décret du 28 mars 2013 et par l'arrêté « filtre ». À titre d'exemple, le Nir peut être utilisé par les acteurs autorisés dans les champs du travail et de l'emploi, dont la DGEFP, « pour l'accomplissement de leurs missions de suivi, de pilotage, d'études et d'évaluation des politiques publiques en matière d'emploi et de formation professionnelle, ainsi qu'à des fins d'études,

¹⁵⁴ La DAJ des ministères sociaux a indiqué à la mission qu'un groupe de travail, piloté par le Secrétariat général du Gouvernement (SGG), étudiait à l'heure actuelle la possibilité d'assouplir les formalités administratives nécessaires pour créer un traitement de données. À ce stade, un décret simple, soumis à l'avis de la Cnil, reste nécessaire pour créer un traitement de données relevant de plusieurs chefs de service ne relevant pas du même ministre.

¹⁵⁵ Prévue par le RGPD, l'AIPD est un outil de responsabilisation des organismes qui créent un traitement de données. Elle est obligatoire pour les traitements susceptibles d'engendrer des risques élevés. Elle se décompose en trois parties : une description détaillée du traitement mis en œuvre ; l'évaluation de la nécessité et de la proportionnalité du traitement (finalité, données et durée de conservation, information et droits des personnes...) ; l'étude des risques sur la sécurité des données et leurs impacts potentiels sur la vie privée.

notamment de suivi de parcours et des dispositifs qu'ils mettent en œuvre »¹⁵⁶. Il peut également être utilisé par la Caisse des dépôts pour assurer le partage d'informations entre les organismes finançant la formation et les organismes de formation sur les parcours des stagiaires et apprentis¹⁵⁷, sans que cette finalité soit prévue par l'arrêté « filtre ».

[12] **Par ailleurs, la liste des organismes habilités à accéder au Nir via la DSN n'est pas totalement cohérente avec celle des acteurs autorisés à utiliser cette donnée dans le cadre du décret n°2019-341 du 19 avril 2019.** Ainsi, l'Agefiph, qui reçoit les Nir des bénéficiaires de l'obligation d'emploi des travailleurs handicapés via la DSN, conformément à l'arrêté « filtre », ne figure pas parmi les acteurs autorisés à l'utiliser. À l'inverse, les opérateurs de compétences (Opcv) sont autorisés à utiliser le Nir pour la gestion du compte personnel de formation et des contrats d'apprentissage, mais ils ne peuvent pas aujourd'hui, en droit, accéder aux données individuelles de la DSN.

1.2.2.2 Un accès indirect à des données issues de la DSN est autorisé par le code du travail pour des usages et un périmètre d'utilisateurs plus étendus

[13] **Les membres du service public de l'emploi peuvent accéder à une partie des données à caractère personnel issues de la DSN via le « système d'information concernant les demandeurs d'emploi et salariés ».** Prévu par l'article R. 5312-38 du code du travail, ce système d'information (SI)¹⁵⁸, placé sous la responsabilité de Pôle emploi, est alimenté en particulier par les données de la DSN auquel l'opérateur a accès conformément à l'arrêté « filtre ». Ces données, issues du traitement de données dénommé « Déclaration sociale nominative Pôle emploi » (DSNPE), concernent plusieurs catégories d'individus, dont les demandeurs d'emploi en cours ou radiés depuis moins de six mois pour lesquels l'employeur procède à la déclaration d'une DSN mensuelle¹⁵⁹. Or l'article R. 5312-43 dispose que les membres du service public de l'emploi (SPE) sont destinataires de tout ou partie des données à caractère personnel issues de ce SI « à raison de leurs attributions respectives et dans la stricte limite des informations dont ils ont à connaître dans le cadre de l'exercice de leurs missions ». C'est sur la base de ces dispositions que la DAJ des ministères sociaux a estimé que l'Unédic, qui demandait à la DSS un accès direct aux données individuelles de la DSN, pouvait, en tant que membre du SPE et sous la seule responsabilité de Pôle emploi, accéder indirectement à ces données¹⁶⁰.

[14] **Un accès indirect aux données issues de la DSN est aussi prévu par le code du Travail dans le cadre du dispositif « Agora » visant à favoriser le partage d'informations sur les parcours des bénéficiaires d'une formation professionnelle.** Ce partage d'informations entre les organismes de

¹⁵⁶ Voir le 10° du C de l'article 2 du décret n°2019-341 du 19 avril 2019.

¹⁵⁷ Voir le 12° du C de l'article 2 du décret n°2019-341 du 19 avril 2019. Ce partage d'informations, introduit par la loi n° 2016-1088 du 8 août 2016 relative au travail, à la modernisation du dialogue social et à la sécurisation des parcours professionnels, est prévu par l'article L. 6353-10 du code du Travail.

¹⁵⁸ Article issu du décret n°2016-729 du 1^{er} juin 2016 relatif au système d'information concernant les demandeurs d'emploi et salariés mis en œuvre par Pôle emploi. Ce texte avait pour objet de mettre le SI de Pôle emploi en conformité avec les exigences de la Cnil.

¹⁵⁹ Pôle emploi réceptionne aussi de la part de la Cnav les informations issues de la DSN relatives à tout salarié pour lequel l'employeur procède à une déclaration « signalement de fin de contrat » et à tout salarié pour lequel une entreprise de travail temporaire effectue une déclaration de mission via une DSN mensuelle.

¹⁶⁰ Analyse de la DAJ communiquée à la DSS le 22 octobre 2021.

formation, les organismes financeurs ou chargés du conseil en évolution professionnelle, et la Caisse des dépôts inclut des données individuelles relatives à l'emploi des stagiaires et apprentis d'une formation professionnelle¹⁶¹. Ces données, issues de la DSN, correspondent à celles dont la Caisse des dépôts est destinataire, conformément à l'arrêté « filtre » au titre de sa mission d'alimentation du Compte personnel de formation (CPF). Bien que la plateforme Agora, présentée par la Caisse des dépôts comme le « *data hub* de la formation », soit à ce stade logée au sein du SI CPF, l'utilisation des données de la DSN à une fin non prévue par les textes relatifs à la DSN justifierait, selon la DGEFP, une mise en cohérence de ces derniers avec les dispositions du code du Travail.

[15] **Enfin, la création du « passeport de compétences » et du « passeport de prévention », qui mobilisent également des données issues de la DSN, pourrait aussi conduire à élargir l'accès à ces données à des fins non prévues par les textes relatifs à la DSN.** Prévu par la loi n°2018-771 du 5 septembre 2018 pour la liberté de choisir son avenir professionnel, le « passeport d'orientation, de formation et de compétences », opéré par la Caisse des dépôts, donnera la possibilité aux titulaires d'un CPF de faire certifier les données relatives à leur carrière, leurs formations et leurs diplômes. La Caisse des dépôts envisage à date d'utiliser les données issues de la DSN pour recenser une partie de ces informations. L'exploitation des flux DSN est aussi prévue par la Caisse des dépôts dans le cadre du passeport de prévention, créé par la loi n°2021-1018 du 2 août 2021 pour renforcer la prévention en santé au travail. L'objectif est que les entreprises, qui renseigneront dans ce passeport les informations sur les formations relatives à la santé et à la sécurité au travail dispensées à leur initiative aux salariés, puissent retrouver dans cet espace leurs salariés.

1.2.2.3 Avec l'arrêté « Champollion », la Dnum des ministères sociaux remet en cause le cadre juridique strict de la DSN

[16] **L'arrêté du 21 septembre 2022 de la Dnum crée un traitement automatisé de données à caractère personnel dit « Champollion » qui, au regard de ses finalités très larges, constitue une sorte de « manifeste » en faveur de la circulation et de l'exploitation des données dans le champ des directions du ministère en charge du Travail.** Placé sous la responsabilité de l'administrateur ministériel des données de ce ministère, en l'espèce la Dnum, ce traitement a pour finalité de « *structurer et de mettre à disposition des administrations relevant du champ du travail, de l'emploi et de l'insertion, des données pour l'accomplissement de leurs missions* ». Les données à caractère personnel issues de la DSN font partie de celles pouvant être enregistrées dans ce traitement¹⁶². Trois finalités de nature générique sont définies par le texte : simplifier les démarches administratives ; favoriser le recours à l'intelligence artificielle et son développement ; participer à des expérimentations relatives à l'utilisation des données. Ces finalités, extrêmement vagues au demeurant, excèdent largement celles prévues par le décret du 28 mars 2013 sur la DSN. Elles ne

¹⁶¹ Voir l'article L. 6353-10 du code du Travail.

¹⁶² L'arrêté « Champollion » prévoit un champ large de catégories de données pouvant être enregistrées dans le traitement de données en visant à l'article 3, outre celles issues de la DSN, celles enregistrées dans « *des traitements pour lesquels les administrations relevant du champ du travail, de l'emploi et de l'insertion sont responsables de traitement ou destinataires des données* ». Selon l'arrêté, ces catégories de données, collectées directement ou indirectement par la DGEFP, la DGT, la Dares ou la Dnum, sont celles relatives à l'état-civil, la situation familiale, la vie professionnelle et extraprofessionnelle, d'ordre économique, financier et social.

recourent pas non plus l'objet du projet « Champollion », mené au sein de la Dnum par des entrepreneurs d'intérêt général, qui mobilise des données issues de la DSN pour identifier les entreprises ayant un recours abusif aux contrats courts.

[17] **Ce texte, élaboré à l'initiative de la Dnum, avec un faible degré d'association de la DSS, sans consultation préalable de la Dares et de la Cnil¹⁶³, et sans prise en compte de l'ensemble des remarques de la DAJ, soulève différents points d'attention sur le plan juridique :**

- si elle reçoit les flux DSN provenant de la Cnav au titre de sa fonction de direction des SI, la Dnum n'est pas, en droit, destinataire des données individuelles issues de la DSN, ce qui explique que l'arrêté indique que ces données sont collectées « *directement ou indirectement* » par la DGEFP, la DGT, la Dares ou la Dnum ;
- la DAJ estime que, malgré les amendements apportés par la Dnum au projet de texte à la suite des échanges entre les deux directions, les finalités du traitement et les catégories de données, outre qu'elles contreviennent aux dispositions du décret et de l'arrêté « filtre », restent trop larges : en cas de contrôle ou de contentieux, les caractéristiques essentielles du traitement pourraient être considérées comme insuffisamment proportionnées¹⁶⁴ ;
- enfin, la DAJ rappelle que les traitements ultérieurs mis en œuvre par les destinataires des données de « Champollion » devront faire l'objet, le cas échéant, du texte réglementaire et de la formalité adéquats.

1.3 Des règles spécifiques encadrent l'usage des données de la DSN retraitées à des fins d'information générale

1.3.1 L'usage des données de la DSN traitées par la statistique publique obéit à des standards propres, relevant de textes de niveaux national et européen.

[18] **Le service statistique public est soumis aux obligations de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (LIL).** La loi vise à garantir à toute personne physique le contrôle de l'usage des données qui la concernent, et ce faisant, à concilier le développement de l'informatique des données personnelles avec le respect des droits fondamentaux. Elle s'inscrit, depuis 2018, dans un cadre juridique européen harmonisé défini par le Règlement général sur la protection des données (RGPD). Celui-ci s'articule autour de deux principes fondamentaux, un principe de transparence – information des personnes concernées et réponse à leurs demandes, documentation des traitements – et un principe de maîtrise et de limitation des risques d'impact des traitements sur la vie privée des personnes concernées.

[19] Compte tenu de la nécessité d'effectuer les traitements statistiques qui permettent de piloter la qualité et les analyses au niveau individuel, la majorité des informations sur les personnes traitées par la statistique publique ne peuvent pas être anonymisées, mais au mieux

¹⁶³ La DSS a indiqué à la mission avoir été informée par la Dnum et la DAJ de l'existence de ce projet de texte, sans avoir été consultée sur sa version définitive avant sa publication.

¹⁶⁴ Mail adressé par la DAJ à la Dnum le 4 août 2018.

pseudonymisées. Elles relèvent donc dans la grande majorité des traitements de la catégorie des données personnelles. Les résultats statistiques qui sont diffusés une fois les traitements réalisés sont, eux, agrégés et font l'objet d'une pose du secret statistique afin de ne pas identifier des individus. Les données individuelles sont quant à elles diffusées dans le cadre sécurisé du CASD.

[20] **Le secret statistique exclut que des données collectées à des fins statistiques puissent être utilisées ou réutilisées à des fins de décision à l'égard des personnes concernées – personnes physiques comme personnes morales.** En particulier, il exclut strictement toute utilisation ou réutilisation de ces données à des fins de contrôle fiscal ou de répression économique. Il permet d'assurer aux personnes physiques le respect de la confidentialité de leur vie personnelle et familiale, et aux entreprises le respect du secret commercial (les informations transmises ne seront pas mises à la disposition de leurs concurrents).

1.3.2 Les données statistiques sont réutilisables à des fins d'information générale ou de recherche, mais pas *a priori* à des fins de pilotage, même si des réflexions sont en cours pour faire évoluer ce cadre

[21] **Les données protégées par le secret statistique peuvent être communiquées au sein du SSP (Insee et services statistiques ministériels), sans procédure formelle lorsqu'elles visent à produire des statistiques à des fins d'information générale.** Elles peuvent aussi être communiquées –après avis du Comité du secret statistique – en réponse à des demandes formulées à des fins d'information générale, d'évaluation ou de recherche scientifique ou historique d'intérêt public. En pratique, elles sont alors accessibles via le CASD, après habilitation par le Comité du secret (cf. annexe n°1). Par contre, jusque-là, il n'était pas envisagé que ces données soient communiquées à des fins de pilotage des politiques publiques ou plus généralement aux acteurs de la politique de la donnée.

[22] **Dans son délibéré du 31 janvier 2022 sur les relations entre statistiques publiques et politiques de la donnée¹⁶⁵, l'Autorité de la statistique publique (ASP) a cependant souligné la nécessité de faire évoluer ce cadre¹⁶⁶ :** « L'Autorité appelle de ses vœux un cadre de collaboration formalisé entre les services en charge de l'administration des données et les différents SSM, assorti de la part de ces derniers d'une offre de service adaptée au contexte de chaque ministère. Une telle offre de service, mettant à profit l'expertise et l'expérience acquises par les statisticiens publics en matière de structuration et de traitement des données, participe de l'extension de la fonction d'appui à la conduite des politiques ministérielles prévue par la charte des services statistiques ministériels de 2019. L'ASP estime toutefois important que le développement de ces collaborations s'effectue non seulement dans le respect des principes du Code des bonnes pratiques (notamment celui du secret statistique afférent aux données individuelles), mais aussi au bénéfice conjoint du pilotage des politiques ministérielles et du service statistique public, en privilégiant des objectifs comme l'amélioration des concepts et des nomenclatures utilisés, la constitution de bases de

¹⁶⁵ <https://www.autorite-statistique-publique.fr/delibere-sur-les-relations-entre-statistiques-publiques-et-politiques-de-la-donnee/>

¹⁶⁶ Ces exigences valent à la fois pour les informations produites par le système statistique public (SSP) et pour les statistiques labellisées ou reconnues à « visée d'information générale » dans le cadre de la « démarche statistique publique » lancée par l'ASP.

données aisément accessibles et documentées ou le développement de nouvelles potentialités d'exploitation et d'appariement à des fins d'études. »

[23] **Les conditions dans lesquelles les données administratives retraitées par le SSP, comme celles issues de la DSN, peuvent faire l'objet d'une « rétrocession » aux acteurs administratifs ont fait l'objet de réflexions entamées à la mi-2022, notamment suite à l'avis de l'ASP.** Globalement, ces réflexions conduisent à proposer un cadre en deux temps, distinguant les retraitements de nature corrective ou améliorative des données sources et les traitements de nature statistique (notamment ceux qui incorporent des objets couverts par le secret statistique)

2 Les modalités d'accès aux données de la DSN ne répondent pas à tous les besoins des utilisateurs

2.1 Le processus d'instruction des demandes d'accès aux données de la DSN est jugé trop lent

2.1.1 La DSS est chargée d'instruire ces demandes dans le cadre d'un processus imbriqué avec celui de l'évolution de la norme DSN

[24] **Le processus de révision de l'arrêté « filtre » est imbriqué avec les évolutions de la norme DSN.** La DSS, en tant que tutelle du Gip-MDS, instruit et arbitre chaque année, en lien avec le Gip-MDS, les demandes d'évolution de la norme dite NÉODES (Norme d'Échanges Optimisée des Données Sociales) qui détermine les rubriques que doivent renseigner les déclarants en DSN¹⁶⁷. Les évolutions de la norme sont retracées chaque année dans un cahier technique publié un an avant son entrée en vigueur. Pour élaborer chaque nouvelle version de ce cahier, la DSS interroge, avant le mois de mars N-2 par rapport à l'entrée en vigueur d'une nouvelle version du cahier technique, l'ensemble des membres du Gip-MDS sur leurs besoins¹⁶⁸, d'une part, d'évolution de la norme DSN et, d'autre part, d'accès à des données individuelles qui seraient déjà en DSN (l'accès à des données agrégées ne requiert pas d'évolution du cadre juridique). En effet, la prise en compte de ce double besoin potentiel nécessite l'évolution d'un même texte, l'arrêté « filtre », qui est publié après recueil de l'avis de la Cnil. Une même équipe constituée de deux ETP au sein du bureau en charge du recouvrement à la DSS instruit ces deux catégories de besoins selon la même temporalité et constitue le dossier pour la Cnil¹⁶⁹. Les demandes d'accès aux données

¹⁶⁷ La gouvernance de la norme DSN pourrait évoluer dans les prochains mois en lien avec la mise en place d'une maîtrise d'ouvrage stratégique interministérielle, positionnée auprès de la DSS, et de la réactivation du Comité de normalisation des données sociales (CNDS) telles que recommandées par le rapport de l'Igas sur la gouvernance du Gip-MDS (juillet 2022).

¹⁶⁸ La DGEFP coordonne la réponse à cette sollicitation de la DSS pour les organismes et administrations relevant du ministère en charge du Travail.

¹⁶⁹ Ce dossier distingue la transmission de nouvelles données déclarées en DSN ; la transmission de données DSN à des organismes ou administrations qui ne les recevaient pas auparavant ; l'extension des données transmises à des organismes et administrations qui recevaient déjà des données de la DSN ; les précisions apportées aux données de la DSN déjà transmises aux organismes et administrations.

individuelles de la DSN, dont l'instruction représente un enjeu et une charge moindres par rapport à l'évolution de la norme DSN, sont ainsi tributaires de cette dernière.

2.1.2 Le délai de modification de l'arrêté-filtre, indispensable pour accéder à des données individuelles, est jugé trop long par rapport aux besoins

[25] **Le calendrier relativement distendu de révision de l'arrêté « filtre » ralentit l'accès aux données individuelles de la DSN sans que ce délai s'explique par des difficultés inhérentes aux données demandées.** La DSS a indiqué à la mission avoir des difficultés, en raison d'un manque de moyens internes pour instruire les demandes liées à la DSN, à tenir l'objectif d'une révision annuelle de l'arrêté « filtre ». Révisé chaque année entre 2013 et 2016, il n'a ensuite été révisé qu'à deux reprises, en 2019 puis en 2022, augmentant le délai entre la demande initiale et l'accès aux données. À titre d'exemple, la DGT a communiqué à la mission les archives d'échanges de mail entre elle-même, la DSS, et la DGEFP qui traduisent un délai de deux ans entre son expression, relayée en mars 2017 à la DSS par la DGEFP, d'un besoin d'accès à des données individuelles de la DSN, et sa prise en compte, intervenue à la faveur de l'arrêté « filtre » du 7 mars 2019. L'Agence de services et de paiement (ASP) a également transmis à la mission des éléments dont elle a déduit un délai de plus deux ans entre sa communication à la DGEFP en décembre 2019 d'une demande d'accès à plusieurs rubriques de la DSN, et leur prise en compte dans l'arrêté « filtre » du 9 février 2022.

2.1.3 L'objectif de simplification associé à la DSN tend à exclure les évolutions de la norme à des fins d'information générale ou de pilotage

[26] **Sans que la mission ait instruit spécifiquement les besoins d'évolution de la norme DSN en lien avec l'usage de ses données à des fins statistiques ou de pilotage, elle note que ce sujet a été peu soulevé par les interlocuteurs rencontrés.** Ainsi, les représentants de l'Insee ont indiqué à la mission qu'ils jugeaient inenvisageable de demander une évolution de la norme DSN à la seule fin de produire des statistiques¹⁷⁰. De leur côté, la DGEFP et l'ASP ont fait part de leurs difficultés à obtenir un retour de la part de la DSS sur la demande d'ajouter de nouvelles rubriques dans la DSN à des fins de pilotage, en particulier concernant les nouveaux dispositifs de politique publique. Ces remarques font écho au constat posé par l'Igas dans son rapport sur la gouvernance du Gip-MDS quant à la difficulté pour la DSS, en l'absence d'instance collégiale d'échange sur l'évolution de la norme DSN, de prioriser les demandes issues de la sphère travail-emploi, *a fortiori* quand elles ne contribuent pas directement à la simplification des démarches des entreprises.

¹⁷⁰ En cas de besoin, la statistique publique privilégie alors la réalisation d'enquêtes complémentaires sur des échantillons représentatifs, s'appuyant cependant sur les déclarations existantes afin d'alléger la charge de réponse pour les répondants

2.2 La DSS n'est pas en mesure d'instruire et d'aiguiller de façon pertinente les demandes d'accès aux données de la DSN

2.2.1 La DSS n'a pas les moyens ni l'expertise pour instruire correctement les demandes d'accès aux données de la sphère travail-emploi

[27] **En tant que responsable du traitement de données DSN, la DSS assure de facto une fonction de guichet d'accès aux données individuelles ou agrégées issues de cette déclaration.** Cette fonction, imbriquée, comme évoqué *supra*, avec sa mission centrale, celle d'assurer l'évolution de la norme DSN dans le respect des objectifs associés à ce dispositif, est assurée dans une certaine mesure « par défaut ». En effet, la DSS a indiqué à la mission qu'elle considérait que cette fonction ne devrait pas nécessairement lui incomber.

[28] **Le processus d'instruction des demandes d'accès aux données individuelles ou agrégées de la DSN présente plusieurs faiblesses :**

- il repose sur des expressions de besoin de nature et de qualité très hétérogènes : aucun format-type d'expression de besoin n'a été communiqué à la mission, les demandes adressées à la DSS obéissant en pratique à des formats très divers (du *mail* peu détaillé à la note identifiant précisément les flux de données à transmettre, élaborée le cas échéant en lien avec le Gip-MDS) ;
- son calendrier, imbriqué donc, s'agissant des demandes d'accès aux données individuelles, avec celui de l'évolution de la norme DSN, n'est pas encadré : aucun engagement de réponse dans un délai donné n'est affiché, le tableau de suivi des demandes reçues en 2022 communiqué à la mission (cf. tableau ci-dessous) ne précisant d'ailleurs pas les dates de réception des demandes ;
- aucune grille d'analyse permettant d'apprécier la conformité au cadre juridique des demandes d'accès aux données de la DSN n'a été communiquée à la mission, à laquelle la DSS a indiqué qu'elle ne faisait qu'examiner si les demandes d'accès aux données de la DSN sont conformes aux finalités du traitement telles que définies par le décret du 28 mars 2013. L'arbitrage est *in fine* pris par la chefferie de bureau et, pour les demandes plus sensibles, par la sous-direction ;
- les moyens alloués à ce processus, de l'ordre de 2 ETP positionnés dans le bureau en charge du recouvrement, qui assurent également l'évolution de la norme DSN, seraient, selon la DSS, insuffisants pour instruire ces demandes, qui pour certaines proviennent d'acteurs en charge de politiques publiques éloignées du champ d'intervention de la DSS.

[29] **L'examen d'un échantillon de demandes d'accès aux données de la DSN transmises à la DSS en 2022 corrobore à la fois le potentiel de développement des usages de la DSN et les faiblesses de ce processus d'instruction (cf. tableau ci-dessous) :**

- cet échantillon traduit d'abord la variété des usages possibles des données issues de la DSN, au sein du ministère en charge du Travail, qu'il s'agisse d'alimenter des services aux salariés et aux entreprises (passerport compétences, passerport de prévention) ou de suivre les effets d'une politique publique (retour à l'emploi des allocataires du RSA ou des bénéficiaires du Contrat d'engagement jeune), mais aussi à l'extérieur, comme en témoigne les demandes

d'accès aux données adressées par la Direction générale des entreprises (DGE) ou encore par la Mission interministérielle de coordination anti-fraude ;

- si le tableau de bord communiqué par la DSS à la mission n'indique pas les dates de réception des saisines, la mission note que la très grande majorité des demandes restaient, au 12 décembre dernier, en cours d'analyse, y compris pour des saisines datant de plus de six mois selon les informations recueillies par ailleurs par la mission : ainsi la demande de la Caisse nationale de solidarité pour l'autonomie (CNSA) d'accéder aux données individuelles de la DSN pour calculer le taux d'encadrement par Ehpad date du 19 avril 2022, dans le contexte de l'affaire Orpéa ; la demande de la DGE d'utiliser des données agrégées pour alimenter le « *data hub* » sur les aides d'État daterait quant à elle de septembre 2022 ;
- la mission relève enfin qu'au regard du nombre de demandes d'accès aux données agrégées restant au 12 décembre 2022 « en cours d'analyse », ces dernières, qui ne nécessitent pourtant pas de modifier l'arrêté « filtre », ne semblent pas bénéficier d'un traitement accéléré par rapport aux demandes d'accès aux données individuelles, ce dont témoigne le calendrier d'instruction de la demande formulée par la DGEFP d'accéder à des données agrégées pour le suivi des allocataires du CEJ, à la requête pourtant expresse du cabinet du ministre en charge du Travail (l'accord de la DSS a été transmis le 25 juillet 2022 pour une demande communiquée par la DGEFP le 18 janvier 2022).

Tableau 1 : Échantillon de demandes d'accès aux données (individuelles ou agrégées) de la DSN pour un nouvel usage adressées à la DSS en 2022

Organismes	Besoin	Nature des données demandées	Statut de la demande au 12/12/2022
Caisse nationale de solidarité pour l'autonomie (CNSA)	Suivi du taux d'encadrement dans les Ehpad	Individuelles	En cours d'analyse
Data insertion	Suivi du RSA au niveau départemental	Individuelles	Abandonnée
DGEFP	Alimentation du passeport de compétences	Individuelles	En cours d'analyse
DGEFP	Alimentation de la plateforme Agora	Individuelles	En cours d'analyse
DGEFP	Alimentation du passeport de prévention	Individuelles	En cours d'analyse
DGEFP	Réalisation de missions de prospection par les Opco	Individuelles	En cours d'analyse
Direction générale des entreprises (DGE)	Alimentation de la plateforme impact.gouv.fr	Individuelles	En cours d'analyse
Mission interministérielle de coordination anti-fraude (Micaf)	Lutte contre la fraude aux cotisations sociales et le travail illégal	Individuelles	En cours d'analyse

DGE	Alimentation du Data Hub Aides d'État	Agrégées	En cours d'analyse
DGE	Alimentation de l'algorithme de « Signaux faibles »	Agrégées	En cours d'analyse
DGEFP	Suivi du dispositif CEJ	Agrégées	Acceptée
DGEFP	Amélioration de l'algorithme de « La Bonne alternance »	Agrégées	En cours d'analyse
Fonds d'expérimentation contre le chômage de longue durée et territoires	Analyse des taux de retour à l'emploi des allocataires du RSA	Agrégées	En cours d'analyse

Source : Mission, à partir des éléments transmis par la DSS.

2.2.2 La DSS mobilise le Gip-MDS pour répondre aux demandes d'accès aux données sans capacité à orienter vers les bases les plus adaptées aux besoins

[30] **La DSS, qui ne dispose pas d'une vision exhaustive de l'ensemble des bases contenant des données issues de la DSN, et encore moins de leurs avantages et inconvénients respectifs, tend à orienter systématiquement les utilisateurs vers le Gip-MDS dont elle assure la tutelle.** Ces demandes sont d'ailleurs adressées souvent directement au Gip-MDS pour pré-instruction. Le Gip-MDS mobilise la base PIQ (« Projet d'indicateur qualité ») pour les projets impliquant la production d'indicateurs agrégés et la base Dad-E pour ceux nécessitant l'identification au niveau individuel (cf. annexe n° 1). Ces bases sont aussi mobilisées par le Gip-MDS pour le compte de la DSS aux fins de suivi de la montée en charge de certains dispositifs (prime inflation, prime de partage de la valeur ajoutée).

[31] **Le développement des usages de la DSN fait partie des cinq axes stratégiques inscrits dans la feuille de route 2022-2025 du Gip-MDS telle que validée par son conseil d'administration.** Néanmoins, la DSS a indiqué à la mission qu'elle n'estimait pas pertinent que le Gip-MDS alloue à cet axe plus de moyens que ceux qui y sont actuellement consacrés, de l'ordre de quatre ETP, sans justification de ce calibrage et sans que les critères de hiérarchisation des demandes d'accès aux données reposant sur le Gip-MDS soient précisés.

[32] **En pratique, une partie des demandes d'accès aux données de la DSN, y compris aux données individuelles, échappent à l'instruction de la DSS et sont directement prises en charge par un autre acteur.** Ce phénomène s'explique sans doute en partie par le délai de réponse de la DSS aux demandes d'accès. En effet, en l'absence de réponse, une partie des demandeurs se tournent alors vers d'autres acteurs disposant déjà des données issues de la DSN. À titre d'exemple, la mission a relevé que le Gip Inclusion, qui avait au départ adressé à la DSS sa demande d'accès aux données individuelles de la DSN, demande abandonnée selon le tableau de bord daté du 12 décembre dernier, s'est finalement tourné vers la Dnum des ministères sociaux pour accéder à ces données. La responsable de la start-up d'État « Signaux Faibles », qui avait également soumis à la DSS une demande d'accès aux données issues de la DSN, a indiqué à la mission avoir finalement obtenu ces données par l'intermédiaire de la Dares, grâce au soutien au premier trimestre 2022 du cabinet de la ministre en charge du Travail.

3 La stratégie de valorisation des données administratives questionne les modalités de régulation appliquées à la DSN

3.1 Le développement de l'usage des données administratives est devenu un enjeu central pour le pilotage des politiques publiques

3.1.1 L'assouplissement de l'accès aux données administratives est un préalable à leur valorisation

[33] Depuis une quinzaine d'années, et plus récemment sous l'impulsion de l'Union européenne¹⁷¹, le partage des données administratives est encouragé à travers :

- une obligation d'échange entre administrations de « *toutes les informations ou données strictement nécessaires pour traiter une demande présentée par le public ou une déclaration transmise par celui-ci en application d'un texte législatif ou réglementaire* »¹⁷² ;
- l'extension du principe « dites-le nous une fois », sur la base duquel il est possible de « *donner accès aux informations ou données de l'entreprise à tout organisme autorisé à en connaître* »¹⁷³ ;
- l'obligation, sauf exceptions¹⁷⁴, pour les administrations de communiquer les documents administratifs qu'elles détiennent aux autres administrations qui en font la demande pour l'accomplissement de leurs missions de service public¹⁷⁵ ;
- le principe de proactivité de l'administration dans l'utilisation, à des fins d'amélioration des services aux usagers, des informations et données détenues¹⁷⁶.

¹⁷¹ Le règlement UE 2022/868 sur la gouvernance européenne des données (Data Governance Act), adopté le 30 mai 2022 et qui sera applicable en septembre 2023, vise en particulier à favoriser l'utilisation des données détenues par le secteur public.

¹⁷² Loi n°2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit.

¹⁷³ Loi n°2012-387 du 22 mars 2012 relative à la simplification du droit et à l'allègement des démarches administratives.

¹⁷⁴ Loi n°2016-1321 du 7 octobre 2016 pour une République numérique, voir les articles L. 311-5 et L. 311-6. Parmi ces exceptions figurent, par exemple, les documents administratifs dont la consultation ou la consultation porteraient atteinte au secret de la défense nationale ou encore à la conduite de la politique extérieure du Gouvernement.

¹⁷⁵ L'article 14 de cette loi a aussi prévu la mise à disposition, en vue de faciliter leur réutilisation, de jeux de données de référence qui présentent le plus fort impact économique et social. Étalab référence ces jeux de données (qui incluent par exemple la base Sirene des entreprises et de leurs établissements, ou encore le Répertoire opérationnel des métiers et des emplois) sur le portail data.gouv.fr.¹⁷⁶ Voir la loi n°2022-217 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale qui a aussi élargi et assoupli encore les possibilités de partage d'informations et de données entre administrations en supprimant la nécessité de recourir à un décret en Conseil d'État pour les domaines, procédures et données concernés

¹⁷⁶ Voir la loi n°2022-217 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale qui a aussi élargi et assoupli encore les possibilités de partage d'informations et de données entre administrations en

[34] **Néanmoins, ce cadre juridique resterait encore difficile à interpréter et perfectible.** Ainsi, dans son étude sur l'intelligence artificielle et l'action publique réalisée à la demande du Premier ministre¹⁷⁷, le Conseil d'État souligne que l'administration n'est pas propriétaire des données qu'elle détient : *« elle n'est est que la gardienne, et son devoir est de les partager pour qu'en soit tirée toujours plus de valeur, dans l'intérêt général »*, précisant que *« ce partage ne doit avoir d'autre borne que la nécessaire confidentialité de certaines données et, en surplomb, la nécessité de ne pas fragiliser la confiance dans les institutions et collectivités publiques auxquelles les données sont remises par des tiers »*. Considérant que les données publiques sont le *« carburant »* de la mise à profit par l'administration de l'intelligence artificielle, il appelle à une clarification et à un assouplissement du cadre juridique applicable aux données administratives (i) en étendant l'article 1^{er} de la loi pour une République numérique aux données (et non seulement aux documents administratifs formalisés) et (ii) en prévoyant que les secrets protégés par la loi ne soient opposables à l'administration demanderesse *« que si les inconvénients de la communication excèdent l'intérêt public que présente cette communication, eu égard à l'usage que cette dernière entend en faire »*. Enfin, le Conseil d'État propose l'ouverture d'une réflexion *« sur la création d'un régime particulier de partage intra-public facilité de jeux de données de haute valeur identifiés par la Dinum »*.

3.1.2 La valorisation des données administratives contribuerait à la transformation, en particulier numérique, de l'action publique

[35] **Depuis la création en 2011 de la mission Étalab¹⁷⁸, la valorisation des données au service de la transformation publique s'inscrit dans le cadre d'une politique publique de la donnée.** Le rapport de la mission Bothorel¹⁷⁹, en décembre 2020, a souligné la nécessité de relancer cette politique en développant en particulier le partage des données entre acteurs publics, le cas échéant en encourageant la création de *« hubs »* sectoriels ou intersectoriels. Le rapport insiste sur la nécessaire adaptation du rôle de la statistique publique face à l'émergence d'une demande de données publiées en temps réel, en appelant ces acteurs à *« s'emparer de la démarche d'ouverture plutôt que de l'invalider a priori »*, en particulier pour prévenir les *« interprétations fallacieuses »* par les différents utilisateurs. Enfin, le rapport rappelle que *« la qualité d'une donnée n'est pas un absolu »* et se conçoit *« au regard de la finalité des traitements qui la consomment »*.

[36] **Dans la continuité du rapport Bothorel, la circulaire du Premier ministre du 27 avril 2021 affirme l'ambition renouvelée du Gouvernement en matière d'exploitation, d'ouverture et de**

supprimant la nécessité de recourir à un décret en Conseil d'État pour les domaines, procédures et données concernés

¹⁷⁷ « Intelligence artificielle et action publique : construire la confiance, servir la performance », étude du Conseil d'État adoptée en assemblée générale plénière du 31 mars 2022.

¹⁷⁸ Étalab est un département de la Dinum assurant, au titre des missions de l'Administrateur général des données, des algorithmes et des codes sources, la coordination, la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée. Étalab favorise en particulier la diffusion et la réutilisation des informations publiques. Il développe et anime la plateforme d'open data data.gouv.fr destinée à rassembler et à mettre à disposition librement une partie des informations publiques. Ce département est aussi responsable du programme « Entrepreneurs d'intérêt général » qui recrute chaque année pendant dix mois des talents du numérique pour résoudre des défis d'intérêt général aux côtés des administrations.

¹⁷⁹ « Pour une politique publique de la donnée », décembre 2020.

circulation des données¹⁸⁰. Ce texte demande aux administrations de « *rechercher en permanence la meilleure circulation de la donnée* » et de mettre en place dans chaque ministère un administrateur des données chargées d'élaborer une feuille de route dans ce domaine. Les administrateurs des données sont également invités à animer les écosystèmes de réutilisateurs des données et codes sources du ministère et les éventuels contributeurs pour identifier les jeux de données à forte valeur ajoutée, définir d'éventuels standards et alimenter le partage.

3.2 Le projet de *Labor data hub*, qui reste un objet à définir, vise à susciter un usage élargi des données administratives telles que la DSN

3.2.1 Dans le cadre de la feuille de route du MTEI, la création d'un « *hub* » est l'un des vecteurs identifiés pour développer l'usage des données

[37] **Publiée en septembre 2021, la feuille de route des données, des algorithmes et des codes sources du ministère du Travail, de l'Emploi et de l'Insertion (MTEI) a fait l'objet, jusqu'à la fin de la mandature, d'un portage politique fort.** Ce portage s'est traduit en particulier par la réunion, à deux reprises, sous la présidence de la ministre, d'un comité ministériel des données¹⁸¹. Il a aussi conduit à accélérer à la demande de la ministre, toujours selon la Dnum, le lancement de deux projets inscrits dans cette feuille de route :

- d'une part, l'appariement, effectif depuis septembre 2022 en raison du délai lié au conventionnement entre les acteurs, entre les données de Pôle emploi, de la DSN et de la Cnaf afin de reconstituer les trajectoires professionnelles des allocataires de l'assurance chômage et des minima sociaux¹⁸², dans le cadre d'une base dite « MiDAS » hébergée au CASD ;
- d'autre part, un projet d'utilisation de la DSN dit « Champollion », engagé en septembre 2022 avec l'arrivée à la Dnum de trois entrepreneurs d'intérêt général financés en partie par Étalab, visant à aider l'inspection du travail à mieux identifier les entreprises abusant du recours aux contrats courts.

¹⁸⁰ La circulaire concerne également la circulation des algorithmes et des codes sources publics.

¹⁸¹ Ce comité s'est réuni sous la présidence de la ministre le 22 juillet et le 9 novembre 2021. Il s'est réuni une troisième fois, sous la présidence de sa directrice de cabinet, en avril 2022. Malgré les demandes formulées par la mission, la Dnum n'a pas communiqué à la mission de compte rendu de ces réunions.

¹⁸² Dans la sphère statistique, un appariement désigne une grande base de données issues de différents systèmes d'information et dont le rapprochement est permis grâce à des informations communes (ou « identifiants ») sur les individus ou entités suivis. Le CASD joue le rôle de tiers de confiance dans le cadre d'un tel appariement. MiDAS croise trois sources données administratives : les contrats salariés dans MMO et post-Sismmo (données DSN de la Dares) ; l'indemnisation des demandeurs d'emploi (données FHS et FNA de Pôle emploi) ; les bénéficiaires de minima sociaux (données Allstat-FR6 de la Cnaf). La convention, signée en juin 2022, prévoit, à l'automne 2022, une première mise à disposition des données appariées sur la période 2017-2021. Trois vagues se succéderont en 2023. Ensuite, l'appariement se fera sur une fréquence annuelle de 2024 à 2026 avec une profondeur temporelle croissante, couvrant jusqu'à 2017-2025. Voir : [Suivre les trajectoires en emploi salarié des bénéficiaires de l'assurance chômage et des minima sociaux | Dares \(travail-emploi.gouv.fr\)](https://www.dares.fr/actualites/2022/11/la-feuille-de-route-des-donnees-et-des-codes-sources-du-ministere-du-travail-de-l-emploi-et-de-l-insertion-a-fait-l-objet-d-un-portage-politique-fort).

[38] **Cette feuille de route se fonde sur un objectif de développement de l'usage des données, renvoyant à des chantiers ultérieurs le lien à établir avec les besoins métiers des administrations.** Ainsi, concernant la DSN, la feuille de route prévoit deux « *taskforces* » :

- la première est relative au cadrage du « *Labor data hub* », défini comme « *la plateforme de données et de services du travail, de l'emploi et de la formation pour faciliter le partage et promouvoir l'utilisation des données au sein de la sphère sociale (...) il réunit à la fois des éléments techniques (entrepôt), des services et des outils pour former un espace de travail commun autour des données, et plus particulièrement pour creuser les usages de l'intelligence artificielle (...)* », l'objectif fixé étant d'établir le cadre juridique et de choisir la plateforme technique de ce « *hub* » à un an avec, comme cas d'usage prioritaire, l'évaluation de l'assurance chômage à travers l'appariement des données de Pôle emploi, de la DSN et de la Cnaf ;
- la seconde, dédiée à « *la DSN et ses usages* », vise à construire, grâce aux données issues de la DSN, de nouveaux services à destination des entreprises et à créer de nouveaux savoir-faire, notamment avec la Dares, pour « *amplifier l'usage de la donnée dans la conception, le pilotage et l'évaluation des politiques publiques* ».

[39] **Le bilan de la feuille de route, réalisé en octobre 2022, traduit le net essoufflement d'une partie des chantiers relatifs à la valorisation des données.** Si ce document met en avant les progrès réalisés sur des projets spécifiques, comme « Champollion », il indique que différentes actions prévues par la feuille de route restent inabouties, telles que la création d'un catalogue des données et des algorithmes, la mise en place de modalités d'échanges normées et partagées¹⁸³, et celle du « *Labor data hub* ». Le comité ministériel des données, piloté par la Dnum, ne s'est pas réuni entre avril 2022, et fin mars 2023. La nouvelle directrice de la Dnum a indiqué à la mission que l'enjeu serait aujourd'hui de mieux imbriquer les chantiers « données » inscrits dans cette feuille de route avec les besoins métiers des directions et opérateurs du ministère. Des échanges seraient programmés à cette fin entre la Dnum et les parties prenantes à la feuille de route.

3.2.2 En dépit du cadrage réalisé par la Dnum, le *Labor data hub* reste à ce jour un objet administratif non identifié

[40] **À partir de l'objectif fixé par la feuille de route « données », la Dnum, appuyée par un prestataire¹⁸⁴, a réalisé un cadrage du « *Labor data hub* » (LDH) au premier trimestre 2022.** Les documents communiqués par la Dnum à la mission montrent que ce cadrage s'est centré :

- d'une part, sur la recherche de cas d'usages des données, en particulier la DSN, à travers des entretiens réalisés avec les directions métiers du ministère et Pôle emploi ;
- d'autre part, sur l'instruction de différents scénarios d'architecture pour favoriser la mise à disposition de ces données, du plus centralisé (la création d'un « lac de données » sous la responsabilité de la Dnum) au moins centralisé (un « *data mesh* », modèle dans lequel les

¹⁸³ S'agissant de ces deux dernières actions, le document de la Dnum renvoie à la création à venir d'une « *taskforce cartographie* ».

¹⁸⁴ Le prestataire sélectionné par la Dnum, Octo Technology, a facturé cette prestation d'une durée de trois mois au montant de 99 960,00€ TTC.

données conservent chacune leur propre infrastructure avec un partage entre acteurs favorisés par des standards communs et des règles d'interopérabilité).

[41] **Si ce cadrage a permis de mieux identifier les attentes des différentes parties prenantes, il présente des faiblesses qui rendent difficile un arbitrage sur la création du LDH.** Comme le montrent les graphiques ci-dessous issus du document de restitution du cadrage, les entretiens menés ont certes confirmé l'intérêt des directions du ministère pour un usage élargi des données à des fins de pilotage (malgré parfois des interrogations sur la qualité des données) et pour un accès facilité à ces données. Toutefois, l'examen du dossier de cadrage communiqué à la mission soulève aussi plusieurs points d'attention :

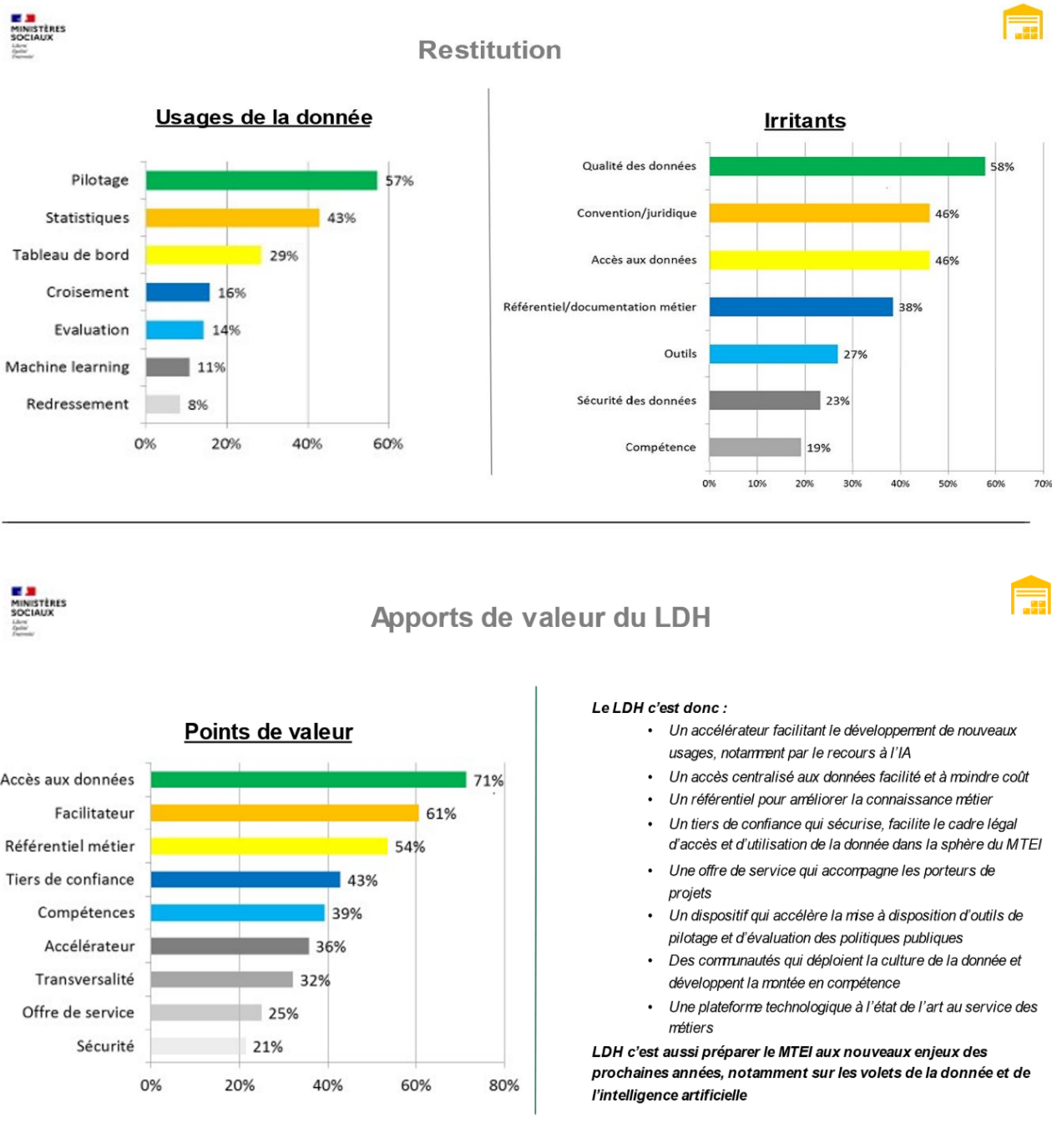
- les comptes rendus d'entretiens, confirmés par les échanges menés par la mission avec ces acteurs, révèlent que ni la DSS, ni le Gip-MDS, ni le CASD n'ont été consultés par l'équipe projet, ce qui contribue à expliquer le fait que le cadrage fait en grande partie l'impasse sur les modalités actuelles d'accès à la DSN, qu'il s'agisse du cadre juridique régulé par la DSS, ou sur l'accès à ces données via les bases sous la responsabilité de la DSS ou du SSP (via le CASD) ;
- si les acteurs rencontrés ont souligné leur souhait d'utiliser davantage les données, dont la DSN, à des fins de pilotage, ils n'identifient pas dans les échanges menés avec l'équipe projet de cas d'usage évidents, ce qui peut traduire un niveau de maturité faible, ou à tout le moins inégal entre les acteurs, sur l'usage de la donnée dans les politiques publiques : ainsi, le cas d'usage initialement identifié pour constituer le « *Minimum viable product* » (MVP)¹⁸⁵ du LDH, qui visait à utiliser plusieurs catégories de données pour aider à identifier les entreprises ayant un recours illégal au travail détaché, s'est avéré finalement ne pas correspondre aux besoins de la DGT ;
- en termes de gouvernance, le cadrage s'est davantage centré sur les questions liées à l'architecture des données (sans faire au préalable une cartographie des différentes bases de données existantes), que sur le sujet, non moins important, du rôle attribuable à chaque acteur sur la base des fonctions attendues du LDH : régulation de l'accès aux données, y compris le cas échéant via l'évolution des textes ; instruction des besoins exprimés par les directions métiers ; aiguillage vers la base de données la plus pertinente ; accompagnement des utilisateurs, y compris de ceux dont les besoins restent latents et qu'il faut faire émerger.

[42] **À défaut d'arbitrage à ce jour, le LDH reste donc un objet administratif non identifié, ce qui questionne aussi son point de départ.** En effet, l'utilisation par la feuille de route du terme de « *hub* » renvoyait sans doute à des modèles extérieurs au ministère, ceux du « *Health data hub* », voir celui du « *Bercy hub* », qui n'avaient pas fait au préalable l'objet d'une instruction. Elle aurait permis de déterminer si ce terme renvoyait à un modèle et, le cas échéant, si celui-ci était transposable au MTEI. La mission relève à cet égard que l'équipe en charge du cadrage du LDH n'a pas rencontré non plus les responsables d'autres « *hubs* » de données ministériels, comme les deux cités précédemment créées en 2019. La Dnum a indiqué à la mission que les travaux relatifs au LDH étaient à ce stade suspendus mais que, s'agissant de son infrastructure, elle excluait dans

¹⁸⁵ Un MVP ou produit minimum viable est la version la plus minimaliste d'un produit conçu avec un minimum de moyens pour tester le produit auprès d'un maximum d'utilisateurs.

tous les cas la création d'un entrepôt unique de données, au profit d'un modèle plus fédéral de type « *data mesh* ».

Graphique 1 : Extraits du support de restitution du cadrage du LDH (part des entretiens réalisés abordant ces items)



Source : Dnum

3.2.3 Les « *hubs* » de données identifiés par la mission dans d'autres ministères confirment la variété des fonctions et modèles d'organisation possibles

[43] La mission n'a pas identifié de modèle de « *hub* » ministériel de données sur la base duquel pourrait être construit le « *Labor data hub* » :

- aucun modèle « type » de « *hub* » n'a été diffusé auprès des administrateurs de données, par exemple par la Dinum ;
- aucune cartographie, et *a fortiori* aucune analyse comparative, des « *hubs* » existants n'a été identifiée par la mission ;
- aucun exemple à l'étranger de « *hub* » de données administratives sur l'emploi n'a été communiqué non plus par les interlocuteurs interrogés à ce sujet¹⁸⁶.

[44] Les trois exemples de « *hub* » identifiés par la mission traduisent bien l'hétérogénéité des objets que ce terme tend à masquer :

- le « **Health data hub** », créé sous la forme de Gip, vise à favoriser l'accès aux données de santé aux projets d'intérêt public, y compris ceux portés par des acteurs privés : si l'objectif était initialement que les données soient hébergées par une plateforme gérée par ce « *hub* », les données restent à ce stade dans des bases accessibles sous la responsabilité de l'assurance maladie et du CASD¹⁸⁷ ;
- le « **Bercy hub** », sous la responsabilité de l'administrateur général des données, positionné auprès du secrétariat général des ministères économiques et financiers, joue essentiellement un rôle d'animation en matière de valorisation des données des différentes directions, lesquelles présentent des moyens et un niveau de maturité hétérogènes dans ce domaine : une plateforme dite « Nubonyxia » qui, sans héberger les données, fournirait un espace sécurisé pour les valoriser, est en cours de création ;
- enfin, le projet « **Innovations, données et expérimentations en éducation** » (IDEE), porté par le service statistique du ministère en charge de l'Éducation nationale, se présente aussi comme un « *hub* », plutôt ciblé sur le monde de la recherche, en ce qu'il vise à favoriser l'accès aux données administratives détenues par le ministère via un centre sécurisé, avec un accent mis sur leur documentation et leur structuration ainsi que sur l'animation de la communauté des utilisateurs.

¹⁸⁶ La mission a en particulier interrogé le département de l'Insee en charge des relations internationales.

¹⁸⁷ Le « *Health data hub* » (HDH) joue néanmoins le rôle de guichet unique pour l'accès aux données. Un comité éthique et scientifique (Cesrees) est sollicité pour avis. Selon le HDH, une moyenne de 300 projets sont soumis au Cesrees chaque année. Le projet initial de créer une plateforme permettant d'héberger de façon centralisée l'ensemble des données de santé a été mis en suspens en 2020 en raison de la recherche d'un hébergeur autre que le « *cloud* » de Microsoft qui soulevait des questions en termes de souveraineté.

3.3 En pratique, le CASD constitue un « *hub* » intersectoriel d'accès à des données dont une partie sont issues de la DSN

[45] **S'il s'adresse notamment aujourd'hui aux chercheurs pour des travaux d'études et à des fins d'évaluation, le CASD n'en est pas moins un « hub » de données qui pourrait être utilisé à des fins plus larges.** Créé en 2010 à la suite de la modification de la loi n°51-711 du 7 juin 1951 permettant l'utilisation par les chercheurs des données d'entreprises, le CASD est un Gip assurant l'hébergement, la fourniture de logiciels experts en traitement de données et la mise à disposition de ces dernières, via une procédure payante de boîtier informatique (SD-Box) et de bulle sécurisée garantissant le contrôle préalable de toutes les données transmises aux utilisateurs authentifiés. Le CASD offre aussi la possibilité d'apparier les données issues de la DSN avec d'autres bases de données, en particulier les bases fiscales (cf. annexe n°2). Selon les données communiquées par le CASD, les données issues de la DSN sont les plus mobilisées par les utilisateurs, avec plus de deux cents projets de recherche en cours et plus de 1 000 utilisateurs.

[46] **Bien que le CASD mette à disposition des chercheurs une documentation sur les données et organise des formations en lien avec les producteurs de données, il ne dispense pas cependant à proprement parler une animation collective.** Il met cependant en relation les chercheurs avec les producteurs en cas de demande. Un chercheur a aussi indiqué souhaiter que le CASD mette à leur disposition des données brutes de la DSN qui leur permettraient en particulier d'accéder à des variables absentes des bases statistiques pour enrichir ses travaux.

3.4 L'assouplissement des modalités d'accès aux données administratives doit être concilié avec les règles relatives à la protection des données

3.4.1 Les utilisateurs des données DSN sont chacun responsables de la conformité des usages au RGPD

[47] **Peu abordée par les travaux de cadrage du LDH, la protection des données personnelles constitue pourtant un enjeu à concilier avec le développement des usages de la DSN.** Si la sensibilité des données issues de la DSN n'est pas du même ordre que celle des données de santé, elles n'en restent pas moins des données personnelles. De plus, au sein de ces données, celles relatives par exemple à la rémunération, aux accidents du travail, aux indemnités journalières ou encore au statut de travailleur handicapé, doivent faire l'objet d'une attention particulière. Comme évoqué *supra*, le RGPD repose sur une logique de responsabilisation des acteurs en charge de traitement de données. Les responsables de ces traitements doivent se conformer au RGPD à travers la rédaction d'une analyse d'impact relative à la protection des données (AIPD), et d'apprécier s'ils en respectent les principes, en particulier¹⁸⁸ :

- le principe de finalité : le responsable d'un fichier ne peut enregistrer et utiliser des informations sur des personnes physiques que dans un but bien précis, légal et légitime ;

¹⁸⁸ <https://www.cnil.fr/fr/cnil-direct/question/quels-sont-les-grands-principes-des-regles-de-protection-des-donnees>.

- le principe de minimisation : les données à caractère personnel collectées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées ;
- le principe d'une durée de conservation limitée : il n'est pas possible de conserver des informations sur des personnes physiques dans un fichier pour une durée illimitée ;
- le principe de sécurité et de confidentialité : le responsable du fichier doit garantir la sécurité et la confidentialité des informations qu'il détient, et décrire comment il met en œuvre la sécurisation des données.

[48] **En dehors des contrôles qui peuvent être effectués par la Cnil, aucune régulation ex post de la conformité des traitements de données issues de la DSN n'est mise en œuvre.** La déléguée à la protection des données (DPO¹⁸⁹) a indiqué à la mission qu'une partie seulement de ces traitements était inscrite au registre des traitements des ministères sociaux prévu par le RGPD. La DPD a aussi souligné qu'elle n'avait pas de visibilité sur les AIPD éventuellement produites en cas d'usage des données issues de la DSN. Elle a relevé à cet égard que l'AIPD relative au traitement DSN lui-même, tel que prévu par le décret du 28 mars 2013, restait à date inachevée.

3.4.2 En pratique, la mission a constaté des standards variables selon les utilisateurs des données issues de la DSN

[49] **Alors que la Dares et l'Insee réalisent de façon systématique des AIPD et recourent à des tiers de confiance, en l'espèce le CASD, pour appairer la DSN avec d'autres données, la Dnum et le Gip-MDS s'alignent sur des standards moins exigeants.** Ainsi, le projet « Champollion » ne fait pas l'objet à date d'une AIPD. La Dnum a aussi procédé directement à l'appariement de listes d'allocataires du RSA transmises par trois Conseils départementaux via le Gip Inclusion avec des données issues de la DSN. Le Gip-MDS a de son côté apparié des listes de bénéficiaires du CEJ communiquées par la DGEFP avec les données issues de la DSN.

¹⁸⁹ Data protection officer.

LISTE DES PERSONNES RENCONTREES

Services du Premier ministre

Direction interministérielle du numérique

M. Pierre Pezziardi, conseiller auprès de la directrice

France Stratégie

M. Cédric Audenis, commissaire général adjoint

M. Antoine Naboulet, directeur adjoint du département Travail, emploi, compétence

M. Clément Dherbécourt, chef de projet, département Société et politiques sociales

Haut Conseil au financement de la protection sociale

M. Dominique Libault, président

Mme Nathalie Guilhembet, secrétaire générale

Secrétariat général des ministères sociaux

Direction du numérique

Mme Anne Jeanjean, directrice du numérique, administratrice ministérielle des données pour le champ travail/emploi

M. Vincent Gaillot, directeur de cabinet

M. Jérémie Vallet, sous-directeur aux projets et services numériques

M. Laurent Bastide, chef de la mission transformation numérique

M. Xavier Béchade, directeur de projets, adjoint au chef du bureau de la valorisation des données

Mme Éléonore Brizay, conseillère transformation numérique

Service à compétence nationale en charge du SI des ARS

M. Mark Hallak, chef du projet Diamant

Mme Annick Le Pape, responsable du département aide à la décision

Direction des affaires juridiques

Mme Pearl Nguyễn Duy, directrice

Mme Romy Bouyahia, adjointe au chef de bureau du droit du numérique et des données

Mme Danièla Parrot, déléguée à la protection des données des ministères sociaux

Ministère du Travail, du Plein Emploi et de l'Insertion

Cabinet du ministre

M. Philippe Zamora, conseiller marché du travail et assurance chômage

Haut-commissariat à l'emploi et à l'engagement des entreprises

M. Thibaut Guilluy, haut-commissaire, préfigurateur de France Travail

M. Thibaut Romatet, directeur de cabinet

Délégation générale à l'emploi et à la formation professionnelle

M. Samuel Berger, sous-directeur du financement et de la modernisation

M. Boris Supiot, adjoint au sous-directeur

M. Olivier Barrat, chef de la mission du pilotage et de la performance

M. Vincent Plouvier, chef de projet DSN

M. Sébastien Vassord, directeur de projet « Agora »

M. Laurent Gaullier, adjoint au sous-directeur des politiques de formation et du contrôle

M. Guillaume Jaspart, chef de projet « réforme de la collecte de la formation professionnelle »

Direction générale du Travail

Mme Annaïck Laurent, directrice adjointe

Mme Catherine Lissarague, cheffe de la mission coordination des systèmes d'information

M. Hervé Legrand, chef de projet SITERE

M. Yoann Blanluet, chargé d'études au bureau du pilotage du SI d'inspection du travail

Mme Laëtitia Collombet, chargée d'études statistiques au bureau du pilotage du SI d'inspection du travail

M. François-Pierre Constant, adjoint à la cheffe de bureau DASIT 1

Direction de l'animation de la recherche, des études et des statistiques

M. Michel Houdebine, directeur

Mme Anne-Juliette Bessone, cheffe de service, adjointe au directeur

M. Dorian Roucher, sous-directeur de l'emploi et du marché du travail

M. Bertrand Marc, chef du département Emploi

Mme Alexandra Louvet, cheffe du département Formation professionnelle et alternance

Mme Claire-Lise Dubost, adjointe au chef de département de l'insertion professionnelle

Ministère de la santé et de la prévention

Direction de la sécurité sociale

M. Franck Von Lennep, directeur

M. Morgan Delaye, chef de service

M. François Godineau, expert de haut niveau

M. Laurent Lenière, adjoint à la sous-directrice du pilotage du service public de la sécurité sociale

Mme Caroline Pereira, cheffe du bureau du recouvrement

Mme Eléa Giraud, chargé de mission pilotage DSN

M. Naël Desaldeleer, chargé de mission – chef de projet pilotage de la DSN

Direction générale de l'offre de soins

M. Philippe Charpentier, sous-directeur des ressources humaines du système de santé

Mme Nathalie Gouge, chargée de mission « politique statutaire et indemnitaire de la FPH »

Mme Isabelle Pinto, chargée d'études et de recherche

Mme Marie-Odile Thévenon, cheffe du bureau « données de pilotage et aide au pilotage »

M. Frédéric Albert, responsable « méthodes et outils au service de la décision »

M. Alexis Sagiroglu, chargé de mission

Direction de la recherche, des études, de l'évaluation et des statistiques (Drees)

M. Fabrice Lengart, directeur

M. Benoît Ourliac, sous-directeur de l'observation de la santé et de l'assurance maladie

M. Olivier Léon, adjoint au sous-directeur de l'observation de la solidarité

M. Claude Gissot, directeur de projet AMDAC

Health data hub

Mme Stéphanie Combes, directrice

Ministère des solidarités, de l'autonomie et des personnes handicapées

Direction générale de la cohésion sociale

Mme Florence Allot, cheffe de service, adjointe au directeur

M. Denis Darnand, sous-directeur en charge de l'inclusion sociale, de l'insertion et de la lutte contre la pauvreté

M. Julien Chauvel, adjoint à la cheffe de bureau emploi et politiques salariales

M. Maxime Laroque, chargé de mission

M. Gérald Maubuleau, adjoint au chef de la mission des SI

Autorité de la statistique publique (ASP)

Mme Mireille Elbaum, présidente

Mme Pascale Pollet, rapporteur

M. Alain Maurand, chargé de mission

Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique

Institut national de la statistique et des études économiques (Insee)

Mme Christel Colin, directrice des statistiques démographiques et sociales

M. Vladimir Passeron, chef du département de l'emploi et des revenus d'activité

M. Jean-François Eudeline, chef adjoint du département

Mme Sylvie Lagarde, directrice de la méthodologie et de la coordination statistique et internationale

Mme Sylvie Eghbal-Teherani, cheffe de la mission de la coordination statistique

Conseil National de l'Information Statistique (Cnis)

Mme Françoise Maurel, secrétaire générale

M. François Guillaumat-Tailliet, secrétaire général adjoint

Direction générale des finances publiques (DGFIP)

M. Denis Boissault, chef du département des études statistiques fiscales et du SSM Finances publiques

Secrétariat général

M. Guillaume Coldre, AMDAC, responsable du « Bercy Hub »

Centre d'accès sécurisé aux données (CASD)

M. Kamel Gadouche, directeur

Ministère de l'Éducation nationale

Direction de l'évaluation, de la prospective et de la performance

Mme Fabienne Rosenwald, directrice

M. Thierry Rocher, adjoint au sous-directeur des évaluations et de la performance scolaire

Gip-MDS

Mme Elisabeth Humbert-Bottin, directrice générale

M. Nicolas Simon, directeur des offres et de la transformation digitale

M. Arnaud Vasserot, responsable des projets Data

Urssaf CN

M. Alain Gubian, directeur des statistiques, des études et de la prévision

M. Cyrille Hagneré, adjoint au directeur

Mme Anne-Laure Zennou, cheffe du département des études statistiques et animation du réseau

M. Maël Buron, adjoint à la cheffe de département,

Mme Gaëlle Priou

Unédic

Mme Lara Müller, directrice des études et analyses

M. Vincent Roberti, directeur des services numériques et de la stratégie de la donnée

Mme Clémence Taillet, cheffe de cabinet du DG

M. Baptiste Constanzo, chargé de mission

Cnam – direction des risques professionnels

Mme Anne Thiebeauld, directrice des risques professionnels

M. Pascal Jacquetin, directeur adjoint

Mme Marie-Claire Pirot, juriste

Caisse des dépôts et consignations

Mme Laure de la Bretèche, directrice déléguée des politiques sociales

M. Ronan Mahieu, directeur des études et des statistiques

M. Philippe Treilhou, responsable d'études économiques et financières

Agence de services et de paiement

Mme Sophie Chaize-Pingaud, directrice de l'Emploi, de l'Environnement et des Politiques Sociales

M. Philippe Nauleau, directeur adjoint

Mme Claire Amelin, chargée de mission contrôle interne

Mme Christine Roux, directrice de projet DSN

Mme Claire Magne, chargée de mission contrôle interne

Agefiph

M. Didier Eyssartier, directeur général

M. Hugues Defoy, directeur de la mobilisation du monde économique et social

Mme Véronique Bustreel, directrice de l'innovation, évaluation et stratégie

Catherine Larramendy, chargée de mission

Mme Lydie Macrez, directrice administrative et financière

Direction générale de Pôle emploi

Mme Charline Nicolas, directrice générale adjointe en charge de la stratégie et des affaires institutionnelles

M. Cyril Nouveau, directeur des statistiques, des études et de l'évaluation

M. Richard Ruot, directeur en charge du développement et de l'ancrage des pratiques au sein de la direction de l'offre de services

M. Stéphane Campion, responsable des responsables extérieures du programme « Intelligence Emploi »

France compétences

M. Marc-Antoine Estrade, directeur de l'observation et évaluation

Mme Aline Valette-Wursthén, Coordinatrice de l'animation des observatoires de branche

CCMSA

M. Sébastien Grippi, directeur Métiers entreprises recouvrement

Mme Nadia Joubert, directrice des statistiques, des études et des fonds

M. Marc Parmentier, chef du département Cotisations, tarifications et fonds

Mme Karine Danthez, cheffe de service Cotisation et tarifications

M. David Foucaud, chef du département Synthèses

M. Alexis Guyonvarch, chef du service Data et publications

Institut des politiques publiques (IPP)

Antoine Bozio, directeur

OPCO et observatoires de branches

Atlas

M. Stéphane Phan, responsable du département « prospective et projets innovants »

M. Antoine Decaris, directeur « ingénierie et innovation »

M. Stéphane Nicolaï, responsable « collecte »

Uniformalion

Mme Elisa Braley, adjointe du directeur de l'appui aux branches professionnelles, responsable du pôle projets et études

Mme Dorothée Garnier, responsable du service « statistiques »

Akto

Mme Sandra Laurol, responsable du pôle « observatoires et certifications »

M. Simon Cantarel, chef de projets études et statistiques

Opcommerce

Mme Armelle Bechieau, directrice des branches et de l'observatoire

Mme Fanny Coste, ex-responsable de l'observatoire

Mme Florence Mange, responsable de l'observatoire

Réseau des Carif-Oref

M. Ludovic Bertrand, directeur

Mme Ophélie Costenoble, *data analyst*

Start-up d'État/projets de plateforme

La Bonne alternance

Mme Anne Becquet, directrice de la mission interministérielle pour l'apprentissage

M. Léo Radisson, product manager

Signaux Faibles

Mme Élodie Quezel, directrice opérationnelle

Data hub aides d'État

M. Vincent Dortet, responsable du projet

Plateforme Impact

M. Aurélien Claeys, intrapreneur « impact »

Diagoriente

M. Vincent Sommé, responsable produit

M. Alexis Akinyemi, docteur en psychologie sociale

Immersion facilitée

Mme Nathalie Reyre, *product owner*

Gip Plateforme de l'inclusion

M. Arnaud Denoix, directeur

Place des entreprises

M. Mathieu Gens, directeur

Mme Adeline Latron, cheffe de produit

Invenis

M. Pierre-Louis Picot, Account Manager

M. Edouard Nahama, Data analyst, en charge de Ramses

LETTRE DE MISSION



**MINISTÈRE
DU TRAVAIL,
DU PLEIN EMPLOI
ET DE L'INSERTION**

*Liberté
Égalité
Fraternité*

Benjamin MAURICE

Le Directeur du cabinet

Paris, le **07 OCT. 2022**

Nos Réf. : D-22-021209

Objet : Usages statistiques et à des fins de pilotage de la DSN dans le champ du travail, de l'emploi et de la formation professionnelle

Monsieur le Chef de l'Inspection générale des Affaires sociales,
Monsieur le directeur général de l'INSEE,

La déclaration sociale nominative (DSN) s'est mise en place progressivement suite à la loi du 22 mars 2012. Gérée par le GIP-MDS (modernisation des déclarations sociales), elle a permis de remplacer l'ensemble des déclarations périodiques ou événementielles et diverses formalités administratives adressées précédemment par les employeurs à une diversité d'acteurs (CPAM, Urssaf, Agirc-Arrco, organismes complémentaires, Pôle emploi, Centre des impôts, Caisses des régimes spéciaux, etc.). À compter de 2022, son extension à la fonction publique contribue à compléter de façon importante le champ de l'emploi salarié en France traité par la DSN.

Outre son usage premier pour gérer les comptes individuels en matière de protection sociale, la DSN devient progressivement le socle des statistiques d'emploi et de salaires du service statistique public (SSP). L'INSEE et la DARES ont ainsi constitué deux systèmes d'information statistiques principaux appuyés sur la DSN (SIERA et SISMMO) pour alimenter l'ensemble des besoins de connaissance statistique, qu'ils soient conjoncturels ou structurels, sur les stocks ou sur les flux, ou pour enrichir d'autres bases de données. Depuis 2017 l'ACOSS est également destinataire des DSN au titre de ses missions de recouvrement des cotisations sociales, et co-produit avec l'INSEE et la DARES les estimations trimestrielles d'emploi.

En outre, la feuille de route numérique 2021-2023 du ministère du travail envisage la DSN comme « un gisement de données transverses au champ travail, emploi, insertion, qui peut contribuer encore plus largement en matière de pilotage de politique publique, d'effectivité du recours aux aides publiques ou encore de simplification administrative ».

Monsieur Jean-Luc TAVERNIER
Directeur général de l'INSEE
18 BOULEVARD ADOLPHE PINARD
75675, PARIS CEDEX 14

Monsieur Thomas AUDIGÉ
Chef de l'inspection générale des affaires sociales
39 43 QUAI ANDRÉ CITROËN
75739, PARIS CEDEX 15

Tél : 01 40 50 80 00
127 rue de Grenelle - 75350 PARIS 07 SP

Le traitement de vos données est nécessaire à la gestion de votre demande et entre dans le cadre des missions confiées aux ministères sociaux.
Conformément au règlement général sur la protection des données (RGPD), vous pouvez exercer vos droits à l'adresse ddc-rsps-cab@ddc-social.gouv.fr ou par voie postale.
Pour en savoir plus : <https://travail-emploi.gouv.fr/ministere/articulation-des-personnelles-et-cookies>

Par la dimension de ce gisement de données, la DSN suscite de nombreuses attentes des administrations ou d'autres organismes par exemple pour produire des indicateurs de performance et de suivi des dispositifs de politique d'emploi ou de formation, ou encore pour améliorer le ciblage de mesures ou de contrôles administratifs. S'agissant de données individuelles à caractère personnel comportant le NIR, le tri des données fournies à chacun des organismes est subordonné à ses missions propres et ses usages doivent être conformes au règlement général de protection des données (RGPD), notamment dans le cadre de la réalisation d'appariements avec d'autres jeux de données. L'arrêté du 9 février 2022 fixe ainsi la liste des variables des déclarations DSN transmises aux administrations et organismes compétents.

Outre cette diffusion des déclarations individuelles, le GIP-MDS produit des tableaux de bord et indicateurs agrégés pour répondre à des demandes émanant d'acteurs divers, comme par exemple des fédérations professionnelles.

Nous souhaiterions qu'une mission conjointe soit diligentée afin, dans le champ du travail, de l'emploi et de la formation professionnelle :

- d'établir un état des lieux des usages de la DSN au sein du système de statistique publique ;
- d'identifier les principaux indicateurs produits sur la base de la DSN par les divers acteurs y ayant recours, à des fins d'information générale - et donc susceptibles d'être examinés par l'Autorité de la statistique publique - ou de pilotage opérationnel ;
- d'étudier le cadre d'utilisation de ces données à des fins de pilotage, en particulier par les directions d'administration centrale chargées de l'animation des politiques publiques travail, de l'emploi et de la formation professionnelle ;
- la question de la mise à disposition d'un public large ou restreint, à travers par exemple le Labor Data Hub sera également prise en compte ;
- d'analyser l'adéquation entre les besoins et l'offre du SSP et celle des autres opérateurs.

Les conclusions sont attendues à la fin de l'année 2022.

Je vous remercie d'accepter cette mission, et vous prie de croire, Messieurs, à l'assurance de ma considération distinguée.



Benjamin MAURICE