

Développement sous R de la base de données DEEP (Données Economiques des Elevages de Porcs)

Bérengère Lecuyer*

Sandrine Lunven†

Résumé

Le pôle économie de l'IFIP (Institut du Porc), avec l'aide de TAC ECONOMICS, a procédé à une refonte de son système d'information en 2020, visant notamment à automatiser l'enregistrement de données économiques et sectorielles. Une quarantaine de connecteurs ont été développés permettant d'extraire automatiquement des données provenant de sources différentes, nationales (INSEE, France AgriMer, RNM etc.) et internationales (Europe, Brésil, Chine, etc.) et de les enregistrer dans la base DEEP. Cette présentation a pour objectif de présenter en détails les aspects techniques de ce projet, développés avec le logiciel R.

Mots-clefs : Connecteurs Data – Economie – Automatisation – Base de données

Développement

Le Pôle Economie de l'IFIP (Institut du Porc) analyse les évolutions des marchés et les tendances économiques et sociales de long terme de la filière porcine. Pour ce faire, il collecte des données issues de nombreuses sources de divers horizons géographiques (France, Europe, Monde) et de différents types (statistiques publiques, bases de données privées, entreprises, presse, fédérations, etc.). Il produit également ses propres indicateurs, analyses et tableaux de bord.

Historiquement, le système d'information du pôle économie était basé sur des saisies manuelles dans une base SQL, des macros VBA et de nombreux fichiers Excel. En 2020, le pôle économie a souhaité restructurer ce système en utilisant le potentiel de R pour automatiser la collecte et le stockage des données, supprimer les doublons et faciliter les analyses et leur diffusion. Cette nouvelle base de données porte le nom DEEP pour Données Economiques des Elevages de Porcs et sa réalisation a été confiée à Tac Economics.

DEEP contient des données économiques sur la filière porcine (volume, cheptel, prix, taux de change, commerce international etc.) provenant actuellement d'une quarantaine de sources nationales et internationales, à fréquences différentes (quotidien, hebdomadaire, mensuel etc.). Des connecteurs développés sous R permettent d'extraire automatiquement les données provenant de ces sources en utilisant les API existantes pour certaines sources (BCE, INSEE, Eurostat, Destatis etc.) ou d'autres moyens (webscrapping, extraction de fichier csv, excel, pdf etc.). Ces connecteurs traitent ensuite les données, les transforment en séries temporelles et les enregistrent dans la base DEEP. L'exécution des mises à jour est planifiée

* Data Manager, IFIP – Institut de Porc, berengere.lecuyer@ifip.asso.fr

† Responsable des projets Data, TAC ECONOMICS, sandrine.lunven@taceconomics.com

automatiquement et quotidiennement. Le développement de cette base permet une grande flexibilité, notamment dans l'ajout de nouveaux connecteurs et également l'intégration de nouvelles séries, zones géographiques ou produits.

Des programmes R ont aussi été développés lors de ce projet pour faciliter l'interrogation de la base (à la fois pour des initiés mais également des débutants) : recherche d'informations dans la base et traitement de données. Ces fonctions d'interrogation/recherche/traitement font appel à l'API spécifiquement développée lors de cette refonte. Cette présentation vise ainsi à montrer comment R permet d'exploiter un système d'information de manière performante et accessible pour des organismes professionnels et des industriels.