

Comité sectoriel de la sécurité sociale et de la santé  
Section “Santé”

CSSSS/15/211

**DÉLIBÉRATION N° 15/079 DU 17 NOVEMBRE 2015 PORTANT SUR LA COMMUNICATION DE DONNÉES À CARACTÈRE PERSONNEL CODÉES RELATIVES À LA SANTÉ PAR TROIS HÔPITAUX À CERTAINS CENTRES DE RECHERCHE UNIVERSITAIRES DANS LE CADRE DE L'ÉTUDE SCIENTIFIQUE INTITULÉE « INTEROPERABILITY OF MEDICAL DATA THROUGH INFORMATION EXTRACTION AND TERM ENCODING » (IMEDIATE)**

La section santé du Comité sectoriel de la sécurité sociale et de la santé (dénommée ci-après « le Comité sectoriel »);

Vu la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, en particulier l'article 37;

Vu la loi du 8 décembre 1992 *relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel*;

Vu la loi du 21 août 2008 *relative à l'institution et à l'organisation de la plate-forme eHealth*;

Vu la demande d'autorisation du Centre de traitement automatique du langage de l'UCL;

Vu le rapport d'auditorat de la Plate-forme eHealth du 4 novembre 2015;

Vu le rapport de monsieur Yves Roger.

Émet, après délibération, la décision suivante, le 17 novembre 2015:

## I. OBJET DE LA DEMANDE

1. Le Centre de traitement automatique du langage (CENTAL) de l'Université catholique de Louvain planifie conjointement avec le 'Pole of Research on Information and Services Management and Engineering' de l'UCL, le Centre de recherche en linguistique appliquée de l'Institut Supérieur de traducteurs et d'interprètes de la Haute école de Bruxelles (TERMISTI) et l'Ecole de Santé Publique (ESP) de l'Université libre de Bruxelles une étude scientifique sur la possibilité d'exploiter de manière conjointe les données structurées et non structurées des systèmes d'information médicale. La recherche porte le nom d' « Interoperability of Medical Data through Information extraction and Term Encoding » (iMediate).
2. Par « données structurées », on entend les rapports encodés sous forme de listes de codes liés à des terminologies ou ontologies médicales (SNOMED, ICD9/10, etc.). Par « données non structurées », on entend les textes cliniques (rapport d'hospitalisation, lettre de sorties, courriers, rapports d'imagerie, etc.) rédigés sous forme de texte libre.
3. Le défi de la présente recherche est de repérer automatiquement la terminologie présente dans les données non structurées. Les deux grands objectifs sont :
  - linguistique et terminologique: étude de la variété et de la variation des termes dans les corpus des différents hôpitaux;
  - informationnel: développer des applications de recherche d'information ou de catégorisation automatique.
4. Afin de pouvoir réaliser l'étude, les données à caractère personnel codées relatives à la santé seront communiquées aux centres de recherche par trois hôpitaux (CHIREC, Cliniques universitaires Saint-Luc et ERASME), d'une part, sous la forme de données cliniques non structurées et, d'autre part, sous la forme de données structurées. Sont visées dans la phase actuelle de l'étude, par hôpital, les données à caractère personnel contenues dans les dossiers disponibles relevant des domaines de la gastro-entérologie et de la médecine interne. Un corpus plus transversal sera constitué dans une phase ultérieure (hors du champ d'application de la présente demande).
5. Etant donné qu'il n'est pas nécessaire de disposer de l'identité des patients concernés pour l'étude, un tiers de confiance (TTP) supprimera les données d'identification dans les dossiers des patients avant la communication des données de santé aux chercheurs. Ce processus se déroulera comme suit:
  - Tout hôpital attribue un numéro aléatoire unique à chaque patient.
  - Le tiers de confiance se rend dans chaque hôpital et supprime dans les fichiers de données sélectionnés les données à caractère personnel susceptibles de donner lieu à l'identification directe (nom, prénom, adresse, numéro de téléphone, numéros d'identification, etc.). L'identification de l'hôpital et des prestataires de soins est également supprimée. Seule la date de naissance de l'intéressé est conservée de manière codée: la date de naissance est remplacée par un nombre entier déterminant un nombre de jours depuis une date de référence secrète (uniquement connue par le tiers de confiance). Le tiers de confiance procède au codage du numéro aléatoire attribué par

l'hôpital de sorte qu'il soit le seul à connaître le lien entre le numéro codé et le numéro attribué par l'hôpital. Le tiers de confiance qui dispose d'un médecin responsable, doit également réaliser une analyse « small cell » et doit, si nécessaire, effectuer les traitements nécessaires (agrégations, etc.) dans les fichiers de données afin d'exclure toute identification indirecte des personnes concernées.

- Tout hôpital contrôle la qualité de la dé-identification réalisée par le tiers de confiance et approuve l'utilisation des fichiers de données en question.
- Le tiers de confiance transmet les données à caractère personnel codées au coordinateur du projet de recherche (CENTAL).
- Le coordinateur du projet de recherche transmet enfin, de manière sécurisée (chiffrée), les données à caractère personnel codées aux chercheurs des trois centres de recherche concernés par la réalisation de l'étude.

6. La S.A. EarlyTracks interviendra comme tiers de confiance. EarlyTracks est experte en *text mining* et *information retrieval*, un domaine de l'ingénierie linguistique dont l'objectif est d'extraire, de structurer et de traiter des informations enregistrées sous forme de textes. Comme requis, le demandeur et le tiers de confiance déclarent qu'ils satisfont aux conditions fixées par le Comité sectoriel et qui incombent à tout tiers de confiance chargé de coder des données relatives à la santé, telles que décrites dans la délibération n° 14/059 du 15 juillet 2014<sup>1</sup>, plus précisément comme suit:

- Le TTP doit être suffisamment indépendant vis-à-vis de l'émetteur (des émetteurs) des données à caractère personnel à coder et du (des) destinataire(s) des données à caractère personnel codées.

Bien que EarlyTracks soit une spin-off de l'Université catholique de Louvain, il s'agit d'une entité indépendante qui est autonome par rapport au responsable du traitement, CENTAL. EarlyTracks dispose d'un centre d'exploitation propre. EarlyTracks et CENTAL disposent d'effectifs de personnel totalement distincts ainsi que de deux parcs informatiques indépendants. Aucun agent de CENTAL n'a accès au parc informatique de EarlyTracks et vice versa.

- Le TTP doit effectuer le codage au moyen de techniques qui rendent raisonnablement impossible la conversion des données à caractère personnel codées en données à caractère personnel non codées. Ce codage doit en principe concerner toutes les personnes identifiées ou identifiables, c'est-à-dire tant les personnes auxquelles les données à caractère personnel relatives à la santé s'appliquent (ex. des patients) que les personnes qui fourniraient les données (ex. des prestataires de soins).

EarlyTracks supprime toutes les données à caractère personnel susceptibles de donner lieu à une identification directe. EarlyTracks remplacera le numéro d'identification aléatoire attribué par les hôpitaux par un numéro codé au moyen d'une table de conversion. Le codage a trait aux patients, aux hôpitaux et aux prestataires de soins qui sont intervenus. La transmission des données entre les hôpitaux et EarlyTracks et entre

---

<sup>1</sup> Délibération n° 14/059 du 15 juillet 2014 relative à la communication de données à caractère personnel codées relatives à la santé dans le cadre du projet Thales.

EarlyTracks et CENTAL se fait au moyen d'une clé USB (sans l'utilisation d'un réseau informatique).

- Le TTP doit, en toute connaissance de cause, veiller à ce que l'ensemble de données à caractère personnel codées mis à la disposition ne permette raisonnablement pas une réidentification (il s'agit du « small cell risk »). Cela signifie que le TTP doit, le cas échéant, disposer de compétences spécifiques afin de pouvoir juger de la possibilité de réidentification au moyen de données à caractère personnel codées relatives à la santé.

EarlyTracks garantit l'exécution d'une étude « small cell risks » sur la base du texte des documents cliniques. Cette analyse comportera deux phases. La première sera automatique et visera l'identification des dossiers faisant état de maladies rares (seuil communément admis de 1/2000 en Europe, d'après l'INSERM) et reposera sur la liste d'Orphanet<sup>2</sup>. La deuxième sera manuelle et basée sur des échantillons aléatoires constitués pour chaque type de document (lettre de sortie, compte-rendu d'opération, etc.). Les textes seront analysés sous la supervision du médecin responsable travaillant pour EarlyTracks. Cette tâche aura pour but l'identification de tout risque potentiel de type « small cell ». Tout problème identifié sur l'échantillon fera systématiquement l'objet d'un screening sur l'ensemble des données. Les données posant des problèmes de type « small cell » et représentant donc un risque possible de réidentification seront supprimées du corpus. Dès qu'un document est jugé problématique, tous les documents liés au même patient sont supprimés. Seules les données validées au terme de l'analyse « small cell » seront transmises par EarlyTracks (TTP) au CENTAL pour utilisation dans le cadre du projet iMediate.

- Le traitement de données à caractère personnel relatives à la santé par un TTP doit être effectué sous la surveillance et la responsabilité d'un professionnel des soins de santé, de préférence un médecin.

Le Comité sectoriel a reçu l'identité du médecin concerné.

- Le TTP doit désigner, en interne ou en externe, un conseiller en sécurité de l'information et en protection de la vie privée.

Le Comité sectoriel a reçu l'identité du conseiller en sécurité de l'information. La désignation de l'intéressé en qualité de conseiller en sécurité de l'information a, dans le passé, déjà fait l'objet d'un avis positif du Comité sectoriel.

- Le TTP doit veiller au respect correct de la législation en matière de protection de la vie privée et doit entreprendre toutes les actions nécessaires afin d'en assurer le respect. Le TTP ne peut pas utiliser les données qu'il a traitées dans le cadre de sa fonction de TTP pour des finalités autres que les finalités spécifiques qui lui ont été confiées. Le TTP doit prendre les mesures techniques et organisationnelles appropriées qui sont nécessaires à la protection des données à caractère personnel contre la destruction accidentelle ou non autorisée, contre la perte accidentelle ainsi que contre la

---

<sup>2</sup> [http://www.orpha.net/consor/cgi-bin/Education\\_AboutRareDiseases.php?lng=FR](http://www.orpha.net/consor/cgi-bin/Education_AboutRareDiseases.php?lng=FR)

modification, l'accès et tout autre traitement non autorisé de données à caractère personnel. Le TTP doit détruire les données que les responsables du traitement initial lui ont fournies, dès qu'il a réalisé sa mission de codage. Le TTP doit effectuer les traitements en toute transparence.

EarlyTracks et CENTAL ont, comme requis, déclaré sur l'honneur qu'ils satisfont à ces conditions.

## II. COMPÉTENCE

7. En vertu de l'article 42, § 2, 3<sup>o</sup>, de la loi du 13 décembre 2006 *portant dispositions diverses en matière de santé*, la section Santé du Comité sectoriel de la sécurité sociale et de la santé est en principe compétente pour l'octroi d'une autorisation de principe concernant toute communication de données à caractère personnel relatives à la santé.
8. Le Comité sectoriel s'estime dès lors compétent pour se prononcer sur la présente demande d'autorisation.

## III. TRAITEMENT

### A. ADMISSIBILITÉ

9. Le traitement de données à caractère personnel relatives à la santé est en principe interdit, et ce conformément au prescrit de l'article 7, § 1<sup>er</sup>, de la loi du 8 décembre 1992 *relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel* (dénommée ci-après la loi relative à la vie privée).

L'interdiction ne s'applique cependant pas, notamment lorsque le traitement est nécessaire à la recherche scientifique et est effectué conformément à l'arrêté royal du 13 février 2001 *portant exécution de la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel*<sup>3</sup>. Il en va de même lorsque le traitement est nécessaire à la promotion et à la protection de la santé publique y compris le dépistage<sup>4</sup>.

10. A la lumière de ce qui précède, le Comité sectoriel est par conséquent d'avis qu'il existe un fondement pour le traitement des données à caractère personnel relatives à la santé envisagé.

### B. FINALITÉ

11. L'article 4, § 1<sup>er</sup>, de la loi relative à la vie privée autorise le traitement de données à caractère personnel uniquement pour des finalités déterminées, explicites et légitimes.

<sup>3</sup> Art. 7, § 2, k), de la loi relative à la vie privée.

<sup>4</sup> Art. 7, § 2, d), de la loi relative à la vie privée.

12. La présente étude est réalisée par le Centre de traitement automatique du langage (CENTAL) de l'Université catholique de Louvain en collaboration avec le « Pole of Research on Information and Services Management and Engineering (PRISME) » de l'UCL, le Centre de recherche en linguistique appliquée de l'Institut supérieur de traducteurs et d'interprètes de la Haute école de Bruxelles (TERMISTI) et de l'Ecole de Santé Publique (ESP) de l'Université libre de Bruxelles.
13. Les objectifs du traitement sont clairement définis, à savoir une étude sur la possibilité d'exploiter de manière conjointe les données structurées et non structurées des systèmes d'information médicale. L'étude a pour objectif d'étudier la variété et la variation des termes dans le corpus des différents hôpitaux en vue du développement d'applications de recherche d'informations ou de catégorisation automatique.
14. Le demandeur souligne, par ailleurs, que le projet correspond à l'action 13 du Plan d'action eSanté 2013-2018 « réalisation d'une politique de terminologie nationale »<sup>5</sup>.
15. Compte tenu des objectifs du traitement tels que décrits ci-dessus, le Comité sectoriel estime que le traitement des données à caractère personnel précitées poursuit effectivement des finalités déterminées, explicites et légitimes.

### **C. PROPORTIONNALITÉ**

16. L'article 4, § 1<sup>er</sup>, 3<sup>o</sup>, de la loi relative à la vie privée dispose que les données à caractère personnel doivent être adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont obtenues et pour lesquelles elles sont traitées ultérieurement.
17. Les données à caractère personnel recueillies pour la présente étude portent, d'une part, sur des données cliniques non structurées (rapports d'hospitalisations, lettres de sortie, correspondance, rapports d'imagerie, ...) et, d'autre part, sur des données structurées provenant de trois hôpitaux (CHIREC, Cliniques universitaires Saint-Luc et ERASME). Sont visés par hôpital tous les dossiers disponibles relevant des domaines de la gastro-entérologie et de la médecine interne.
18. Le demandeur déclare que les données non structurées sont nécessaires afin de pouvoir étudier la manière dont la terminologie utilisée peut être repérée automatiquement. Les chercheurs utiliseront à cet effet des ontologies médicales de base qu'ils compléteront par un travail d'enrichissement terminologique.
19. Le second objectif du projet est d'automatiser la structuration des documents médicaux et l'attribution de catégories d'encodage (termes descripteurs). Le but ultime étant de concevoir des méthodes capables de réconcilier les données structurées et non structurées. Il est donc capital pour le projet de disposer de ressources permettant d'entraîner les modèles de classification ou les algorithmes de recherche d'information.

---

<sup>5</sup> Le Plan d'action 2013-2018 a été actualisé en octobre 2015 (Roadmap 2.0). La documentation est disponible à l'adresse suivante: <http://www.plan-egezondheid.be/>

20. Compte tenu des finalités du traitement, le Comité sectoriel estime que la communication envisagée est pertinente et non excessive.
21. Conformément à l'article 4, § 1<sup>er</sup>, 5<sup>o</sup>, de la loi relative à la vie privée, les données à caractère personnel ne peuvent être conservées sous une forme permettant l'identification des personnes concernées que pendant une durée n'excédant pas celle nécessaire à la réalisation des finalités pour lesquelles elles sont obtenues.
22. Le demandeur déclare que les données codées seront conservées jusqu'à la fin de l'étude, plus précisément jusque fin 2017. Le Comité sectoriel est d'accord avec ce délai de conservation.

#### **D. TRANSPARANCE**

23. L'article 14 de l'arrêté royal du 13 février 2001 portant exécution de la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel prévoit que le responsable du traitement de données à caractère personnel, collectées à des fins déterminées, explicites et légitimes, ou l'organisation intermédiaire doit, préalablement au codage des données, communiquer certaines informations à la personne concernée.
24. Le responsable du traitement est toutefois dispensé de fournir ces informations lorsque l'information de la personne concernée se révèle impossible ou implique des efforts disproportionnés. Compte tenu du nombre important de personnes dont les données à caractère personnel seront traitées, le Comité sectoriel estime que la communication impliquerait des efforts disproportionnés.

#### **E. MESURES DE SÉCURITÉ**

25. En vertu de l'article 17 de la loi relative à la vie privée, le responsable du traitement doit, avant de mettre en œuvre un traitement entièrement ou partiellement automatisé, faire une déclaration à la Commission de la protection de la vie privée. Le demandeur est dès lors tenu de faire le nécessaire.
26. Conformément à l'article 7, § 4, de la loi relative à la vie privée, le traitement de données à caractère personnel relatives à la santé peut uniquement être effectué sous la surveillance et la responsabilité d'un professionnel des soins de santé. Même si cela n'est pas strictement requis par la loi relative à la vie privée, le Comité sectoriel estime qu'il est préférable de traiter de telles données sous la responsabilité d'un médecin<sup>6</sup>. Ce qui est le cas en l'espèce. Le Comité sectoriel rappelle que lors du traitement de données à caractère personnel, le professionnel des soins de santé ainsi que ses préposés ou mandataires sont soumis au secret.

---

<sup>6</sup> Voir la délibération du Comité sectoriel de la sécurité sociale et de la santé n°07/034 du 4 septembre 2007 relative à la communication de données à caractère personnel au Centre fédéral d'expertise des soins de santé en vue de l'étude 2007-16-HSR « étude des mécanismes de financement possibles pour l'hôpital de jour gériatrique ».

27. Conformément à l'article 16, § 4, de la loi relative à la vie privée, l'Université libre de Bruxelles doit prendre toutes les mesures techniques et organisationnelles requises pour protéger les données à caractère personnel. Ces mesures doivent garantir un niveau de protection adéquat, compte tenu, d'une part, de l'état de la technique en la matière et des frais qu'entraîne l'application de ces mesures et, d'autre part, de la nature des données à protéger et des risques potentiels.
28. Pour garantir la confidentialité et la sécurité du traitement de données, tout organisme qui conserve, traite ou communique des données à caractère personnel est tenu de prendre des mesures dans les onze domaines d'action suivants liés à la sécurité de l'information: politique de sécurité; désignation d'un conseiller en sécurité de l'information; organisation et aspects humains de la sécurité (engagement de confidentialité du personnel, information et formations régulières du personnel sur le thème de la protection de la vie privée et sur les règles de sécurité); sécurité physique et de l'environnement; sécurisation des réseaux; sécurisation logique des accès et des réseaux; journalisation, traçage et analyse des accès; surveillance, revue et maintenance; système de gestion des incidents de sécurité et de la continuité (systèmes de tolérance de panne, de sauvegarde, ...); documentation<sup>7</sup>.
29. Le demandeur déclare que les conditions suivantes sont remplies:
- Un responsable médical assume la responsabilité générale de la protection des données.
  - Un conseiller en sécurité de l'information a été désigné au niveau de l'institution.
  - Les risques encourus par les données à caractère personnel traitées ont été évalués et les besoins de protection y afférents ont été fixés.
  - Un document écrit (police de sécurité) définissant les stratégies et les mesures retenues pour la protection des données a été établi. Le Comité sectoriel a reçu un exemplaire du plan de sécurité pour le projet de recherche iMediate.
  - Tous les supports possibles sur lesquels sont enregistrées les données à caractère personnel traitées, ont été identifiés.
  - Le personnel interne et externe concerné par le traitement de données à caractère personnel a été informé, eu égard aux données traitées, des obligations de confidentialité et de protection découlant à la fois des différentes dispositions légales et de la politique de sécurité.
  - Les mesures de protection appropriées ont été prises afin d'empêcher tout accès non autorisé ou tout accès physique inutile aux supports contenant les données à caractère personnel traitées.
  - Les mesures nécessaires ont été prises afin de prévenir tout dommage physique qui risquerait de compromettre les données à caractère personnel traitées.
  - Les différents réseaux couplés au matériel traitant les données à caractère personnel sont protégés.
  - Une liste actuelle des différentes personnes compétentes qui ont accès aux données à caractère personnel dans le cadre du traitement, a été établie. Le Comité sectoriel a

---

<sup>7</sup> « Mesures de référence en matière de sécurité applicables à tout traitement de données à caractère personnel », document rédigé par la Commission de la protection de la vie privée.



reçu une liste limitative des collaborateurs qui ont accès aux données à caractère personnel codées.

- Un mécanisme d'autorisation d'accès a été conçu de sorte que les données à caractère personnel traitées et les traitements qui y ont trait, soient uniquement accessibles aux personnes et applications qui y sont expressément autorisées.
- Un système d'information assurant un traçage, un dépistage et une analyse permanents de l'accès de personnes et d'entités logiques aux données à caractère personnel traitées a été mis au point.
- Un contrôle de la validité et de l'efficacité dans le temps des mesures techniques ou organisationnelles instaurées a été prévu.
- Une documentation concluante a été établie concernant l'organisation de la protection du traitement envisagé. Cette documentation est tenue à jour.

**30.** Le Comité sectoriel rappelle qu'il est interdit, conformément à l'article 6 de l'arrêté royal du 13 février 2001 *portant exécution de la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel*, d'entreprendre toute action visant à convertir les données à caractère personnel codées qui ont été communiquées en données à caractère personnel non codées. Le non-respect de cette interdiction peut donner lieu à une amende en vertu de l'article 39, 1°, de la loi relative à la vie privée. Le Comité sectoriel rappelle également qu'en cas de condamnation du chef d'infraction à l'article 39, le juge peut prononcer la confiscation des supports matériels des données à caractère personnel formant l'objet de l'infraction (fichiers manuels, disques et bandes magnétiques, ...) ou ordonner l'effacement de ces données. Le juge peut également interdire de gérer, personnellement ou par personne interposée, et pour deux ans au maximum, tout traitement de données à caractère personnel<sup>8</sup>.

---

<sup>8</sup> Article 41 de la loi relative à la vie privée.

Par ces motifs,

**la section santé du Comité sectoriel de la sécurité sociale et de la santé,**

autorise, selon les modalités de la présente délibération, la communication de données à caractère personnel codées relatives à la santé par les Cliniques universitaires Saint-Luc, ERASME, CHIREC aux centres de recherche CENTAL, PRISME, TERMINISTI et ESP dans le cadre de l'étude scientifique iMediate.

Yves ROGER  
Président

Le siège du Comité sectoriel de la sécurité sociale et de la santé est établi dans les bureaux de la Banque Carrefour de la sécurité sociale, à l'adresse suivante: Quai de Willebroeck 38 - 1000 Bruxelles (tél. 32-2-741 83 11).