

Automatisation en R de la recherche et de la structuration de codes (CIM-10, CCAM, ATC) à partir de sources externes

Nelly Le Guen*, Sandrine Morin*

Résumé

La Haute Autorité de Santé (HAS) développe des indicateurs de qualité et de sécurité des soins (IQSS) à partir des bases médico-administratives (PMSI : Programme de médicalisation des Systèmes d'information - SNDS : Système National des Données de Santé) en lien avec les professionnels de santé, les patients et les usagers.

Les algorithmes d'identification des populations cibles et des événements d'intérêt mobilisent des variables codées selon des nomenclatures standardisées (Exemple : CIM-10). Ces codes sont spécifiés en amont par les chefs de projets scientifiques dans des fichiers Excel. Jusqu'ici ils étaient directement utilisés dans des scripts R, augmentant les risques d'erreurs et compliquant la mise à jour lors de l'ajout de nouveaux codes.

La problématique consiste à concevoir un dictionnaire de codes sous format Excel pour générer automatiquement, dans R, des objets par catégorie de codes.

Méthode :

- Formations et accompagnement des chefs de projet scientifiques par les data-managers et les statisticiens ;
- Structuration du dictionnaire de codes standardisé exploitable par R ;
- Développement de fonctions de lecture de plusieurs onglets EXCEL intégrées dans un objet unique (liste) ;
- Génération automatique, dans R, de vecteurs par catégories d'actes ou diagnostics.

Résultats :

Cette automatisation permet de créer un dictionnaire de codes unique et standardisé, générer automatiquement des vecteurs pour chaque événement d'intérêt et supprimer les listes de codes écrites en dur dans les scripts R.

Le processus a été testé avec succès et déployé.

Conclusion :

Cette approche centralise et sécurise l'utilisation des codes, facilite leur mise à jour et réduit significativement le risque d'erreur, tout en améliorant l'efficacité des développements. Elle est transposable à l'ensemble des indicateurs produits par la HAS avec un gain de temps et une réduction importante du risque d'erreur.

Perspective :

Déploiement généralisé de ce processus.

Mots-clefs (3 à 5) : Automatisation – Dictionnaire des codes – Sécurisation – Optimisation des traitements

* Haute Autorité de Santé, Service Evaluation et Outils pour la qualité et sécurité des soins - contact.iqss@has-sante.fr