

PACINFO: OÙ SONT PASSÉS LES 10 MILLIARDS DE LA PAC?

LE 13 JANVIER 2010 NICOLAS KAYSER-BRIL

9,94 milliards, c'est un netbook pour chaque Français. C'est le train gratuit pour tous pendant 6 mois. C'est le feu d'artifice du 14 juillet (celui du Champ de Mars, hein, pas de Agen) tous les jours de l'année dans toutes les capitales de région. C'est aussi ce que reçoit la France tous les ans au titre [...]

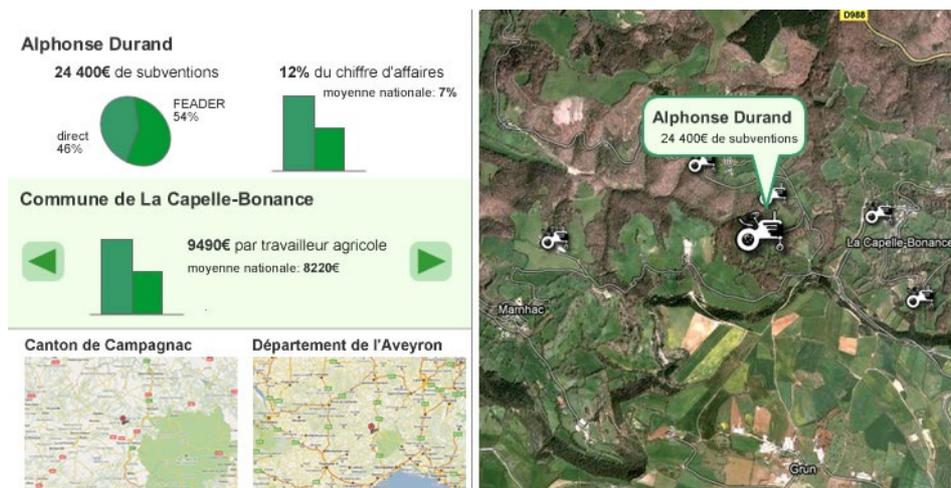
9,94 milliards, c'est un netbook pour chaque Français. C'est le train gratuit pour tous pendant 6 mois. C'est le feu d'artifice du 14 juillet (celui du Champ de Mars, hein, pas de Agen) tous les jours de l'année dans toutes les capitales de région.

C'est aussi **ce que reçoit la France tous les ans** au titre de la politique agricole commune. OK, on paye moins qu'on ne donne. Mais ça signifie que les autres prétendants à des subventions européennes sont désavantagés (régions, associations) puisque Bruxelles essaye d'aplanir le solde net de chaque pays.

Malgré des sommes colossales et des enjeux qui dépassent largement l'agriculture, les contribuables ne reçoivent que peu d'info de qualité. Depuis maintenant 9 mois, la liste des bénéficiaires de la PAC est en ligne. **Le site farmsubsidy.org l'a même remise en forme**, pour que les internautes ne soient pas obligés de subir le site du ministère de l'agriculture. Malgré cette manne d'info, les journalistes français se sont montrés très circonspects.

Pourtant, l'OLAF (les incorruptibles de l'Union) a trouvé en 2006 que **48% des vaches slovènes** n'existaient que sur les formulaires de demande de subvention. Plus tôt, c'était les Irlandais qui faisaient **tourner leurs moutons d'exploitation en exploitation** pour toucher plus de sous. Selon toute probabilité, les chiffres de la PAC recèlent bien d'autres histoires salaces.

L'année dernière déjà, j'avais lancé un ballon d'essai pour un projet de datajournalisme sur la PAC, sans résultat. J'ai mené depuis une plus grande étude de faisabilité, mais je ne peux pas aller plus loin, faute de temps et d'argent.



Je balance ici mes plans, dans l'espoir qu'un codeur (ou, qui sait, un rédac chef) la reprenne et la développe.

1. RÉCUPÉRER LES DONNÉES

Construire un programme pour récupérer les 135.000 points de données, c'est long. C'est pour ça que farmsubsidy.org a déjà fait le boulot et vous offre une base Access à peu près propre. (Le lien vers le fichier zip si vous le demandez dans les coms).

2. LOCALISER LES BÉNÉFICIAIRES

La base nous donne le nom et la commune du bénéficiaire. On peut donc les chercher dans

l'annuaire.

Les Pages Blanches nous offrent les adresses.

L'API de Google Maps nous permet de les placer sur la carte.

3. TROUVER DES DONNÉES HYPERLOCALES (NIVEAU: FACILE)

Pour chaque territoire, il serait intéressant de voir si les subventions reçues s'éloignent de la moyenne nationale.

La base de l'INSEE permet de faire des ratios de subvention par habitant ou par actif au niveau communal.

Le **recensement agricole de 2000** nous donnera, lui, des ratios par agriculteur, par hectare et par vache.

Grâce à de telles données, le journaliste et l'utilisateur peuvent très vite identifier les zones s'éloignant sensiblement de la moyenne. Et commencer à enquêter.

4. TROUVER DES DONNÉES HYPERLOCALES (NIVEAU: DIFFICILE)

société.com ou **Infogreffe** possèdent suffisamment d'infos sur les entreprises pour pouvoir établir des stats significatives sur le ratio subvention/chiffre d'affaire. Mais ces sites sont loin d'être complets ou structurés.

Le **cadastre** en ligne est une mine d'infos sur la structure des communes et des exploitations. Mais je n'ai pas encore trouvé de moyen de l'interroger automatiquement.

Wikipédia et sa petite sœur structurée **dbpedia** peuvent aussi se rendre utile, en donnant par exemple la couleur politique d'une commune.

NICOLASKAYSER-BRIL

le 13 janvier 2010 - 21:39 • SIGNALER UN ABUS - PERMALINK



Pourquoi prendre Agen comme ville pourrie? J'ai tapé "ville pourrie" dans Facebook, et Agen gagne, devant Béziers.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

CAROLINE GOULARD

le 14 janvier 2010 - 1:26 • SIGNALER UN ABUS - PERMALINK



Joli programme. J'aimerais pouvoir te dire que je relèverai un jour le défi...

Plus sérieusement, autant il me semble très intéressant d'établir des comparaisons par territoire (Commune de la Capelle Bonnance subvention par ouvrier agricole supérieure de 1270€ à la moyenne nationale) et de construire des visualisations permettant de les mettre en valeur.

Autant je suis sceptique sur l'éditorialisation de données nominatives, individuelles (Alphonse Durand, 24 000€ de subvention, et sa maison pointée sur Googlemap). Seule l'enquête de terrain permet ici de déterminer s'il est normal que cet Alphonse Durand touche ce montant de subvention ou s'il compte deux fois les mêmes moutons. Quel est alors l'intérêt d'attirer l'attention sur ce Mr Durand avant d'avoir fait l'enquête ?

Il me semble indispensable que ces données (même nominative) soient mises à la disposition du grand public, que la base de données soit libre d'accès. Mais je crois qu'un projet de database journalisme doit avoir, en plus, une vraie politique éditoriale, que cette réflexion éditoriale ne peut se satisfaire d'une mise en image indiscriminée des données disponibles, qu'elle doit, au contraire donner des clés d'analyse de ces données, par la construction d'indicateurs et de visualisations pertinents, selon les résultats de ses enquêtes.

Je me rend compte, en écrivant ce commentaire, que l'acte de réaliser une visualisation revient en fait à un éditorialiser des données, que rendre visible une base de donnée n'est pas neutre : cela ne relève pas de la simple action du web-designer, mais engage aussi une responsabilité journalistique.

[WORDPRESS HASHCASH] The poster sent us '0 which is not a hashcash value.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

ARIANE

le 14 janvier 2010 - 13:13 • SIGNALER UN ABUS - PERMALINK



Il serait intéressant de connaître également la répartition des subventions entre agriculture conventionnelle vs agriculture biologique. On peut le voir?

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

NICOLAS KAYSER-BRIL

le 14 janvier 2010 - 13:22 • SIGNALER UN ABUS - PERMALINK



Ariane,

Bonne idée en effet. Je sais pas si une base liste les exploitations certifiées. Faudrait demander soit au ministère, soit à Ecocert, mais la nature nominative de la certification fait qu'on peut pas invoquer la loi de 1978 sur les données publiques. Au mieux, on peut demander le nombre d'exploitations certifiées par commune.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

OLIVIER B.

le 14 janvier 2010 - 15:14 • SIGNALER UN ABUS - PERMALINK



Très bonne idée, il faudrait que cela soit consultable par beaucoup de monde.

Pouvez vous mettre le lien vers le zip de la base access faite par farmsubsidy.org ?

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

JACK THURSTON

le 15 janvier 2010 - 12:32 • SIGNALER UN ABUS - PERMALINK



Please forgive me for replying in English. My French is approximative at best, though sufficient to understand the article. We're working along these lines already. We did some mapping work on Sweden a while ago (you can see one of the prototypes here: <http://tinyurl.com/ycmfeou> It only works for Sweden and Denmark and is not totally up-to-date in terms of data. But you get the idea.

For France the problem is that we only have geographical locations by 5 digit postcode, not more than this. However, this should give us something to work with as the centre points of postcode zones is known. Moving to more precise location of recipients may require a crowd-sourcing approach. You suggest cross-referencing recipient names with other databases to find address, but this is difficult since most recipients are simply an name, e.g. Pascal Dupont. I can imagine the possibilities for incorrect matches is very great.

For anyone who wants to play with the data we have screen-scraped from the French Ministry site, you're welcome to it, we have made it available on a Creative Commons Attribution License. This means you're free to use and remix it as long as you attribute the source as 'farmsubsidy.org'.

Anyone interested in collaborating with us (and we do have some budget for developer time if we find the right people!) please get in touch. team (at) farmsubsidy (dot) org.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

NICOLAS KAYSER-BRIL

le 15 janvier 2010 - 14:54 • SIGNALER UN ABUS - PERMALINK



Hi Jack,

I just posted this to share the ideas I played with. Now, we need a developer to estimate the proportion of correct matches we can hope for. Crowdsourcing would be a good idea, but we'd need to start small, with support from a local paper or TV. Unlikely to happen, as scrutiny over the agribusiness isn't part of what they do.

Thanks for the link to the SE/DK map. Cool geocoding, but without clustering, it

resembles a population density map :) I think we can add more value to this dataset.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

NICOLASKAYSER-BRIL

le 16 janvier 2010 - 15:59 • SIGNALER UN ABUS - PERMALINK



@Caroline,

Complètement d'accord avec toi sur l'éditorialisation des visualisations. C'est là que le journalisme de données devient différent de la statistique ou de la gestion de base de données.

Concernant la PAC, t'as raison que la majeure partie de la valeur qu'on peut ajoutée, à faible coût pour nous, réside dans les données communales, voire cantonales.

Dans l'hypothèse – hautement improbable ;) – où les 3 premières étapes offrent des résultats hyper fiables et que l'étape 4 fonctionne, on pourrait avoir une carte cadastrale + images satellites, i.e. les outils qu'utilisent les mecs de la Commission pour détecter les fraudes. Reste plus qu'à trouver un moyen de pas faire trop "chasse aux sorcières" et ça peut donner une expérience de crowdsourcing intéressante.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

MIGKA

le 11 septembre 2010 - 17:15 • SIGNALER UN ABUS - PERMALINK



Thanks for the link to the SE/DK map. Cool geocoding, but without clustering, it resembles a population density map :) I think we can add more value to this dataset.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE