

La fabrique de données ou "Oups on a fait un infolab" dixit Simon Chignard

Lors du **Forum des Usages coopératifs de Brest** nous avons co-animé un atelier sur la fabrique de données. Cet atelier a été l'occasion de tester une des étapes de **l'infolab citoyen**. Nous aurons l'occasion de tester l'infolab citoyen grandeur nature lors de "Paris Citoyens" événement de la Mairie de Paris qui se déroulera le 10 novembre.

Nous vous proposons de lire l'article de [Simon Chignard](#) publié sur son blog [donneesouvertes.info](#)

A l'occasion du **Forum des Usages coopératifs de l'Internet à Brest**, j'ai eu le plaisir de co-animer une session consacrée à **la fabrique des données** avec [Loïc Hay](#) de [La Fonderie](#) (agence numérique d'Ile-de-France) et la [Fondation Internet nouvelle génération](#). Ce billet retrace cette expérience pratique de mise en place d'un **infolab**, dans un temps et un lieu déterminé.



De droite à gauche : Denis Pansu (Fing), Loïc Hay (La Fonderie) et moi – crédit photo La Fonderie

1 – La fabrique des données

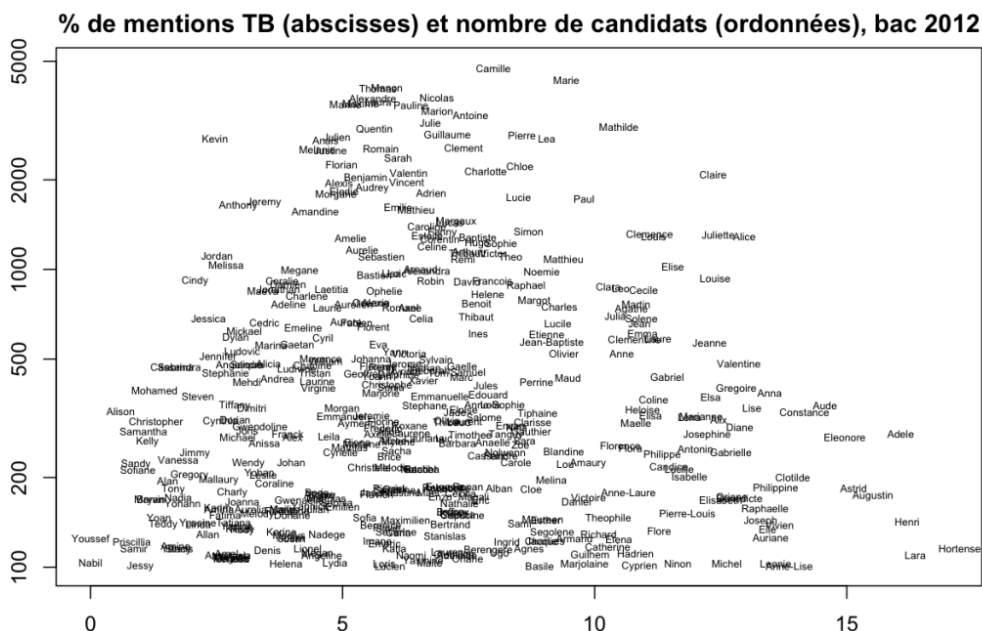
La fabrique des données propose d'**illustrer une démarche de réutilisation** de données ouvertes. De la recherche de la matière première, jusqu'à la réalisation de quelques infovisualisations, cet atelier combine dans un format court (2h30) **une approche critique** (d'où viennent les données ?) **et pratique** (comment les représenter ?).

2 – Le thème retenu : les prénoms

Nous avons retenu **la thématique des prénoms** pour ce premier atelier. Le prénom présente plusieurs avantages :

- d'abord on en a tous un ! (voire deux, trois ou quatre). Chacun peut se sentir concerné par cette thématique, a fortiori s'il a des enfants et s'est donc déjà retrouvé en position de choisir un prénom,
- ensuite, **la matière première est disponible** : les jeux concernant les prénoms les plus populaires sont disponibles sur les portails open data de [Paris](#), [Nantes](#) et [Rennes](#). Plutôt que de râler contre la non-disponibilité des données, utilisons celles qui sont déjà proposées !
- les jeux de données sont **faciles à appréhender et à comprendre**. Nul besoin de savoir développer une application mobile ou d'être un expert de la comptabilité publique pour s'en saisir.

Nous nous sommes ensuite appuyé sur une actualité de ce début juillet : la publication par **Baptiste Coulmont** (sociologue et auteur de "[Sociologie des prénoms](#)" aux éditions La Découverte) d'une **étude sur les prénoms des candidats** au bac ayant reçu la mention très bien. Elle révèle des succès très différents pour les Eleonore et les Jessica, les Augustin et les Kevin.



Prénoms et mentions TB au bac par Baptiste Coulmont (source coulmont.com/blog)

La représentation graphique fait réagir la salle, et elle est surtout pour nous l'occasion de souligner la **confusion fréquente entre corrélation et causalité** – ce n'est pas le prénom qui détermine le résultat au bac (contrairement à ce que laissent penser nombre d'articles de presse qui ont repris l'information) !

Le prénom est un marqueur d'un milieu social ou d'une région. Ainsi, Loïc explique qu'on lui demande souvent quelles sont ses racines bretonnes (réponse : aucune). Bref le prénom laisse imaginer – à tort ou à raison – beaucoup de choses sur celui qui le porte ... et sur celui qui le donne (voire sur celui qui le juge).

3 – D'abord, apprendre à lire les données

Après cette introduction sur les prénoms, j'aborde le **"tronçon commun"** de tous les ateliers que j'anime, c'est-à-dire une courte séquence pour expliquer la différence entre une donnée et une information, une donnée publique et une donnée ouverte... Donner des bases de compréhension me semble plus que jamais indispensable et c'est en tout cas **un pré-requis avant de pénétrer dans la fabrique des données**.

Nous proposons ensuite aux participants de découvrir les jeux de données disponibles sur les portails open data de Paris, Nantes et Rennes. Chacun est invité à suivre les liens à partir de son propre ordinateur. **J'ai volontairement fourni l'adresse des pages descriptives des jeux de données** (et non le lien de téléchargement) or la majorité de nos participants commencent d'abord par télécharger le fichier lui-même... Comment ce fichier a-t-il été constitué ? Que comprend-t-il ? Que nous raconte-t-il ? Quelle est la licence juridique applicable ? On ne peut répondre à aucune de ces questions sans consulter la notice de chaque jeu de données – c'est une démonstration **"par l'exemple"** et une première illustration de l'importance des métadonnées.



crédit photo La Fonderie

Une dizaine de minutes sont consacrées à **une lecture critique et comparée des trois jeux de données**. Les participants notent ainsi que les stratégies de diffusion ne sont pas les mêmes selon les villes. Paris ne distingue pas les filles des garçons pour les naissances intervenues avant 2011 – Camille par exemple est un prénom populaire dans la capitale. Rennes et Paris proposent un fichier consolidé pour plusieurs années, alors que Nantes a scindé chaque année dans un fichier spécifique – un moyen pas bien méchant mais pas discret non plus de "gonfler" artificiellement le nombre de jeux de données disponibles...

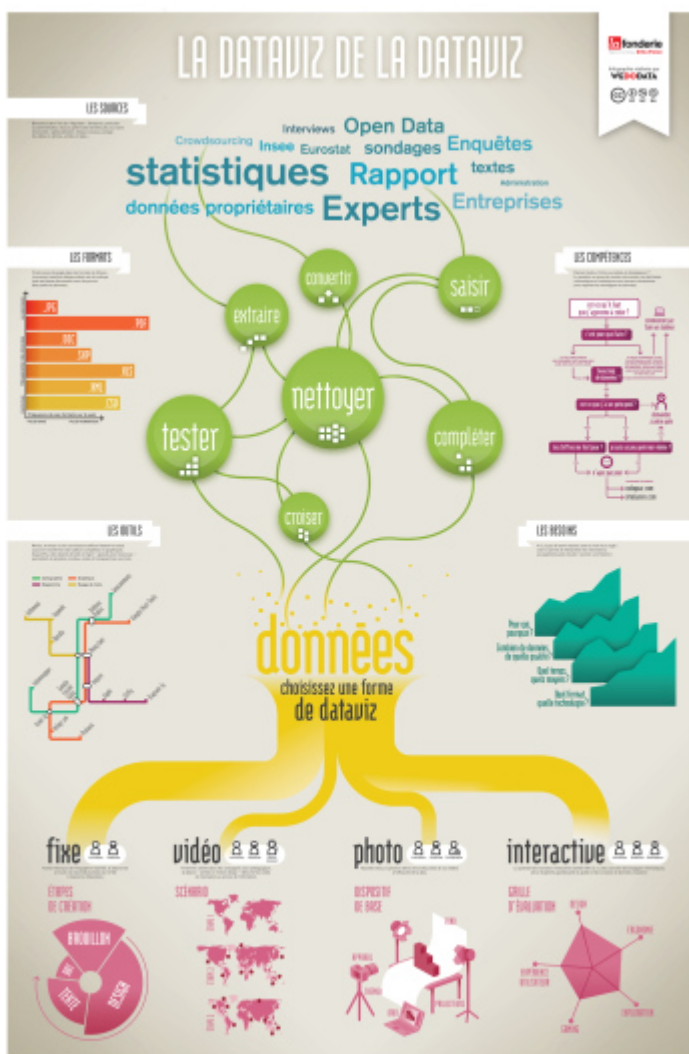
On constate aussi que d'une manière générale **les prénoms les plus populaires** – ceux qui figurent dans le top10 – **sont souvent les mêmes dans les 3 villes** : Emma, Manon, Matthis, ...

Les participants remarquent aussi, sur les portails de Rennes et Nantes, la mention d'une **soi-disant recommandation de la CNIL** sur les prénoms ayant été donnés moins de 6 fois au cours de l'année considérée (nous reviendrons dans un prochain billet sur cette "recommandation"... l'histoire vaut vraiment le détour !). Cela signifie en pratique que les fichiers ne comportent pas tous les prénoms donnés afin de respecter la vie privée des individus. Cela nous amène à **évoquer rapidement les problématiques d'anonymisation** à partir des données personnelles.

Ayant bien fait le tour de notre matière première, de ses atouts mais aussi de ses limites, je passe la main à **Loïc Haÿ** pour la suite de l'atelier. Maintenant que nous savons "lire" les données, on passe au niveau supérieur : l'écriture.

4 – Ensuite, apprendre à écrire

Loïc montre tout d'abord **deux exemples de visualisations** que l'on peut réaliser facilement : des "nuages de tag" reprenant les **150 prénoms les plus populaires** à Rennes et Nantes pour l'année 2008. Il explique ensuite comment les réaliser à partir du site wordle.net.



"La Dataviz de la dataviz" par WeDoData pour Expoviz – La Fonderie

La Fonderie, agence numérique Ile de France est à l'origine de l'exposition **Expoviz** consacrée à la visualisation de données. A cette occasion, l'agence WeDoData a réalisé le poster "**La Dataviz de la dataviz**" que Loïc nous détaille. Il insiste notamment sur la grande diversité des modes de représentation possibles des données (dont la photovisualisation). La parole est ensuite donnée à la salle : *et vous, comment aimeriez-vous représenter les données concernant les prénoms ?*

Léa Lacroix explique le travail qu'elle a réalisée pour son site **LesP titsRennais**, on évoque l'idée d'une photographie de petites Emma, Manon et Louise sur les marches d'un escalier, pour illustrer le classement qui change d'une année sur l'autre. **L'idée de classement revient souvent** et nous cherchons donc de l'inspiration du côté des **résultats sportifs**... Un participant nous fait à juste titre remarquer que l'on

devrait d'abord définir ce que l'on cherche à montrer – avant de chercher le bon outil pour le faire !

Loïc présente **différents outils de représentation de données** dont Many Eyes. Certains sont accessibles au plus grand nombre, d'autres réclament plus de temps pour les maîtriser.

5 – Oops, on a fait un infolab !

Revenons maintenant sur le titre de ce billet, "**oops, on a fait un infolab**". Le concept d'infolab a connu récemment un regain d'intérêt suite à l'article d'Internet Actu "[Avons-nous besoin d'infolabs ?](#)", article qui reprend les réflexions en cours à la Fondation Internet nouvelle génération sur **les modes d'appropriation des données**. Notre atelier brestois s'est d'ailleurs conclu par une intervention de **Denis Pansu** de la FING sur ce propos.

On sent bien que la problématique de l'animation autour de l'open data, de son accès à un public plus large que les seuls développeurs suscite de nombreuses réflexions – le sujet était d'ailleurs central lors de [la semaine européenne de l'open data](#). La Fonderie avec [Expoviz](#), ou moi-même avec les [ateliers autour des données de mobilité](#), **nous expérimentons de nouveaux formats d'animation et de transmission...**

Initialement une blague partagée avec Loïc, le titre de ce billet traduit aussi une conviction : ce dont nous avons avant tout besoin ce sont des médiateurs motivés (et si possible compétents)... qu'ils travaillent ou pas dans un "infolab".