

RICA, Outil d'interrogation et traitements SAS via le Web.

Jean-Marc Rousselle

Laboratoire d'Economie Forestière, UMR INRA-AgroParisTech, NANCY



2èmes journées de recherches en sciences sociales

INRA SFER CIRAD

11 & 12 décembre 2008 – LILLE, France

1. Le RICA.

1.1. Descriptif

Le Réseau d'Information Comptable Agricole (RICA), en tant qu'opération statistique communautaire, est instauré en France, en 1968, par application du règlement 79/65/CEE. Il a pour objectif de fournir des informations sur le fonctionnement technico-économique des exploitations agricoles pour suivre leur revenu et éclairer les décisions de la PAC. Le RICA couvre l'ensemble des exploitations agricoles d'une dimension économique supérieure à 9600 UCE "1986" soit 12 équivalents hectares de blé et employant au moins 0,75 UTA. La production des exploitations agricoles appartenant au champ de l'enquête représente près de 90 % de la production de la branche agriculture. Les règlements communautaires fixent le nombre minimum d'exploitations à sélectionner par pays membre (de l'ordre de 7100 pour la France avec une ventilation régionale précise).

L'échantillon est extrapolé, en s'appuyant sur les données des enquêtes de structure, afin d'être représentatif de l'agriculture professionnelle.

Le RICA est une base de données très utilisée par les chercheurs notamment à l'INRA. Les recherches portent sur les disparités de revenus, l'économie de la production, la modélisation des exploitations, l'analyse des effets de la PAC. De nouvelles questions concernent l'environnement, ce qui nécessite une adaptation du RICA.

1.2. Accès aux données

Historiquement les bases de données du RICA étaient hébergées sur les gros systèmes (mainframe) IBM de l'INSEE. Afin de pouvoir accéder à ces données depuis un poste INRA, il était indispensable d'utiliser un logiciel spécifique qui transforme un Micro-ordinateur de type PC sous Windows en terminal reconnu par les systèmes IBM (type terminal 3278). Une fois ce terminal configuré, il était nécessaire ensuite de se connecter sur le réseau interne de l'INSEE en transitant via le réseau privé (et payant) TRANSPAC. Une fois connecté sur le système de l'INSEE, le RICA pouvait alors être interrogé sous l'environnement d'exploitation spécifique IBM.

1.3. Organisation des données

Les bases de données du RICA sont stockées sous le format spécifique du logiciel statistique SAS. L'organisation des ces données est de type hiérarchique. Dispersées dans plusieurs tables SAS suivant leur fonction, il est nécessaire de bien connaître cette organisation pour pouvoir les manipuler. Les tables SAS référençant les informations comme, les caractéristiques générales, le bilan comptable, et les résultats des exploitations agricoles contiennent une observation par exploitation. Par contre, les autres tables comme les données sur les animaux, les produit animaux, les végétaux, les produits végétaux et les produits végétaux transformés, peuvent avoir plusieurs observations par exploitation, en fait autant que de produits ou d'animaux. Toutes ces informations peuvent être réorganisées grâce à une variable commune à toutes ces tables, qui est en fait un code d'identification de l'exploitation.

Il est donc indispensable de bien connaître l'organisation des données du RICA avant de pouvoir l'exploiter.

2. SAS

Les bases de données du RICA sont créées et enregistrées au format propriétaire de SAS (Statistical Analysis System). Donc pour éviter les transformations de celles-ci, SAS est le langage naturel pour exploiter les données. Ce qui fait que tout utilisateur potentiel des informations du RICA se doit de connaître le langage SAS. N'ayant pas d'environnement graphique sous le système IBM de l'INSEE, le nouveau produit « Enterprise guide » disponible depuis la version 9 de SAS n'est pas disponible. Il est donc nécessaire de connaître les principes, les règles, le fonctionnement et surtout la syntaxe du langage SAS.

3. L'idée

Comme il est possible de le constater, l'accès aux données du RICA est assez lourd, voire complexe. Non seulement d'un point de vue matériel par l'achat du logiciel d'émulation de terminal IBM et un abonnement onéreux au réseau TRANSPAC, mais également par un investissement assez important de la part de l'utilisateur qui doit connaître l'organisation des bases mais aussi maîtriser la syntaxe du langage de programmation de SAS.

Le but de l'environnement d'interrogation du RICA développé au LEF de Nancy, est d'essayer de se soustraire au maximum de ces contraintes. Normalement une personne qui ne connaît ni le langage SAS ni l'organisation des bases RICA est capable de pouvoir sortir des résultats assez rapidement et simplement. Seul l'utilisation d'un navigateur WEB comme Firefox ou Internet Explorer est nécessaire.

4. L'outil

4.1. Développement

L'outil a été développé en PHP, langage permettant de générer du code HTML. Le produit est également couplé avec une base de données sous MySQL qui recense toutes les caractéristiques de la base de données RICA, comme le nom des variables disponibles, dans quelle tables SAS elles se trouvent, sous quelle catégorie elles sont référencées, leur période de validité etc. . Initialement le logiciel tournait sur un serveur dédié WEB qui envoyait des requêtes SAS à une seconde machine hébergeant SAS, laquelle machine renvoyait les résultats à la première. Maintenant, avec l'utilisation de produit permettant de transformer n'importe quel PC en pseudo serveur WEB (comme par exemple EasyPHP, LAMP ou WAMP), l'ensemble du dispositif peut être exploité sur une seule et même machine. Si bien que pour pouvoir exécuter le logiciel, il suffit d'avoir une machine pouvant faire tourner SAS, d'y copier les bases du RICA et la transformer en serveur WEB.

4.2. Installation

Pour pouvoir utiliser le logiciel, la machine doit donc avoir le logiciel SAS installé, les bases de données du RICA doivent être accessibles depuis cette machine. Il existe un CD contenant le reste des outils nécessaires pour une utilisation sous un environnement Windows, à savoir le logiciel easyPHP, les programmes du logiciel lui même, ainsi qu'une notice détaillée d'installation au format PDF qu'il suffit de respecter. L'installation a déjà été faite sur plusieurs sites différents et n'a posé

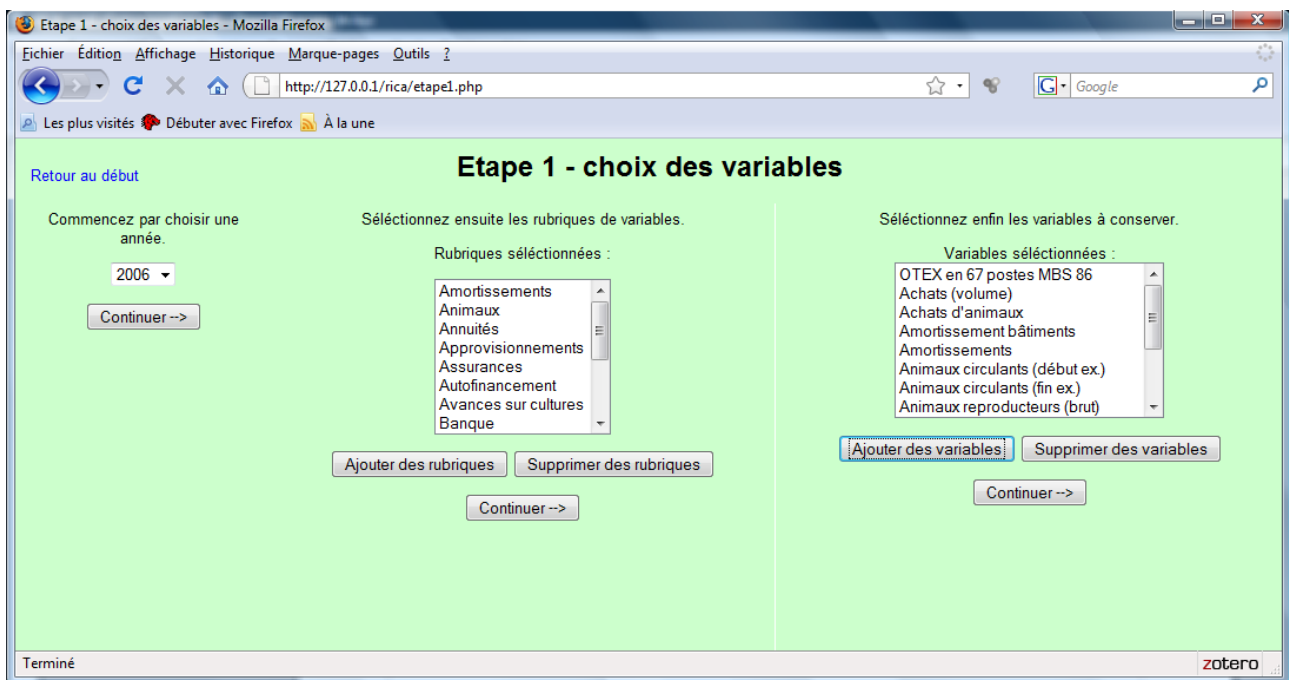
aucun problème.

4.3. Utilisation

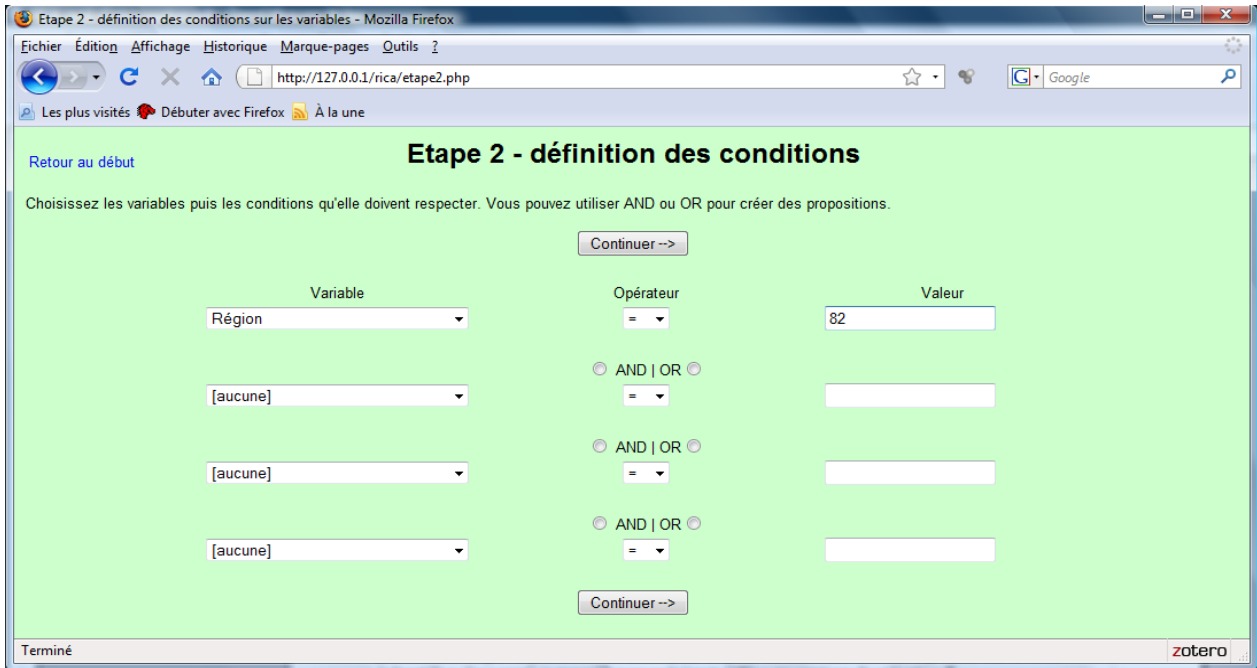
Nous avons essayé de développer l'outil grâce à notre expérience de l'utilisation et de l'exploitation des données du RICA, mais en ayant comme soucis principal de le rendre le plus simple possible d'utilisation. Le but recherché étant qu'il soit à la portée de personnes non familiarisées avec le langage SAS ou avec l'organisation des données.

L'utilisateur se retrouve en fait devant trois fenêtres qui s'enchaînent, et demandent sur quelles données précises il souhaite travailler, sur quelle sélection de l'échantillon et enfin, quels résultats ou sorties il souhaite obtenir.

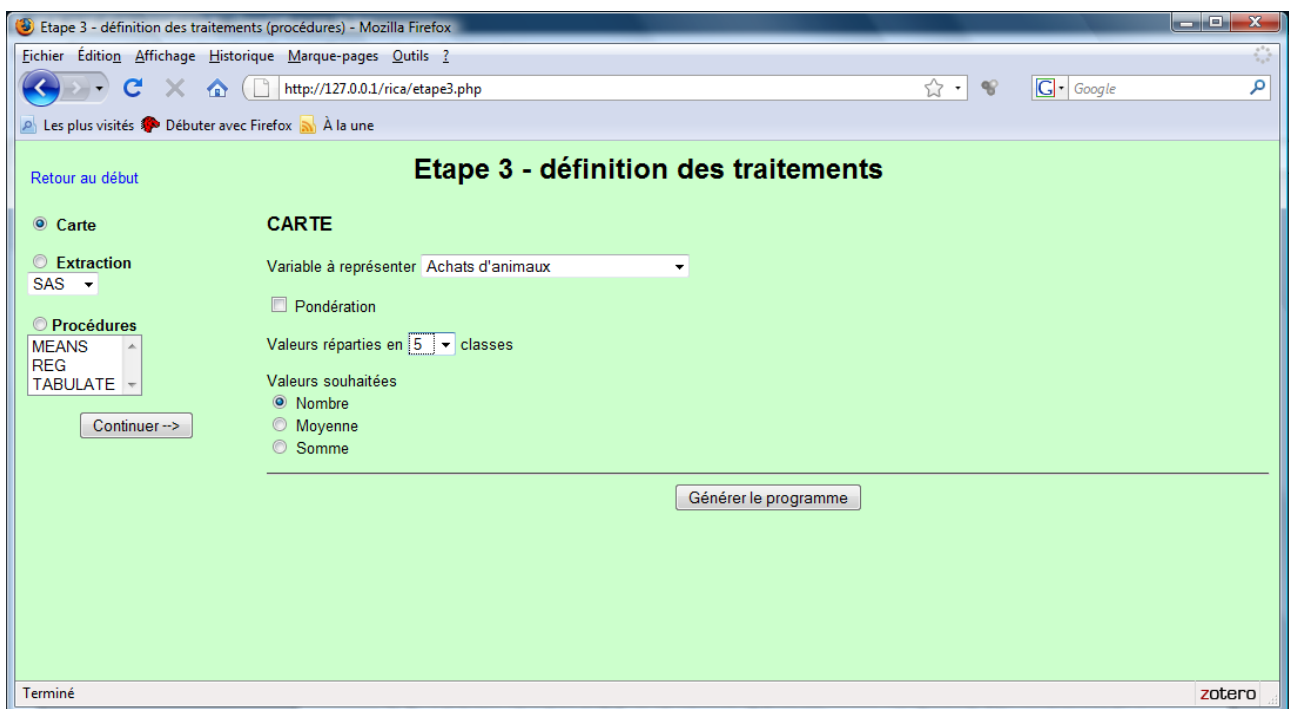
D'une manière plus précise, la première étape d'interaction avec le logiciel permet de demander à l'utilisateur sur quelle année il souhaite travailler, les rubriques de variables qui l'intéressent et enfin les variables elles mêmes.



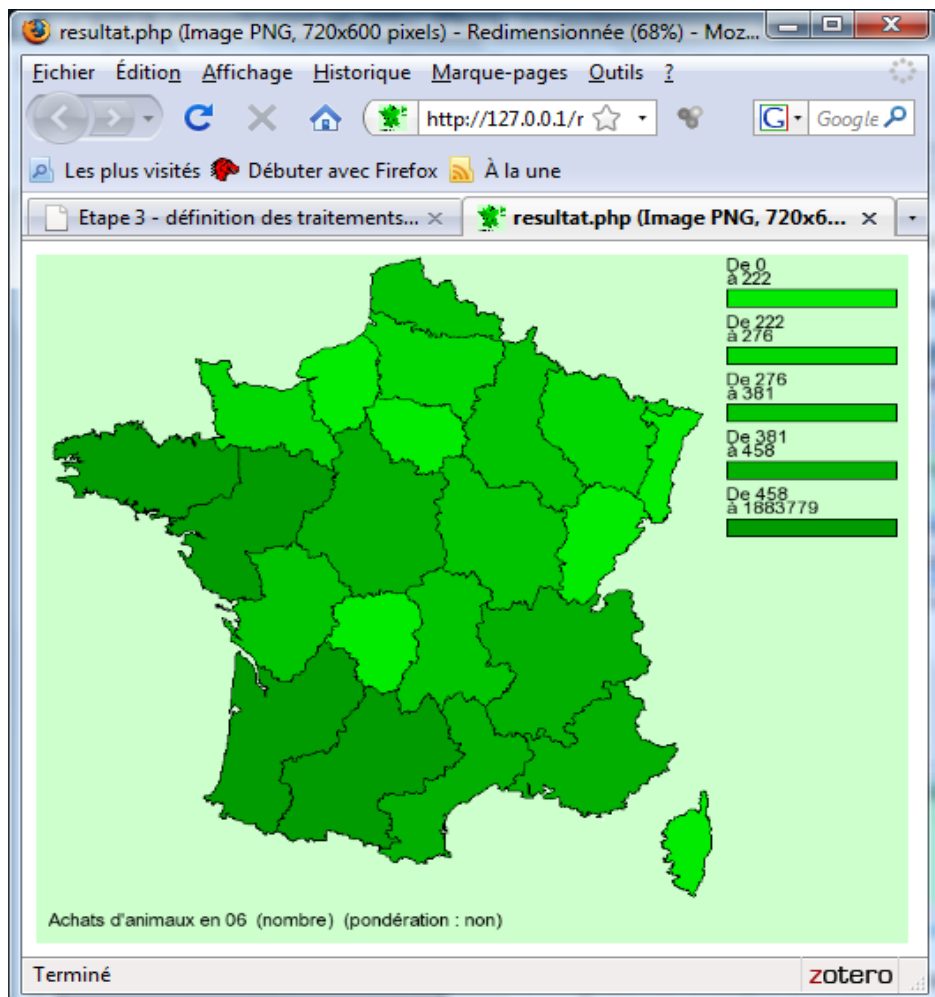
La seconde étape va permettre à l'utilisateur de pouvoir faire une sélection précise du sous échantillon sur lequel il va vouloir travailler en précisant différents critères de sélection.



Et pour finir, la troisième étape, où l'utilisateur va pouvoir choisir le traitement qu'il attend recevoir de l'exploitation des données ainsi sélectionnées. Il pourra par exemple demander des statistiques de base, faire une régression, demander également d'obtenir une extraction de ses données en format SAS ou en format récupérable sous un tableur. Il pourra enfin demander la génération d'une carte des régions françaises représentant des informations sur une donnée particulière.



Dans cet exemple il obtiendra une carte comme celle-ci.



4.4. Limites

Comme on peut le constater le logiciel ne prétend pas vouloir remplacer l'utilisation du logiciel SAS pour exploiter les données du RICA, mais propose un outil simple qui peut permettre à des personnes qui souhaitent avoir une première approche des bases du RICA de pouvoir le faire sans forcément avoir besoin de trop s'investir.

L'utilisation de cet outil reste quand même limitée à l'exploitation des données générales du RICA, car il n'est pas capable pour le moment d'exploiter les données de productions des exploitations comme les animaux, les produits animaux, les végétaux ou les produits végétaux. Il ne peut également travailler que sur une seule année à la fois

5. Conclusion et évolution

Le produit pourrait évoluer sans problème en y ajoutant des fonctionnalités comme par exemple travailler sur plusieurs années simultanément et en proposant la possibilité de constituer un échantillon constant. Il pourrait également se voir ajouter des traitements statistiques supplémentaires, ou donner les possibilités à l'utilisateur l'accès aux données concernant les animaux et les végétaux entre autre (avec la possibilité de créer ses propres agrégations).

L'intérêt de l'utilisation d'un tel outil réside essentiellement dans la possibilité d'avoir un premier contact avec les données des bases du RICA et ceci assez rapidement et en tout cas beaucoup plus facilement que par l'exploitation habituellement complexe des données.