

LES MAUVAISES DONNÉES DES MARCHÉS

LE 12 JUIN 2012 MARIE COUSSIN

Le portail français d'ouverture des données publiques, data.gouv.fr publie sur son site des données truffées d'erreurs, donc inutilisables. Le problème, identifié il y a plusieurs semaines, n'a toujours pas été traité. Nous avons demandé à Etalab, la mission chargée de la coordination du site, de s'expliquer.



Depuis peu, le site data.gouv.fr¹ met en avant sur sa page d'accueil trois sets de données particulièrement attractifs : la liste des attributaires des marchés publics d'État pour les années 2008, 2009 et 2010.

ACTUALITÉS

MÉTHODE ET ORGANISATION POUR FACILITER LE RECENSEMENT DES DONNÉES PUBLIQUES

Etalab, mission interministérielle chargée de l'ouverture des données publiques, placée sous (...)

□ □ □ □ ■ □

SUGGESTION DE RECHERCHE LES PLUS RECHERCHÉS

- Principales infractions en vigueur
- Composition des cabinets ministériels
- Inventaire du patrimoine immobilier de l'Etat
- Financement et dépenses de la sécurité sociale
- Statistiques du commerce extérieur
- Attributaires des marchés publics de l'Etat
- Effort financier de l'Etat en faveur des PME
- Pôles de compétitivité
- Avis de rappel de produits
- Adresse des évènements culturels français
- Adresses diplomatiques et consulaires

Etalab | [f](#) | [t](#) | Mentions légales | CGU | Actualité de data.gouv.fr | RSS

Or ces données contiennent de nombreuses erreurs (montants invalides, cases vides, informations incomplètes, etc.), rendant tout le fichier inutilisable. Une petite boulette qui récuse même la notion de "donnée ouverte" pourtant au cœur même du mouvement **Open Data** dont se revendique data.gouv.fr.

Simon Chignard, vice-président de la Cantine de Rennes, spécialisé sur ce sujet, donne la définition d'une donnée ouverte dans son livre *L'open data, comprendre l'ouverture des données publiques* :



Pour qu'une donnée soit ouverte, elle doit répondre à trois grands critères :

- **techniques** : les données brutes doivent être exploitables de manière automatique (i.e. par des programmes informatiques) et mise à disposition dans des formats les plus ouverts possibles et non propriétaires (par exemple : on privilégie le format .csv à .xls d'Excel),
- **juridiques** : les licences doivent clarifier les droits et les obligations des détenteurs et des réutilisateurs de données, elles doivent être les plus ouvertes possibles (par exemple : obligation d'attribution ou de partage à l'identique),
- **économiques** : peu ou pas de redevances tarifaires (susceptibles de constituer des freins à la réutilisation), tarification maximale au coût marginal, ...



Les explications d'Etalab

L'interlocuteur de référence sur ce sujet est Etalab, **mission** mise en oeuvre par le Premier Ministre François Fillon, en mai 2011 pour la création et de la gestion de **data.gouv.fr**. Bien que créée sous le gouvernement précédent, Etalab reste néanmoins en charge du portail, avec l'élection de François Hollande. La **Charte de déontologie** du gouvernement Ayrault, signée par les ministres, fait clairement mention de leur engagement envers l'Open Data :



Plus généralement, le gouvernement a un devoir de transparence. Il respecte scrupuleusement les dispositions garantissant l'accès des citoyens aux documents administratifs. Il mène une action déterminée pour la mise à disposition gratuite et commode sur internet d'un grand nombre de données publiques.



Owni a donc demandé à Etalab des explications sur les erreurs rencontrées. **Romain Tales**, responsable du recensement des données publiques, précise le parcours de publication sur data.gouv.fr :



Conformément à la circulaire du 26 mai 2011 instituant Etalab, un coordinateur a été nommé par ministère pour définir et transmettre les données à publier. Ce coordinateur initie les différents contacts dans les services, les directions, les bureaux, reliés à son ministère. Il peut ensuite créer une "équipe" qui a accès au back office du site. Cette "équipe" est globalement autonome sur le back office. Une personne est responsable de charger le jeu de données à mettre en ligne. Ce jeu de données est ensuite soumis à validation : relecture, vérification du contenu, des méta-données, etc. Jusqu'à trois personnes peuvent valider ce jeu de données avant que le responsable de publication de data.gouv.fr ne le publie définitivement.



La personne pouvant répondre précisément sur les erreurs du jeu de données ne dépend donc pas directement d'Etalab mais du ministère concerné. Dans ce cas, il s'agit du **ministère du Budget, des Comptes publics et de la Réforme de l'Etat**, plus précisément le **SAE** – service des achats de l'Etat. Romain Tales nous assure les avoir contactés et être, plus de 72 heures après notre premier appel, sans retour de leur part.

Owni n'est pas le seul à avoir relevé des incohérences dans ces jeux de données : **des internautes l'avaient d'ores et déjà exprimé** sur le forum mis à disposition par data.gouv.fr. Les commentaires relevaient notamment des données *incomplètes* et "*des valeurs extrêmes étranges*".

Le modérateur d'Etalab poste à chaque fois la réponse suivante (dernière en date : 5 mai 2012) :



Bonjour,
Le Ministère du Budget, des Comptes publics et de la Réforme de l'Etat a pris connaissance de ce dysfonctionnement.
Nous ne manquerons pas de vous tenir informé dès que le problème sera résolu.
Merci pour votre contribution sur data.gouv.fr



Sans éléments précis de réponse sur ce jeu de données, Romain Tales conclut néanmoins :



Le processus que nous avons mis en place est fait pour éviter ce genre de situation. Le problème avec ce jeu de données sur les marchés publics est plutôt l'exception, passée au travers des mailles du filet. Mais l'ouverture et la publication des données est encore quelque chose de nouveau, et de complexe, avec lequel nous devons nous familiariser.



Domage que les mailles du filet aient laissé passer autant d'erreurs², sur un jeu de données aussi emblématique. Florilège des plus importantes.

Des montants surréalistes

Par exemple, le rectorat de Montpellier³ aurait publié un marché ayant pour objet "*la localisation de données pour relier l'ensemble des sites du Rectorat de l'Académie de Montpellier*", attribué à France Telecom, pour un montant de plus de 30 milliards d'euros.

Owni a retrouvé le règlement de consultation accompagnant la publication de ce marché : il stipule que le montant minimum de cette opération sera de 30 000 euros et le montant maximum de 120 000 euros. Voilà d'où viennent donc les 30 milliards d'euros : 30 000 + 120 000 ont été accolés dans le fichier publié par data.gouv.fr.

Hello, you have an old version of Adobe Flash Player. To use iPaper (and lots of other stuff on the web) you need to **get the latest Flash player.**

D'autres montants semblent très suspects dans le fichier :



LA FRANCE
ENTR'OUVERTE

L'État a lancé son site data.gouv.fr. La France, enthousiaste, ouvre donc ses données publiques comme les États-Unis. ...

- plus de 1,4 milliard d'euros a été attribué à l'entreprise AMD Multicom pour "l'impression des documents (brochures, affiches et dépliant) du plan académique de formation et des innovations de l'académie pour 2009", toujours pour le rectorat de Montpellier⁴ ;
- 300 millions d'euros auraient été attribués par la préfecture des Hauts-de-Seine et les sous-préfectures d'Antony et Boulogne-Billancourt, pour de la location et entretien de vêtements.⁵

Le jeu de données 2010 comportant 16 231 lignes, il est impossible de vérifier l'ensemble des montants à la main. Les anomalies relevées sur quelques unes des lignes jettent donc le discrédit sur l'ensemble du fichier.

52 % des marchés

Sur ce même jeu de données, 7 519 lignes sont renseignées sans aucun montant relatif au marché public passé. A ces lignes sans montant, il faut ajouter les 1 010 lignes pour lesquelles la colonne "Montant" est effectivement remplie, mais de manière erronée : "611193.96partieforfaitaire-sansminimaxi" par exemple. Autrement dit, le montant du marché public est inconnu pour 52 % du fichier.

Parfois, des cases vides se cumulent : c'est le cas de deux marchés passés par le Ministère de la Défense avec la société Amesys (que nous **connaissons bien** à *Owni*) mais dont on ne perçoit ni l'objet ni le montant.

Sur des pans entiers du fichier, les colonnes ont été décalées : la colonne "Montant" a ainsi été renseignée avec les codes postaux. Pour un grand nombre de marchés publics, impossible donc de savoir qui a passé le marché. Dommage : on aurait adoré savoir quel ministère ou entité publique s'est réabonnée à la revue "La coiffure de Paris"⁶.

Si les erreurs relevées sur ce fichier ne peuvent servir de référence à la qualité des 352 431 jeux de données publiés sur **data.gouv.fr** ; leur accumulation, sur un fichier aussi emblématique de l'Open Data que les marchés publics et mis en avant sur la page d'accueil du site, mérite réflexion.

Analyse réalisée avec **Nicolas Patte**.

Photo Flickr CC Hans Gerwitz (**by-sa**)

1. Le portail français d'ouverture des données publiques, lancé en décembre 2011 [↔]
2. Pour ressortir les informations ci-dessous, les jeux de données ont été exploités dans Google Refine, qui permet de traiter de grandes masses de données [↔]
3. ligne 7159 du fichier 2010 [↔]
4. ligne 7667 du fichier 2010 [↔]
5. Ligne 7976 du fichier 2010 [↔]
6. Ligne 10963 du fichier 2010 [↔]

JP GILBERT

le 12 juin 2012 - 23:05 • SIGNALER UN ABUS - PERMALINK



Article très intéressant et utile. L'État peut mieux faire.... Mais quand sera affiché le budget de l'État, complet avec recettes et dépenses ? Et non par bribes et chiffres détachés du contexte comme c'est l'habitude ?

VOUS AIMEZ  3

VOUS N'AIMEZ PAS  0

LUI RÉPONDRE

3 pings

Open Data : 20 questions qui se posent au gouvernement français | ITACTU le 18 juin 2012 - 8:11

[...] sur data.gouv.fr souffrent d'une qualité parfois médiocre, budgets publiés en .jpg, données sur les marchés publics erronées, liens html des sites présentés comme des données ouvertes... Faut-il continuer à publier de [...]

Enjeux de la visualisation de données pour l'évaluation...et au delà | La Vigie de l'éval le 19 juin 2012 - 16:20

[...] L'essor du data journalisme permet enfin d'exploiter les données publiques...qui se font rares. On ne s'étonnera donc pas d'une mobilisation croissante des acteurs du journalisme et de l'internet pour « l'open data » dans a peu près tout les pays depuis une

petite dizaine d'année. En France, la reconnaissance gouvernementale de l'open data prend la forme d'Etalab encore embryonnaire, mais c'est un début (un début bien imparfait, comme le souligne OWNI dans son article : Les mauvaises données des marchés). [...]

Rencontres régionales de l'OpenData | Libres à vous — Formations le 3 juillet 2012 - 17:52

[...] site OWNI propose un bon article sur les travers autour de l'OpenData « Les mauvaises données des marchés », ainsi qu'un premier bilan français. Cette entrée a été publiée dans Actualités, [...]