

Outils techniques pour l'exploitation du SNDS

Périmètre des outils d'intérêt

La base principale du Système National des Données de Santé (SNDS) constitue une ressource exceptionnelle puisqu'elle couvre l'ensemble de la population recevant des soins sur le territoire français, et l'ensemble des prestataires de soins, avec une grande profondeur historique, et une possibilité d'interopérabilité avec de nombreuses autres bases de données. Cependant, son exploitation reste freinée par des obstacles techniques récurrents et indépendants de toute thématique médicale : complexité de sa structure relationnelle, multiplicité des tables et des nomenclatures, ou encore règles de gestion implicites (filtres de qualité, gestion des doublons, traitement des majorations tarifaires) que chaque équipe doit aujourd'hui redécouvrir et réimplémenter.

Des outils génériques, complémentaires aux travaux existants

Pour lever ces verrous qui génèrent aujourd'hui de fortes inefficiences pour l'ensemble de l'écosystème, le développement et l'intégration d'« outils SNDS » génériques constituent un levier majeur de mutualisation des efforts au bénéfice de tous. Contrairement aux nombreux travaux existants, centrés sur le ciblage de pathologies spécifiques, ces briques logicielles n'ont pas vocation à répondre à une question médicale donnée. Elles visent à faciliter et fiabiliser les étapes communes à la majorité des projets de recherche sur cette base : extraction, structuration, mise en qualité, requêtage, ainsi que préparation des données pour l'entraînement de modèles statistiques ou d'intelligence artificielle.

Ces outils ont vocation à être mutualisés entre équipes afin d'encapsuler les bonnes pratiques de la communauté. Ils permettront de réduire le risque d'erreurs méthodologiques (omissions de filtres, mauvaises jointures) qui compromettent la validité des résultats, et d'accélérer l'ensemble des usages de la base, quelle qu'en soit la thématique, en réduisant le temps de prise en main pour les nouveaux utilisateurs.

Maîtriser les subtilités du SNDS requiert aujourd'hui plusieurs années d'expérience, une expertise encore trop souvent figée dans des rapports textuels passifs. **Concevoir, posséder et généraliser ces outils est la condition pour rendre cette connaissance directement actionnable et exploiter pleinement le potentiel de nos données.**

Notre projet de cartographie cible exclusivement les **outils logiciels et briques algorithmiques génériques** dédiés à l'exploitation de la base principale du SNDS. Pour entrer dans le périmètre, ces outils doivent répondre aux critères suivants :

- **Une vocation transversale et non médicale** : contrairement aux travaux centrés sur le ciblage de pathologies spécifiques, ces outils sont indépendants de toute thématique de recherche. Ils visent à encapsuler des règles de gestion implicites et transversales (gestion des doublons, filtres de qualité, nomenclatures, majorations tarifaires).
- **Une couverture des étapes clés de la chaîne de traitement** : le périmètre englobe les solutions facilitant et fiabilisant les étapes communes à la majorité des projets (extraction, structuration, mise en qualité, requêtage), ainsi que la préparation des données pour les modèles statistiques ou d'intelligence artificielle.
- **Un potentiel de mutualisation et de standardisation** : les outils recherchés ont vocation à être partagés au sein de la communauté pour transformer une expertise textuelle passive en composants directement actionnables. L'objectif final est

d'identifier des fonctions courantes d'analyse qui, une fois standardisées en mode « clé en main », réduiront le risque d'erreur méthodologique et accéléreront la prise en main du SNDS pour tout type d'utilisateur.

La Plateforme des données de santé (PDS) lance une initiative nationale visant à fédérer la communauté des utilisateurs du SNDS. Cette démarche s'articule autour de deux objectifs concrets :

- Objectif 1 — Réaliser une cartographie systématique des outils existants

Un premier repérage a déjà été engagé, mais il reste partiel et repose sur une méthodologie fastidieuse. L'enjeu est désormais de le systématiser en sollicitant largement l'écosystème, pour identifier les fonctions courantes d'analyse de la base principale du SNDS pour lesquelles des briques algorithmiques standardisées présenteraient un intérêt collectif.

Cette démarche s'appuiera sur une collecte active des contributions de la communauté, via un formulaire diffusé par QR code, relayé à plusieurs occasions : meetups SNDS, Journée Commun du Numérique en santé, communications LinkedIn ou actualités sur le site de la PDS relayées par la communauté, mailings auprès du réseau open science (roadshow, relais existants - notamment dans les directions travaillant directement avec la communauté SNDS -, etc.).

- Objectif 2 — Soutenir la standardisation de ces fonctions courantes

Sur la base des fonctions couramment utilisées ainsi identifiées, une priorisation des besoins sera réalisée. Cette priorisation distinguera deux situations :

- Situation 1 : des outils existent déjà mais restent à un stade de maturité technologique insuffisant pour une réutilisation large. Dans cette situation, l'enjeu sera de soutenir leur consolidation et leur standardisation afin de les rendre interopérables et directement utilisables, en mode « clé en main », par tout utilisateur de la base principale du SNDS.
- Situation 2 : aucune solution satisfaisante n'existe encore, auquel cas l'enjeu sera d'accompagner le développement de nouveaux outils répondant à ce besoin partagé par la communauté.

Cet objectif sera d'abord mis en œuvre sur un nombre restreint de fonctions jugées prioritaires, sélectionnées à l'issue de la cartographie, avant d'envisager une extension à l'ensemble des besoins identifiés. Ces cas pilotes permettront de tester concrètement les modalités de consolidation ou de développement des outils, avant généralisation de la démarche.

La question du financement de ces travaux sera traitée dans un second temps, une fois la faisabilité de la standardisation clairement établie.