



Construire notre patrimoine national des données de santé



Paris, le 05/11/2024

Le gf2i (Groupement français de l'industrie de l'information) est une organisation indépendante qui fédère depuis 1979 l'écosystème de la donnée autour de producteurs et/ou utilisateurs de données publics (France TV, DILA, IGN, ONERA, universités, etc.) et privés (Crédit Agricole, Deloitte, Total Énergies, éditeurs juridiques et scientifiques, Cairn.info, etc.). Leurs représentants y croisent leurs expertises pour faire émerger des solutions concrètes et innovantes pour relever les défis de la transition numérique, notamment pour une ouverture maîtrisée de leurs données.

Le gf2i salue la publication de la stratégie interministérielle pour construire notre patrimoine national des données de santé et son ouverture à consultation publique.

Le gf2i souligne que l'exploitation des données de santé constitue un facteur d'innovation en matière de recherche, de diagnostic, de traitement mais aussi de protection sociale et d'aménagement des territoires.

Dans ce cadre le gf2i soutient :

1 - Une ouverture des données raisonnée qui prend en compte les utilisateurs, les utilisations et la localisation des données (adoption de principes types pour les licences d'exploitation).

2 – Une approche fédérée de l'exploitation des données de santé, en cohérence avec l'approche européenne :

- elle diminue les risques en matière de cybersécurité (une attaque sur un entrepôt centralisateur unique constituerait une catastrophe) ;
- elle permet aux producteurs des données d'en conserver la maîtrise et d'en assurer l'actualisation continue et la maintenance ;
- elle permet de contourner les problèmes de compatibilité des formats et des codages des données ;
- elle évite une étape coûteuse et risquée de duplication sur des infrastructures externes.

Ce fonctionnement ne s'oppose pas à une approche plus centralisée par cas d'usage et à la constitution d'un catalogue centralisé de métadonnées recensant les bases de données, à l'exemple des données de recherche qui fonctionnent sur un principe de centralisation limitée et de fédération pour l'essentiel des données (<https://recherche.data.gouv.fr/fr>).

3 – L'anticipation et la pérennisation des financements nécessaires à cette politique (modèles économiques des entrepôts de données à définir).

4 – La constitution de jeux de données artificielles dont la création permet d'assurer la confidentialité des données, de réduire les coûts (inférieurs à ceux de la collecte de données réelles) et les délais d'intégration de ces données dans les essais cliniques et les études en santé au sens large. Grâce à l'instauration d'une garantie de contrôle humain sur ces données, il est possible d'anticiper les problèmes de qualité et de répétition des erreurs et de les traiter. Chaque fois que cela est possible, le recours systématique à des données artificielles de santé offrirait une garantie en matière de sécurité et de respect de la vie privée des patients.