

# Commission nationale de l'informatique et des libertés

**Délibération n° 2015-162 du 4 juin 2015 portant avis sur un projet de décret en Conseil d'Etat relatif à la création d'un traitement de données à caractère personnel concernant un programme de surveillance de la santé des travailleurs dénommé « COSET » (demande d'avis n° 15006009)**

NOR : CNIX1526476X

La Commission nationale de l'informatique et des libertés,

Saisie par le ministère des affaires sociales, de la santé et des droits des femmes d'une demande d'avis concernant un projet de décret en Conseil d'Etat relatif à la création d'un traitement de données à caractère personnel concernant un programme de surveillance de la santé des travailleurs dénommé « COSET » ;

Vu la convention n° 108 du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel ;

Vu la directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données ;

Vu le code de la sécurité sociale, notamment son article L. 161-28-1 ;

Vu la loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, notamment son article 27-1 (1°) ;

Vu le décret n° 98-37 du 16 janvier 1998 autorisant l'accès aux données relatives au décès des personnes inscrites au répertoire national d'identification des personnes physiques dans le cadre des recherches dans le domaine de la santé ;

Vu le décret n° 2005-1309 du 20 octobre 2005 modifié pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ;

Vu la décision DR-2010-321 du 19 octobre 2010 autorisant l'Institut de veille sanitaire à mettre en œuvre un traitement de données ayant pour finalité l'accès aux données des systèmes d'information existant dans le cadre de la phase pilote du projet COSET-MSA ;

Vu la décision DR-2012-091 du 11 avril 2012 autorisant l'Institut de veille sanitaire à mettre en œuvre un traitement de données ayant pour finalité la mise en œuvre d'un pilote dans le cadre du programme COSET (Cohortes pour la surveillance épidémiologique en lien avec le travail), programme national pour la surveillance épidémiologique en milieu du travail, portant sur l'inclusion d'actifs relevant du régime social des indépendants ;

Vu la décision DR-2014-471 du 28 octobre 2014 autorisant l'Institut de veille sanitaire à mettre en œuvre un traitement de données ayant pour finalité le suivi de cohortes pour la surveillance épidémiologique en lien avec le travail (COSET) visant à décrire l'état de santé des actifs relevant de la Mutualité sociale agricole (COSET-MSA) et du Régime social des indépendants (COSET-RSI) à l'échelle nationale ;

Après avoir entendu M. Alexandre LINDEN, commissaire, en son rapport et M. Jean-Alexandre SILVY, commissaire du Gouvernement, en ses observations,

Emet l'avis suivant :

Le ministère des affaires sociales, de la santé et des droits des femmes a saisi la commission, le 26 février 2015, d'une demande d'avis relative à un projet de décret en Conseil d'Etat (ci-après le « projet ») portant création d'un traitement automatisé de données à caractère personnel relatif à un programme de surveillance de la santé des travailleurs dénommé « COSET ».

Ce traitement à des fins de recherche dans le domaine de la santé portant sur des données parmi lesquelles figure le numéro d'inscription des personnes au répertoire national d'identification des personnes physiques ou « NIR », relève, à ce titre, d'un décret en Conseil d'Etat, pris après avis motivé et publié de la CNIL, en application des dispositions de l'article 27-1 (1°) de la loi du 6 janvier 1978 modifiée.

La CNIL a également été saisie, à l'appui du projet, d'un protocole de recherche portant notamment sur les modalités d'organisation de l'étude et de collecte des données individuelles des personnes concernées, issues du Système national d'information interrégimes de l'assurance maladie (SNIIRAM) mis en œuvre par la Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS) et du Système national de gestion des carrières (SNGC) mis en œuvre par la Caisse nationale d'assurance vieillesse (CNAV), impliquant l'utilisation du NIR.

## **Présentation et conditions de mise en œuvre de l'étude :**

Le programme de cohortes COSET, mis en œuvre par l'Institut de veille sanitaire (InVS) a pour objectifs de décrire l'état de santé des personnes actives selon leur activité professionnelle et d'en analyser l'évolution dans le temps. Il repose sur l'exploitation de données de cohortes d'affiliés aux trois principaux régimes de sécurité sociale que sont le régime général, le régime de la mutualité sociale agricole (MSA) et le régime social des indépendants (RSI).

Concernant le régime général, le programme COSET s'appuiera sur les données des actifs de la cohorte « Constances », en cours de mise en place par l'université de Versailles - Saint-Quentin-en-Yvelines et l'Institut national de la santé et de la recherche médicale (INSERM) en collaboration avec la CNAMTS, qui prévoit de recruter un échantillon de 200 000 personnes sur cinq années afin d'étudier les déterminants professionnels et sociaux de santé.

Concernant les actifs affiliés au régime agricole et au régime des indépendants, l'InVS suit deux cohortes de personnes tirées au sort parmi chacun de ces régimes la cohorte COSET-MSA et la cohorte COSET-RSI.

La cohorte COSET-MSA est constituée à partir d'un échantillon d'environ 35 000 personnes couvrant l'ensemble des professions et des secteurs d'activité du monde agricole. Elle concerne l'ensemble de la population active, âgée de 18 à 65 ans, affiliée à la MSA en France qu'elle soit en activité ou en période d'inactivité (chômage par exemple), salariés agricoles y compris du secteur tertiaire et non salariés (exploitants agricoles, conjoints collaborateurs, aides familiaux), quels que soient la catégorie socio-professionnelle, le secteur d'activité et le type de contrat de travail. Les personnes participant à l'étude feront l'objet d'un suivi postérieurement à la cessation de leur activité professionnelle.

La cohorte COSET-RSI est constituée à partir d'un échantillon d'environ 30 000 personnes couvrant l'ensemble des professions indépendantes relevant du RSI. Elle concerne l'ensemble de la population active, âgée de 18 à 65 ans, affiliée à ce régime en France qu'elle soit effectivement en activité ou bien dans une période d'inactivité y compris les conjoints collaborateurs, quels que soient la catégorie socio-professionnelle, le secteur d'activité et la situation de travail. Les personnes participant à cette cohorte feront également l'objet d'un suivi postérieurement à la cessation de leur activité professionnelle.

Les deux cohortes suivent la même organisation, à savoir d'abord une phase d'inclusion lors de laquelle un courrier postal est envoyé au domicile des personnes sélectionnées par tirage au sort, afin de leur communiquer leur code confidentiel et leur mot de passe temporaire nécessaires à leur inscription sur l'application hébergeant le questionnaire en ligne. A des fins de suivi de l'état de santé et du parcours professionnel, les personnes seront ensuite sollicitées périodiquement par courriel, éventuellement suivi d'un envoi postal, afin de répondre à un questionnaire de suivi en ligne.

Par ailleurs, d'autres informations seront recueillies indirectement, auprès des organismes d'assurance maladie (remboursements de médicaments et consultations, hospitalisations, accidents du travail et maladies professionnelles) et d'assurance vieillesse (périodes d'emploi, profession, type d'activité), concernant les répondants au questionnaire d'inclusion ainsi que 60 à 70 000 personnes tirées au sort parmi les non-répondants au questionnaire.

S'agissant d'un traitement de données à caractère personnel consistant en une recherche dans le domaine de la santé, les cohortes COSET sont soumises aux formalités prévues au chapitre IX de la loi du 6 janvier 1978 modifiée. Ainsi, la commission a d'abord autorisé la réalisation de deux études pilotes en 2010 pour COSET-MSA (décision n° DR-2010-321 du 19 octobre 2010) et en 2012 pour COSET-RSI (décision n° DR-2012-091 du 11 avril 2012). A la suite de cette phase pilote, l'extension nationale des cohortes a été autorisée par la commission dans sa décision n° DR-2014-471 du 28 octobre 2014.

L'InVS souhaite dorénavant pouvoir procéder à l'appariement des données des cohortes COSET-MSA et COSET-RSI avec le SNIIRAM et le SNGC aux fins de suivi « passif » des personnes, grâce au traitement de leur NIR.

#### **Sur la dénomination et la finalité du traitement :**

L'article 1<sup>er</sup> du projet précise que le traitement « *relatif au programme de cohortes nationales pour la surveillance épidémiologique en lien avec le travail dénommé "COSET" [...] a pour finalité de décrire l'état de santé des travailleurs actifs selon leur activité professionnelle et d'analyser son évolution dans le temps.* »

La commission considère que ces finalités sont déterminées, explicites et légitimes au sens de l'article 6-2 de la loi Informatique et Libertés.

#### **Sur la nature des données traitées :**

L'article 2-II du projet précise les catégories de données à caractère personnel enregistrées.

Il s'agit, d'une part, des données collectées directement auprès des personnes participant à l'étude, et, d'autre part, des données collectées au sein des bases de données existantes.

A l'inclusion dans les cohortes COSET-MSA et COSET-RSI, les données d'identification suivantes sont collectées auprès des bases nationales des régimes d'assurance maladie MSA et RSI : nom de famille, nom d'usage, prénoms, sexe, mois et année de naissance, adresse postale.

Les informations suivantes seront ensuite recueillies grâce à un questionnaire autoadministré :

- l'adresse électronique afin de recontacter la personne dans le cadre du suivi ;
- les données socio-démographiques (sexe, mois et année de naissance, nationalité) ;
- les données de santé relatives à l'état de santé général (poids, taille, antécédents personnels, pathologies prévalentes) ainsi que la situation de la personne au regard de pathologies particulières (asthme, troubles musculo-squelettiques, syndromes dépressifs) et ses comportements de santé (consommation de tabac et d'alcool) ;
- situation familiale et situation professionnelle du conjoint ;
- données professionnelles décrivant l'emploi actuel (statut, type de contrat, conditions de travail), l'historique de carrière professionnelle (niveau d'études, emplois occupés, éventuelles interruptions d'activité

professionnelle) et les expositions à certaines nuisances (contraintes organisationnelles, pénibilité, bruit professionnel, nuisances d'origine chimique, physique ou biologique).

D'autres données dites « auxiliaires » issues des différents systèmes d'informations existants seront recueillies afin de documenter, d'une part, les événements de santé et, d'autre part, les événements socio-professionnels :

- (II [3°] de l'article 2) à partir du SNIIRAM : des données relatives aux consommations de soins des personnes, les données hospitalières mentionnées à l'article R. 6113-1 du code de la santé publique et des données relatives à la prise en charge pour affection de longue durée, pour maladie professionnelle, accident du travail ou invalidité ;
- (II [4°] de l'article 2) à partir des fichiers de la CNAV : droits de retraite perçus et parcours socio-professionnel (notamment périodes d'activité, de chômage, rémunération) ;
- (II [5°] de l'article 2) auprès du régime d'assurance maladie dont relève la personne concernée : données sociodémographiques, données relatives à la santé (notamment la prise en charge des affections longue durée, des accidents du travail et maladies professionnelles et des indemnités journalières), données relatives à l'activité professionnelle (notamment aux périodes d'activité, à la nature et aux conditions de cette activité ainsi que, le cas échéant, à la perception d'une retraite agricole) ;
- (II [6°] de l'article 2) les causes médicales de décès auprès du Centre d'épidémiologie sur les causes médicales de décès (CépiDc).

Afin de permettre le suivi de chaque participant au programme COSET, un numéro de « cohortiste » (« NC-DST ») lui est attribué. La commission prend acte de ce que le terme « *identifiant anonymisé* » a été remplacé par le terme « *code de confidentialité* » pour désigner ce numéro de « cohortiste » dans le projet.

En application de l'article 6 [3°] de la loi Informatique et Libertés, la commission considère que les catégories de données à caractère personnel enregistrées sont adéquates et pertinentes au regard des finalités de l'étude telles que prévues à l'article 1<sup>er</sup> du projet.

La commission relève, cependant, que le NIR des participants à l'étude ne figure pas parmi les données traitées à l'article 2 du projet alors que le NIR est utilisé à des fins d'appariement avec différentes bases de données existantes. Dès lors, elle demande que la liste des catégories de données à caractère personnel enregistrées soit complétée afin d'y mentionner le NIR.

#### **Sur les modalités de collecte des données :**

Les données précitées seront collectées selon différentes modalités. Elles seront issues :

- des questionnaires renseignés par les personnes ;
- des régimes d'assurance maladie dont relèvent les participants à l'étude (RSA ou MS1) ;
- du Système national d'information interrégimes de l'assurance maladie (SNIIRAM) géré par la Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS) ;
- du Système national de gestion des carrières (SNGC) géré par la Caisse nationale d'assurance vieillesse (CNAV) ;
- du Centre d'épidémiologie sur les causes médicales de décès (CépiDc), selon les conditions fixées par le décret n° 98-37 du 16 janvier 1998 autorisant l'accès aux données relatives au décès des personnes inscrites au Répertoire national d'identification des personnes physiques dans le cadre des recherches dans le domaine de la santé.

Concernant la collecte de données issues du SNIIRAM, l'article 4 du projet autorise la CNAMTS à utiliser le NIR des participants à l'étude « *en vue d'effectuer un appariement des données relatives aux personnes participant au programme (...) avec les données issues du SNIIR-AM* ».

Dès lors que l'identifiant utilisé dans le SNIIRAM, permettant d'obtenir les consommations de soins individuelles, est issu du résultat d'une double fonction de hachage du NIR, la CNAMTS, grâce au NIR des personnes, pourra extraire les données pertinentes et les transmettre à l'InVS associées à un code de confidentialité. L'InVS pourra ainsi relier les consommations de soins précitées aux données collectées dans le cadre de la cohorte.

Le NIR des personnes est également l'identifiant utilisé au sein des fichiers de la CNAV, de la Caisse centrale de la MSA et de la Caisse nationale du RSI. Afin de permettre l'appariement des données de la cohorte avec celles issues de ces différentes bases, les articles 5 (CNAV) et 6 (régimes d'assurance maladie) du projet autorisent le traitement du NIR à cette fin. En particulier, l'article 5 du projet ainsi que le dossier produit à l'appui de la demande prévoient que les NIR ne sont échangés qu'entre la CNAV, la MSA et le RSI qui sont habilités à le détenir.

La commission en prend acte.

En outre, elle relève qu'un « numéro de confidentialité » différent est attribué aux personnes par chaque régime d'assurance maladie, la CNAMTS et la CNAV lors de la phase d'inclusion et qu'un prestataire est chargé de conserver la correspondance entre ces différents numéros identifiants et l'identité des personnes concernées.

Dans la mesure où cette disposition permet d'empêcher, d'un côté, l'InVS de connaître la correspondance entre l'identité d'une personne et les données de l'étude associées, et, d'un autre côté, les régimes d'assurance maladie, la CNAMTS et la CNAV de connaître le code de confidentialité NC-DST, la commission estime que cette organisation est de nature à garantir la confidentialité des données.

Le recueil des données dites « auxiliaires » sera réalisé tant pour les personnes ayant renseigné le questionnaire d'inclusion que pour un sous-échantillon de non-répondants au questionnaire, constitué par tirage au sort parmi les individus n'ayant pas répondu mais ne s'étant pas opposés au traitement de leurs données pour l'étude.

La commission prend acte de ce que les données auxiliaires ne seront pas recueillies ni s'agissant des personnes ayant exprimé leur opposition à la collecte de ces informations, ni pour celles considérées comme non informées (plis non remis au destinataire).

#### **Sur la durée de conservation des données :**

L'article 10 du projet indique que les données seront « *collectées et conservées pendant vingt ans* ». Il est précisé que « *l'utilisation du numéro d'inscription au répertoire national des personnes physiques n'est autorisée que pendant une durée qui n'excède pas la durée de collecte de données* ».

Compte tenu de l'investissement nécessaire à la constitution et au suivi régulier de la cohorte ainsi que des enjeux de cette étude en termes de santé publique, la commission estime que ces durées de conservation des données n'excèdent pas les durées nécessaires aux finalités pour lesquelles elles sont collectées et traitées, conformément à l'article 6 (5°) de la loi Informatique et Libertés.

#### **Sur les destinataires des données :**

Conformément à l'article 29 (4°) de la loi Informatique et Libertés, l'acte autorisant le traitement doit préciser les destinataires ou catégories de destinataires habilités à recevoir communication des données.

Ainsi, l'article 8 du projet précise que : « *Seuls les personnels de l'Institut de veille sanitaire, nommément désignés et habilités à cet effet par son directeur, sont autorisés à accéder aux données mentionnées à l'article 2 du présent décret, dans la stricte mesure où elles sont nécessaires à l'exercice des missions qui leur sont confiées.* »

La commission relève que des données directement identifiantes sont traitées dans le cadre de l'étude COSET (II [1°], a) de l'article 2 du projet) et qu'un prestataire est chargé d'assurer leur sécurité et leur confidentialité (article 3 [1°] du projet).

Dès lors, la commission demande que l'article 8 du projet soit modifié afin de préciser que seul le prestataire mentionné à l'article 3 (1°) du projet devrait accéder aux données directement identifiantes ou, qu'à défaut, la politique d'habilitation mise en œuvre au sein de l'InVS garantisse que les personnels habilités à accéder aux données directement identifiantes n'ont pas accès aux données de l'étude.

#### **Sur l'information des personnes :**

L'article 2-III du projet prévoit que « *Les personnes sont individuellement informées par courrier de la possibilité d'un recueil auprès de leur organisme de protection sociale des données relatives à leur état de santé et à leur situation professionnelle. Conformément à l'article 57, elles sont spécifiquement informées des modalités d'opposition à la collecte de ces données.* »

En outre, le protocole de recherche prévoit plusieurs modalités d'information des personnes, tant collectives qu'individuelles :

##### **L'information collective :**

Une campagne de communication relative au programme COSET a été lancée par l'InVS et les régimes d'assurance maladie concernés. Un site internet dédié à l'étude informe le public des caractéristiques de l'étude COSET et met à disposition les documents de l'étude.

##### **L'information individuelle :**

Les personnes tirées au sort sont informées par courrier du lancement du programme COSET. Cette note comprend des informations relatives à l'identité du responsable du traitement, aux objectifs, aux modalités de mise en œuvre de l'étude et aux conditions d'exercice des droits des personnes, conformément aux dispositions de l'article 57 de la loi du 6 janvier 1978 modifiée.

En particulier, les personnes y sont informées du caractère volontaire et facultatif de leur participation à l'étude et de l'absence de conséquences d'un refus d'y participer. Elles y sont également informées de la possibilité de mettre fin à leur participation à tout moment.

Un suivi des plis non distribués mis en place lors de l'invitation initiale permet de repérer les personnes qui n'ont pas reçu de courrier d'information. Les données auxiliaires ne seront pas recherchées pour ces personnes qui n'ont pu être informées.

La commission en prend acte.

#### **Sur les droits d'accès, de rectification et d'opposition des personnes :**

Outre l'information spécifique relative à l'étude, le dossier produit à l'appui de la demande d'avis précise que les personnes concernées par l'étude sont informées des modalités pratiques d'exercice de leurs droits. Ceux-ci pourront être exercés par courrier postal, courrier électronique ou téléphone en appelant le numéro vert dédié à l'étude.

La commission prend acte de ces mesures qui sont de nature à garantir l'exercice effectif des droits des personnes.

Cependant, l'article 11 du projet prévoit que : « *Les droits d'accès et de rectification des données, prévus aux articles 39 et 40 de la loi du 6 janvier 1978 susvisée, sont exercés auprès du département scientifique de l'Institut de veille sanitaire, en charge du programme COSET.* »

Ce faisant, la commission relève que les personnels de l'InVS auraient, dès lors, accès aux données directement identifiantes des participants à l'étude dans l'hypothèse où ceux-ci exerceraient leurs droits.

Afin de garantir la confidentialité des données, au regard, notamment de ce que prévoit l'article 8 du projet, la commission estime que les droits des personnes devraient être exercés auprès du prestataire en charge de la gestion des données directement identifiantes, mentionné à l'article 3 (1<sup>o</sup>) du projet.

Ce prestataire, qui serait autorisé en l'espèce à disposer de données directement identifiantes pour accomplir sa mission de conservation de la correspondance entre l'identité et les numéros de confidentialité des personnes concernées pourrait ainsi recevoir les demandes émanant de ces personnes dans le cadre de l'exercice de leurs droits, et transmettre ces demandes à l'InVS après les avoir expurgées des éléments directement identifiants et leur avoir associées le numéro de confidentialité permettant à l'InVS de faire droit aux demandes sans avoir accès à l'identité.

A défaut, cette séparation stricte entre données identifiantes et données de l'étude doit être assurée au sein de l'InVS notamment par la mise en place d'une politique d'habilitation stricte afin que les mêmes personnes n'accèdent pas aux données de l'étude et aux données directement identifiantes.

Dès lors, la commission demande que l'article 11 du projet soit modifié sur ce point.

#### **Sur la sécurité des données et la traçabilité des actions :**

L'article 9 du projet prévoit que l'InVS est responsable de la mise en œuvre des mesures de sécurité des données enregistrées dans le traitement.

L'article 8 du projet prévoit que seules les personnes faisant partie de l'InVS, habilitées par le directeur de l'InVS, peuvent accéder aux données mentionnées à l'article 2.

L'authentification passe par l'utilisation d'un mot de passe d'une longueur minimale de huit caractères pour les participants. Pour les utilisateurs internes à l'InVS, gestionnaires et administrateurs, une longueur minimale de dix caractères est imposée, et le mot de passe doit contenir au minimum une majuscule, une minuscule et un chiffre. Les mots de passe choisis par les participants lors de l'activation de leur compte ne sont pas stockés en clair dans l'application.

La commission rappelle qu'elle recommande, concernant la complexité des mots de passe, d'utiliser des mots de passe disposant d'une longueur minimale de huit caractères pour les utilisateurs ne disposant pas de privilèges élevés sur les données, et de dix caractères pour les autres. En outre, les mots de passe doivent comporter trois types de caractères distincts parmi les minuscules, majuscules, chiffres ou caractères spéciaux. Ils ne doivent pas être stockés en clair dans un fichier ou une base de données, y compris pour les utilisateurs internes à l'InVS.

Les accès à l'application sont tracés. Les traces des participants sont utilisées pour évaluer et améliorer l'outil de recueil de données et, éventuellement, pour être utilisées lors de l'analyse afin de prendre en compte la non-participation.

La commission demande que les données de journalisation soient conservées pendant une durée de six mois glissants.

Les transferts de données sont sécurisés par la transmission par courriel de pièces jointes préalablement chiffrées.

La commission précise que le dossier présenté n'apporte pas d'éléments relatifs aux mesures qui seront mises en œuvre pour assurer la confidentialité des données renseignées par les participants en remplissant les auto-questionnaires directement sur internet. Elle rappelle à cet égard ces échanges doivent être sécurisés et que le protocole SSL dans sa version 3 n'est plus fiable. Il conviendra donc d'utiliser le protocole TLS dans sa dernière version.

Le projet de décret prévoit que certaines données sont conservées chiffrées. Il s'agit :

- des adresses électroniques, conservées par l'InVS ;
- des NIR, nom patronymique, nom d'usage, prénoms, sexe, date et lieu de naissance des personnes tirées au sort, conservées par les régimes d'assurance maladie.

Ces mesures sont satisfaisantes. Toutefois, la commission rappelle qu'une procédure doit être documentée et mise en œuvre afin de s'assurer que seules les personnes habilitées pourront avoir connaissance des clés de chiffrement, notamment en cas de départ d'une personne qui disposait des habilitations d'accès à ces clés, et qu'il convient de vérifier régulièrement les vulnérabilités des algorithmes de chiffrement utilisés.

Enfin, chacun des partenaires s'engage à supprimer définitivement les données à l'aide d'un logiciel disposant d'une certification de premier niveau de l'Agence nationale de sécurité des systèmes d'Information si une procédure n'existe pas déjà dans ces organismes, ou au moyen de la procédure déjà en place.

Sous réserve des observations précédentes, les mesures de sécurité décrites par le responsable de traitement sont conformes à l'exigence de sécurité prévue par l'article 34 de la loi du 6 janvier 1978 modifiée. La commission rappelle toutefois que cette obligation nécessite la mise à jour des mesures de sécurité au regard de la réévaluation régulière des risques.

Les autres points du projet de décret n'appellent pas, en l'état et au regard de la loi du 6 janvier 1978 modifiée, d'autres observations de la part de la commission.

*La présidente,*  
I. FALQUE-PIERROTIN