

## OpenData – Données bien données

OpenData, en français, données ouvertes, est un gros effort national pour que les données publiques soient à la disposition de tous.

Données accessibles, pertinentes et réutilisables, avez-vous dit !!

Certes !

Nulle allusion à la nature des données recherchée. L'informaticien a beau jeu. Donnez lui des chiffres, des mots, des noms en grand nombre, il sera enchanté de fabriquer les outils qui permettront de jongler avec des milliers de mains.

L'ergonome est beaucoup plus humble - Ne me dites pas que l'informaticien est ergonome, il est tout le contraire. Son cerveau, qui manipule à l'envie une multitude de concepts abstraits, fonctionne à l'opposé du cerveau de l'utilisateur, qui est la proie du concret - L'ergonome a constaté que 90% des applications informatiques sont abandonnées faute d'empathie avec les utilisateurs.

Qui est à l'écoute de l'utilisateur ? Qui peut imaginer quelles sont les données qu'il souhaiterait voir dans l'instant de son problème, sous une forme immédiatement compréhensible ?

Ceux qui produisent les données n'ont pas la même vision que ceux qui les publient, qui eux-mêmes n'ont pas la même vision que les utilisateurs dans leur grande diversité.

L'utilisateur cherche des données de façon nébuleuse. Il suppose que la donnée existe, mais il ne sait pas sous quelle forme, ni où elle peut se trouver. Il la recherche dans le cercle étroit de ses relations, n'y trouve que des approximations, voire des contradictions ou la nécessité de développer une procédure soit-disant évidente, mais qui se révèle infertile. Alors, il baisse les bras et s'en remet à son intuition, ou à celle de son supérieur hiérarchique qui travestit à son tour le besoin.

Du coup, le producteur de la donnée ne reçoit pas les demandes des utilisateurs. Il imagine que ses données n'intéressent personne et se démotive dans la publicité et dans la

présentation des données. Le cercle est vicieux : d'un côté "je ne trouve pas ce que je veux, je ne cherche plus" et de l'autre : "personne ne s'intéresse, alors à quoi bon".

C'est un paradoxe souvent vécu : l'utilisateur ne sait pas ce qu'il cherche. A force d'essais-erreurs, il trouvera peut-être quelques éléments, mais au compte-goutte, à un prix exorbitant.

L'autre paradoxe est que l'étude des besoins ne révèle pas les besoins. L'utilisateur ne connaît ni les jeux de données existants ni les outils de fouille des données, car ce n'est pas sa compétence. Il est pour lui difficile d'imaginer les données nécessaires à son métier et donc la forme sous laquelle il pourrait en extraire les évidences.

Un nouveau métier est à inventer : \_\_anticipateur\_\_. Celui qui décèle les problèmes, en concertation avec l'utilisateur, qui identifie les données à exhumer et la forme sous laquelle elles peuvent être présenter pour aboutir à : "Mais bon sang, c'était bien sûr!"

En attendant, il se trouvera quelques créatifs pour trouver fortuitement les bonnes solutions. C'est à ceux-là que je dédie mon commentaire.

Un autre problème concerne la fiabilité et la validité des données. Le contexte de production est en général mal connu de l'utilisateur. Les outils statistiques ne peuvent économiser une phase de qualification préalable des jeux de données traités. Un capteur à défaillance intermittente, une définition imprécise de la grandeur mesurée, un mélange entre blocs de mesure fortement dépendant d'un paramètre instable sont quelques-uns des biais statistiques qui invalide les résultats. Le traitement statistique lui-même est sujet à erreur de programmation qu'il faut savoir piégé par une méthodologie parallèle d'analyse.

Entre l'intuition et la démonstration, cherchons la prudence et l'humilité.