

Alimentation en ligne d'une base de données à partir d'un fichier Excel

M.C. Quidoz^a and J. Fabre^b

^a Plateforme Systèmes d'Information en Ecologie
CEFE
1919 Route de Mende
34293 Montpellier Cedex 5
Marie-claude.quidoz@cefe.cnrs.fr

^b OSU OREME
Université Montpellier 2
Place Eugène Bataillon - CC 060
34095 Montpellier cedex 05
juliette.fabre@univ-montp2.fr

Mots clefs : Tableur, Base de données, Web.

Dans le domaine de l'écologie, les tableurs sont encore de nos jours des logiciels très utilisés. Les chercheurs les utilisent pour travailler sur les données qu'ils collectent sur le terrain (les saisies sont faites sur le lieu de collecte ou de retour dans leur laboratoire). Pour assurer la pérennisation de ces données et leur diffusion, il est nécessaire de les mettre dans une base de données. Le système d'information du CEFE est construit autour du SGBD PostgreSQL.

Pour réaliser le transfert des données stockées dans des tableurs vers notre système d'information, nous avons pris l'habitude d'utiliser le logiciel R qui, grâce aux bibliothèques disponibles (XLConnect, RPostgreSQL), rend cette tâche facilement automatisable.

Nous rencontrons cependant des problèmes lors de l'intégration et cela principalement au niveau des domaines de valeurs, car peu de contraintes existent sur les tableurs. La solution retenue était alors de charger toutes les valeurs dans des champs de type texte et de procéder ensuite à des vérifications à l'aide de commandes SQL. Cette solution est opérationnelle mais n'est pas satisfaisante.

Notre proposition a été de développer des fonctions R pour vérifier les données et refuser leur insertion en cas de présence d'erreurs. Les fonctions de vérification sont de type assez classique : est un entier ? un réel ? la valeur appartient-elle à un intervalle donné ? à une liste d'espèces ? Y a-t-il des valeurs manquantes ? ... L'originalité de notre démarche ne réside pas dans les fonctions R écrites, mais dans le but recherché : nous souhaitons en effet que chaque propriétaire de données soit autonome pour faire ses propres vérifications, d'où la proposition d'utilisation du logiciel R très répandu parmi les chercheurs de notre laboratoire.

De plus, notre base de données étant de plus en plus constituée de données collectées par différents producteurs de données, il était nécessaire que l'insertion des données puisse être réalisée depuis l'extérieur de notre laboratoire. Nous avons donc couplé le mécanisme de vérification avec une interface web en charge du dépôt du fichier sur le serveur, de l'exécution du script R et du renvoi d'information auprès du producteur.