

Gouvernance de l'ouverture de données sensibles, des principes à l'action : aspects juridiques et éthiques dans le domaine de la recherche en santé

Gauthier Chassang, Bettina Couderc, Laurence Mabile, Anne Cambon-Thomsen
INSERM/UT3, UMR1027, Toulouse

L'ouverture totale des données de la recherche publique pose des questions pratiques, juridiques et éthiques particulières. Ici nous nous focalisons sur un élément essentiel à considérer dans la prise de décision quant à l'ouverture des données de la recherche : la sensibilité des informations concernées

COMMENT GERER LA SENSIBILITE DES DONNEES AVANT L'OUVERTURE ?

Désensibiliser ou ne pas ouvrir

Données personnelles

Information se rapportant à une personne physique identifiée ou identifiable.

Données dites « sensibles »: catégorie spéciale incluant les données de santé, données génétiques, biométriques, relatives à la vie sexuelle, à l'origine, aux convictions religieuses, aux croyances, aux convictions politiques, aux condamnations (principe d'interdiction de traitement sauf exception légale)

Principes: Confidentialité, respect de la vie privée, secret médical

Règles pour l'ouverture: consentement préalable de la personne concernée ou anonymisation totale et irréversible dans le respect des standards techniques acceptés (possibilité de valider le process avec la CNIL). Data protection officer (DPO) à consulter. Inscription au registre des traitements.

INDS – Institut national des données de santé: le gardien d'une ouverture maîtrisée des données de santé pseudonymisées (sensibles - accès restreint) et anonymisées (désensibilisées - accès ouvert), au service de l'intérêt public, sur évaluation des projets d'accès.

Textes de références: [RGPD \(UE\) 2016/679](#), [Loi Informatique et Libertés mod. 2018](#), [Loi République Numérique 2016-1321](#), [Art. L.311-6 CRPA*](#)

Données secrètes et classifiées

Ensemble de connaissance réservées à quelques-uns. Secrets des affaires; industriel et commercial (ex: R&D); statistique. Spécificité du secret-défense portant sur toute information (ou matériel) dont la divulgation non autorisée pourrait porter atteinte, à des degrés divers, aux intérêts de l'UE, ou à ceux d'un ou de plusieurs de ses États membres. Ex. Sécurité des Systèmes d'Information; sécurité publique, sûreté de l'Etat, atteinte à la politique extérieure...

Marquage initial des données de la recherche concernée. Mesures spéciales de protection (ex: cration de Zone à Régime Restrictif). Question du double usage. Inéligibilité de principe à l'ouverture.

Principes: Confidentialité et sécurité accrues, protection du potentiel scientifique et technique de la nation (PPST), 4 classes de sensibilités (UE), contrôle des transferts.

Règles pour l'ouverture: analyse de risques et procédure de déclassification des données. Masquage partiel éventuel. Délais légaux. L'ouverture n'est pas préconisée par défaut. L'intérêt public de l'ouverture doit être supérieur à celui de l'étanchéité de la protection de l'information (ex: Cas du SuperVirus H5N1).

Textes de références: [Art. L.311-5 CRPA](#); [Loi n°2018-670](#); [Décision \(UE, Euratom\) 2015/444 de la CE](#); [Décret n°98-608](#); [Instruction Interministérielle n°2102 et 1300](#);

Données soumises au droit d'auteur

Données de recherche et base de données, résultats ou publications et autres documents étant soumis au droit d'auteur.

Principes: protection des œuvres de l'esprit

Règles pour l'ouverture: autorisation de l'auteur/coauteurs ou titulaire des droits d'exploitation. Extraction/Retraitement. Disposition légale. Ouverture du preprint tolérée. Périodes d'embargos possibles pour les publications en format travaillé par l'éditeur (max: 6 mois dans le domaine des sciences, de la technique, de la médecine; 12 mois dans le domaine des sciences sociales, après première publication).

Textes de références: [Art.L.533-4-1 Code de la Recherche](#)

LA SENSIBILITE DES DONNEES A L'UTILISATION : UNE QUESTION IMPREVUE ?

Contrôler ou ne pas contrôler ?

Principe: liberté d'utilisation des données ouvertes

Mesures palliatives des risques de mésusages

- **Le contrôle des modalités d'accès:** consultation sur place, copie, mail, publication en ligne ou autorisation préalable (ex: INDS; European Open Science Cloud?). Quel rôle pour l'administration ? [Art.L.311-9 CRPA](#)

- **Les contrats et licences, un cadre adapté?** L'utilisateur des données ouvertes est contraint de respecter les termes du contrat ou de la licence associée au jeu de données ouvertes. Est-ce suffisant?

- **Les délais légaux d'ouverture des données non-communicables?** (25-100 ans); ou quand la sensibilité s'éteint... [Art. L.311-8 CRPA](#), [L.213-1 et 2 du Code du Patrimoine](#)

Deux exemples questionnant

1- Données génétiques partagées par les citoyens (ex. personal genome project), associées à leurs particularités (origine ethnique, prédispositions à des pathologies, caractéristiques psychologiques, type « colérique », physiques ou comportementales...), en accès illimité et non contrôlé => but : partage de connaissances.

Question: Risque de favoriser des recherches discutables du point de vue éthique? Ex: comparaisons de multiples séquences mettant en évidence des associations avec des « traits humains comportementaux » ex: empathie, paraphilie...

2- L'IA pour la réalisation d'un logiciel permettant le diagnostic, le pronostic de pathologies, la prescription thérapeutique ou d'exams complémentaires, peut être envisagée pour peu que l'on dispose d'un ensemble conséquent de dossiers médicaux.

Question: Risque lié au développement de nouveaux outils de diagnostic médical en ligne utilisant des données ouvertes (ex: sur internet, les réseaux sociaux...) par des opérateurs hors système de santé?

Problèmes:

- Fiabilité des données et donc des résultats: l'ouverture ne garantit pas la qualité et peut engendrer des erreurs
- Resensibilisation de données par croisements, ex: assimiler des comportements à des ethnies; réidentification des personnes
- Contournement des législations/institutions protectrices/déontologie (loi Informatique et Libertés, CNIL; loi Jardé, CPP; Code de la Santé)

Recommandations: contrôler l'accès aux données génétiques et médicales dans leur ensemble, même anonymisées ou ouvertes par l'évaluation des demandes d'accès selon la pertinence scientifique ou médicale et la probité éthique des projets

Questions partagées : les données génétiques et médicales sont des données sensibles en France. Est-ce un frein excessif au développement des innovations (ex: IA) pour la santé? Ou un mal nécessaire?

Une catégorisation malaisée des sensibilités, pas de taxonomie claire et manque de standards opérationnels sur le.s processus de désensibilisation et la gestion des pluri-sensibilités. Qui décide de l'ouverture, qui est responsable? Qui surveille la correcte utilisation? Un équilibre délicat entre ouverture et protection/précaution. Les organismes, notamment établissements publics, doivent se positionner et communiquer les bonnes pratiques aux agents !