

Quand nous serons tous des dataminers

 www.zdnet.fr/blogs/ils-ont-blogue/quand-nous-serons-tous-des-dataminers-39601426.htm

Je souhaitais rebondir sur [le récent et excellent article d'Hubert Guillaud](#), d'InternetActu, relatif à l'intelligence des données. Il y explique en substance que la masse de données brutes en accès libre va être décuplée dans les années à venir, "*par exemple accéder aux données de tel capteur, de telle caméra ou de tel moniteur (...) tracer n'importe quel évènement du monde réel (...) fouille(r) de(s) données (...) depuis chez soi, permettant d'analyser les statistiques de la criminalité ou de la circulation dans sa ville, ou des informations (...)*" sera possible "*Pas seulement des masses d'information statiques et statistiques d'ailleurs, mais des données qui vont être de plus en plus dynamiques, parce qu'elles seront accessibles à distance et en temps réel bien sûr, mais surtout parce que ces données mêmes seront le résultat de flux de données eux-mêmes mouvants. De combinatoires.*"

Cette accessibilité, ainsi que la quantité et la variété de ces données va faire émerger des individus (nous? nos enfants?) que [l'économiste Ian Ayres](#) qualifie dans un ouvrage récent de Super-crunchers. C'est à dire des personnes qui, à titre professionnel ou privé, seront passés expertes dans l'exploration, la manipulation, le mélange, le "trituration" de ces masses de données, dans le but de leur faire dire quelque chose, les rendre intelligibles, en extraire des signaux faibles et émettre des prédictions sur l'avenir qui permettront d'améliorer la prise de décision (quasiment des [Mentats](#) en somme). Il en déduit que l'expertise et l'intuition seront petit à petit remplacées par des modèles statistiques car, à mesure que le temps passe, la quantité de données produite et enregistrée dans le monde, année après année, peut être envisagée comme une forme d'expérience ou de connaissance du passé et utilisée comme telle. Ainsi depuis plusieurs années déjà dans le sport professionnel outre-atlantique (basket-ball, baseball, football américain) l'intuition des entraîneurs est bien souvent remplacée par les statistiques des joueurs au moment du recrutement ou de la constitution de l'équipe (une quarantaine d'autres exemples [sur son site](#)). Pour ma part il me semble que l'intuition sera pourtant au coeur de cette activité et que c'est même elle qui fera la différence entre les bons et les mauvais dataminers. L'intuition c'est la part humaine, subjective, créative, interprétative que l'on applique à la réalité, dans le but de comprendre et d'anticiper les événements. Parce qu'aussi brutes que soient les données c'est toujours l'homme qui, au final, les transformera en informations et en connaissances. Pour cela il aura besoin de trouver les sources de données à exploiter et, comme l'ont montré [Edouard Chi et Peter Pirolli](#), du Xerox Palo Alto Research Center, c'est d'abord à son instinct qu'il se fie pour cela. Quant à la créativité, elle lui permettra d'imaginer des croisements improbables de données, susceptibles de générer de la valeur, comme cela est déjà perceptible avec le phénomène des mashups. La manipulation de données telle que la pratique le héros de Minority Report, tirée d'ailleurs de travaux en cours (voir [ici](#), [là](#) ou encore [là](#)), nous donne une idée de ce que pourront être les interfaces de demain.

Par ailleurs nous avons eu en 2007 une convaincante préfiguration de ce que permettront ces outils avec l'arrivée de services gratuits comme [Yahoo!Pipes](#), [Swivel](#), [ManyEyes](#) d'IBM, ou le tout récent [logiciel Strata](#), de la société Kirix. Ces trois derniers sont déjà des outils de traitement de données particulièrement puissants. ManyEyes vous propose par exemple d'exporter des sets de données, d'y appliquer des traitements statistiques, puis de les représenter en mettant à votre disposition des dizaines de types de visualisations graphiques. Strata est également très innovant puisqu'il permet d'intégrer des données dynamiques et de disposer de graphiques en quasi temps réel.

Après [la veille sur Internet en pleine démocratisation](#), le décisionnel, jusqu'à maintenant apanage de marketeurs et décideurs d'entreprises de taille respectables, a toutes les chances de devenir petit à petit et sous une forme simplifiée, l'outil de Monsieur Tout-le-Monde ([via la console Wii?](#)). Certains métiers vont devoir penser à se réinventer.