
Atelier OpenRefine

Stefan Gaget*¹

¹Metabolic functional (epi)genomics and molecular mechanisms involved in type 2 diabetes and related diseases - UMR 8199 - UMR 1283 – CNRS – France

Résumé

Vous disposez de données sous différents formats et standards ? Vous constatez des erreurs de saisie, des écritures différentes, voire des données manquantes ? Le travail de reprise de données peut être fastidieux. Avez-vous pensé à utiliser OpenRefine ?

OpenRefine permet d'effectuer des opérations de nettoyage en masse, mais aussi d'explorer des jeux de données, de convertir les données d'un format à un autre, d'appliquer des transformations cellulaires de base et avancées, de filtrer et de créer un partitionnement des données avec des expressions régulières, d'extraire des entités nommées sur des champs de texte intégral.

Dans cet atelier, vous découvrirez par la pratique la richesse de ce logiciel et ses fonctions de base pour effectuer une curation de votre jeu de données. Vous créerez un projet, effectuerez des tris, comprendrez les facettes, les filtres, le clustering, la possibilité de reproduire vos actions et bien d'autres possibilités vous donnant envie d'aller plus loin avec cet outil.

*Intervenant