

« ON L'A PAS EN FORMAT ORDINATEUR »

LE 17 DÉCEMBRE 2009 NICOLAS KAYSER-BRIL

Un projet de database journalisme commence toujours par la récolte des données. [...] L'administration adore les bases de données. Et souvent, elle ne rechigne pas à les mettre en ligne.

Nicolas Kayser-Bril, data journaliste et blogueur invité, propose un retour d'expérience sur un projet qu'il publiera au printemps prochain



Un projet de database journalisme commence toujours par la récolte des données. Très opportunément, plein d'organismes étatiques adorent collecter des données, que ce soit le nombre de plantations de kiwi dans le Tarn-et-Garonne (187 d'après **Agreste**) ou le nombre de menaces ou chantage dans le Nord (384 en octobre d'après **Cartocrime**).

L'administration adore les bases de données. Et souvent, elle ne rechigne pas à les mettre en ligne.

Là où le bas blesse, c'est qu'elle fait tout pour que ces données restent cloisonnées, en dehors du web. La preuve ? Impossible de faire un lien vers les ressources que j'ai citées plus haut.

Je travaille en ce moment sur un projet agrégeant des données départementales. Ce qui implique d'appeler chaque préfecture. A quelques exceptions près, les fonctionnaires transmettent les fichiers demandés (faut dire que ce sont des données déjà publiées hors-ligne).

La qualité des documents, en revanche, laisse à désirer. 20% des fichiers sont au format Excel. 10% au format Word. Le reste, du PDF. En enlevant les tableurs faits avec les pieds et en rajoutant les pdfs convertis depuis Excel, environ 25% des données sont structurées, c'est-à-dire lisibles par un ordinateur.

Exemple imaginaire :

Donnée non-structurée : « Les classes de CM2 de l'école du quartier Saint-Jean sont fermées pour cause de grippe A. Les élèves concernés sont priés de rester chez eux, sauf ceux inscrits à la chorale, qui pourront répéter à l'école Jean Moulin. »

Donnée structurée :

Classes fermées				
Nom de l'école	Adresse	Classes concernées	Accueil des élèves	Remarques
Pierre Perret	12 av. St Jean	CM2	Non	Chorale assurée à l'école Jean Moulin

Au-delà du problème de la structuration des données, force est de constater que les administrations ont du mal à passer à l'informatique. Un nombre non-négligeable de

préfectures, environ 10%, n'y comprennent tellement rien que j'ai eu du mal à ne pas mourir de rire au téléphone.

La palme de l'inefficacité revient à celles qui ont pris le temps de *scanner des documents Word imprimés*. Mais le meilleur reste cette réplique d'un charmant fonctionnaire d'Evreux : « Désolé, on l'a pas en format ordinateur ».

Ouvrir les données, c'est bien. Les rendre exploitables, c'est mieux. Dans la plupart des **pays moins corrompus que la France**, l'accès aux données est une question réglée depuis plusieurs années. Le nouveau débat concerne leur format et leur facilité d'accès.

Le Sénat américain **propose le résultat des votes au format XML** depuis mai dernier. Beaucoup plus simplement, le gouvernement suédois tente de rassembler en un même portail l'ensemble des bases de données de ses ministères *en indiquant leur degré de structuration* (opengov.se).

Le problème a apparemment **été discuté** à l'**Opendata Barcamp** à Paris vendredi dernier (mais pas moyen de savoir quelles en ont été les conclusions).

—

» **Article initialement publié sur Databasejournalism**

Des réponses intéressantes aux questions soulevées par l'auteur ont été ajoutées en commentaires /-)

LE DODO

le 6 janvier 2010 - 13:38 • SIGNALER UN ABUS - PERMALINK



*Bravo Nicolas, tes analyses sont toujours pertinentes ! Dommage en effet que l'on n'ait pas les conclusions du dernier data-camp. J'avais prévu d'y aller pour la Voix du dodo, mais ai eu un empêchement de dernière minute. Gardons l'oeil bien ouvert sur cette question de la mise à disposition des données publiques, c'est un sujet stratégique pour l'avenir du journalisme, et de la démocratie.
Tatiana, alias le dodo.*

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

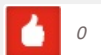
KATRINAMCNEIL35

le 23 août 2010 - 14:47 • SIGNALER UN ABUS - PERMALINK



Different people in every country get the credit loans from different creditors, just because it is fast and easy.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

3 pings

Les tweets qui mentionnent « On l'a pas en format ordinateur » | Owni.fr -- Topsy.com le 17 décembre 2009 - 20:00

[...] Ce billet était mentionné sur Twitter par damien douani, Owni. Owni a dit: #Owni « On l'a pas en format ordinateur » <http://bit.ly/68UoBn> [...]

[WORDPRESS HASHCASH] The comment's actual post text did not contain your blog url (<http://owni.fr/2009/12/17/%c2%ab-on-l-%e2%80%99a-pas-en-format-ordinateur-%c2%bb>) and so is spam.

Data journalism : pourquoi les médias français ne s'y mettent pas ? « Database journalism le 7 mars 2010 - 22:48

[...] Sur ce sujet, l'expérience de Nicolas Kayser-Bril est édifiante voir les articles « On l'a pas en format ordinateur ». [...]

[WORDPRESS HASHCASH] The comment's actual post text did not contain your blog url (<http://owni.fr/2009/12/17/%c2%ab-on-l-%e2%80%99a-pas-en-format-ordinateur-%c2%bb>) and so is spam.

Le Medialab de Cécile » Je suis aussi journaliste de données le 10 avril 2010 - 17:01

[...] culture...) très bien expliquées ici par Caroline Goulard, et «exemplifiées» par la lutte de Nicolas Kayser-Brill pour réussir à obtenir les emplacements des bureaux de votes français, à l'occasion des [...]

[WORDPRESS HASHCASH] The comment's actual post text did not contain your blog url (<http://owni.fr/2009/12/17/%c2%ab-on-l%e2%80%99a-pas-en-format-ordinateur-%c2%bb>) and so is spam.