cchsflow: Une approche ouverte à science pour transformer et combiner les enquêtes sur la santé de la population dans un ensemble de données

## Résumé

**Auteurs:** Warsame Yusuf, Méthodologiste, Institut de recherche de l’Hôpital d’Ottawa. Douglas G. Manuel, Scientifique supérieur et médecin, Institut de recherche de l’Hôpital d’Ottawa. Rostyslav Vyuha, Assistant de récherche, Institut de recherche de l’Hôpital d’Ottawa Carol Bennett, Épidémiologiste, Institut de recherche de l’Hôpital d’Ottawa.

**État Actuel:** L’Enquête sur la santé dans les collectivités canadiennes (ESCC) est l’une des plus grandes enquêtes transversales sur la santé de la population avec plus de 130 000 sondés tous les deux ans et plus de 1.1 million de sondés depuis son début en 2001. Tant que l’enquête reste relativement cohérent, il y a des différences entre des cycles qui pose une challenge majeure pour analyser l’enquête au fil du temps.

**Intervention:** Nous avons développé un processus du nom de *cchsflow* pour transformer et harmoniser les variables de l’ESCC dans des formats cohérents aux travers plusieurs cycles d’enquêtes.

**Résultats:** Le paquet *cchsflow* R développé utilisait les données d’enquête de l’ESCC entre 2001 et 2014. Les feuilles de travail étaient crées pour identifier des variables, leurs noms dans des cycles précédents, leurs structures de catégories et leurs noms des variables finales. Ces feuilles de travail ont ensuite été utilisées pour ré-coder les variables dans chaque cycle de l'ESCC pour générer les ensembles de données harmonisés qui peuvent être combiner dans un ensemble de données constamment étiqueté pour analyser. Le paquet a ensuite été ajouté comme un entrepôt de GitHub pour encourager la collaboration entre les autres chercheurs.

**Implication:** Le paquet *cchsflow* a été ajoute au Comprehensive R Archive Network (CRAN) et contient appui pour plus de 160 variables de l’ESCC, générer un ensemble de données de plus d’un million de sondés. En exécutant les pratiques de sciences ouvertes, *cchsflow* vise de minimiser le montant de temps nécessaires pour nettoyer et préparer les données pour les plusieurs CCHS participants à travers le Canada.

**Mots clés:** Enquêtes de santé, Analyse des données, Science des données, Santé de la population