

# LE DATAJOURNALISME: VECTEUR DE SENS ET DE PROFITS

LE 17 AVRIL 2011 NICOLAS KAYSER-BRIL

Face à l'avalanche d'informations, les techniques de datamining permettent d'extraire du sens de bases de données. La confiance devient la ressource rare, créatrice de valeur. Et les médias peuvent s'en emparer.

Ce post reprend les éléments d'une réflexion amorcée avec Mirko Lorenz et Geoff McGhee dans un article intitulé *Media Companies Must Become Trusted Data Hubs* [en] et présentée à la conférence re:publica XI.

Chaque jour, nous produisons **deux ou trois exaoctets** [en] de données, soit 1 million de téraoctets. Dans le même temps, Facebook et ses 600 millions d'utilisateurs produisent à eux seuls 70 téraoctets, soit à peine 0.007% du total. Pour comparer, un journal papier traditionnel pèse entre 1 et 50 mégaoctets.

Si l'on veut synthétiser toute l'information produite en quelque chose de digeste pour l'utilisateur final, il faut résumer par un facteur de 100 milliards. Pas facile pour un journaliste.

Pour faire sens de cette hyper-abondance de contenus, les professionnels de l'information doivent adopter de nouvelles techniques. Dans un monde analogique, l'enregistrement et la restitution permettent de rendre compte de manière optimale de la réalité. Dans un monde numérique et connecté, la difficulté ne vient plus de l'enregistrement mais de l'extraction de données pertinentes. Pour les exploiter, il devient indispensable de s'approprier des connaissances en analyse de données et en statistiques.

Une fois équipé des bons outils, faire parler des masses de données devient possible. La plupart des opérateurs téléphoniques, tels **Vodafone** [pdf/en], **China Mobile** [en] ou **Verizon** [en], utilisent les données produites par leurs utilisateurs pour prédire les embouteillages, par exemple. De telles techniques pourraient également être utilisées par les journalistes pour prévoir les manifestations.

## Toute information est une donnée

L'aboutissement du datajournalisme reste de penser tout type d'information comme une donnée à mettre en lien et en contexte pour lui donner du sens. Si Vodafone est capable de transformer les informations de communication d'un simple GSM en service de prédiction des bouchons, les professionnels de l'information doivent être capables de réassembler et de manipuler les éléments pris dans l'actualité pour en extraire des tendances et du sens pour le citoyen.

L'information telle qu'on la consomme habituellement, sous forme de textes ou d'images, n'est que très partiellement compréhensible par l'ordinateur. Les masses d'information accumulées par les journalistes restent amorphes une fois publiées, alors qu'elles pourraient être valorisées en fournissant le contexte nécessaire à la compréhension des articles du jour.

Certaines initiatives vont dans ce sens. **L'International Press Telecommunications Council** (IPTC) vient de publier un format de description des données permettant d'identifier clairement les personnes, les lieux et les organisations impliquées dans un article. Ce standard, **rNews**, a été dévoilé le 5 avril. L'IPTC compte parmi ses membres l'AFP, la BBC et 27 autres acteurs majeurs des médias. Son pas en avant va peut-être accélérer le passage des médias au web sémantique et aux données liées.

## Médias liquides

Pour diffuser sur l'ensemble des plateformes et des écrans, les contenus doivent être capables de prendre différentes formes. Du SMS à l'infographie dynamique, l'information doit pouvoir être consommée dans n'importe quelle situation, y compris celles que nous n'imaginons pas encore.

La plupart des rédactions peuvent certainement prendre à leur charge la distribution de leurs contenus sur le web, via leur site. Une petite partie d'entre elles est capable de faire de

même sur l'iPhone, sur Android et sur l'iPad. Et malgré ça, aucun média français (à part **10 minutes à perdre**) n'est capable de se positionner parmi **les 100 app les plus vendues** sur iTunes.

En donnant la possibilité à tout développeur d'accéder à leurs informations, les médias peuvent diffuser beaucoup plus largement leurs contenus, sans se poser la question du support. Tout comme la plupart des titres papier externalisent l'impression et la distribution, les sites web peuvent se débarrasser de la diffusion et laisser les spécialistes – les développeurs – s'en charger.

Les **API** (interfaces permettant aux ordinateurs d'accéder directement à l'information) du Guardian (**Open Platform**) et du New-York Times (**Developer Network**) donnent une idée de ce à quoi pourrait ressembler un média réellement liquide. C'est ainsi que les médias pourront cesser de devenir des répertoires de données pour devenir des points d'échange obligés dans le parcours de l'information.

L'actif des médias, en plus de leur marque, reste leur audience. Sa valeur ne provient plus de sa capacité à être vendue aux annonceurs, mais des possibilités de l'intégrer au processus de création de l'information. Le *crowdsourcing* permet d'augmenter, d'après **l'expression d'Eric Scherer**, l'information préexistante. Que ce soit pour compléter une base de données (comme nous le faisons avec **prixdeleau.fr**) ou pour valider des informations (comme sur **InfluenceNetworks**), l'apport de l'audience différencie profondément un média, enrichisseur de données, d'une base de données « froide ».

## Le marché de la confiance

Les médias ont, jusqu'à présent, évolué sur un marché où ils offraient une information à leurs lecteurs ou spectateurs et une audience à leurs annonceurs. A une époque où l'information était une ressource rare et où les annonceurs ne pouvaient pas toucher directement leur public, les médias créaient de la valeur aux deux bouts de la chaîne.

Aujourd'hui, cette position n'est plus tenable. Les lecteurs peuvent contourner les médias et s'informer via de nouveaux canaux (Wikipédia, Facebook, les sites d'institutions ou d'ONG) et les annonceurs peuvent toucher leur audience sans passer par les médias.

Cette dynamique redessine complètement le marché des médias et de l'information. Pour obtenir l'attention de leurs prospects sur un marché hyper-compétitif, les annonceurs ont intérêt à investir massivement dans la qualité du contenu. Unilever, l'un des plus gros annonceurs traditionnels, pénètre ainsi le marché du divertissement en produisant un jeu vidéo en ligne pour ses glaces Magnum (**Pleasure Hunt**).

Des ONG, qui se seraient auparavant contenté d'envoyer des communiqués de presse aux médias de masse, publient directement les résultats de leurs études et de leurs actions sur leurs sites. Les ONG **sont devenues des médias comme les autres** [en]. De nouveaux acteurs de l'info, comme Wikileaks et OpenLeaks, l'ont d'ailleurs reconnu clairement. Le premier en offrant les télégrammes diplomatiques **évoquant le traité ACTA à la Quadrature du Net** [en], une association militant pour les libertés numériques, le second en faisant en sorte d'accueillir en son sein autant d'ONG que de médias.

Les institutions ne sont pas en reste, même en France. Le Quai d'Orsay prend particulièrement au sérieux son nouveau rôle de média. Sur Twitter, **@francediplo** compte sept fois plus de *followers* que **@France\_Info\_Com**, par exemple. Le ministère des affaires étrangères publie également sa propre série de web-documentaires, **Destinations**.

La rareté, aujourd'hui, ne réside plus dans l'accès à l'audience (ce pourquoi les annonceurs allaient vers les médias) ni dans l'accès à l'information (ce pourquoi les consommateurs se tournaient vers les médias). La rareté, aujourd'hui, réside dans la confiance à accorder à une information.

Dans un univers de contenus surabondants, produits par tous, la valeur vient de la relation de confiance que l'on arrive à tisser avec son audience. Un tel changement de paradigme met sur un pied d'égalité tous les producteurs de contenus, média y compris.

re:publica XI 110413 on Prezi

**XAVIER C.**

le 18 avril 2011 - 11:12 &bullet; SIGNALER UN ABUS - PERMALINK



Plusieurs remarques :

- "Dans un monde analogique, l'enregistrement et la restitution permettent de rendre compte de manière optimale de la réalité. Dans un monde numérique et connecté, la difficulté ne vient plus de l'enregistrement mais de l'extraction de données pertinentes."

Vous dites aussi que pour "rendre compte de la réalité" il faut contextualiser l'info et la relier à d'autre. Bien que je sois d'accord avec ça, il me semble que mettre de côté les méthodes et outils de recueil de l'info de côté du fait de leur surabondance est une erreur. La technique n'est pas neutre, elle fait partie du contexte de l'info, à mon sens.

**nkb: tout à fait d'accord (et merci pour ce commentaire très riche!). La technique n'est pas neutre. Je dis seulement qu'à une époque où les journalistes ne sont plus les seuls à maîtriser la technique, ils sont parfois plus utiles à vérifier les contenus produits par d'autres qu'à aller sur le terrain. Prenez l'exemple d'une manifestation: nombreux sont les participants (ONG, police, passants) à prendre des photos. La présence d'un photographe de presse "professionnel" n'est plus forcément obligatoire.**

L'analyse pertinente et méthodologique ("il devient indispensable de s'approprier des connaissances en analyse de données et en statistiques.") de l'info, la recherche de contexte etc., c'est-à-dire la volonté de créer une info fiable plutôt qu'une info "scoop" rapproche le datajournalisme de l'analyse scientifique. Sans être la même démarche, les deux ont alors tout à gagner à emprunter des outils et des méthodes à l'autre.

Par exemple le fait de systématiquement donner ses sources est une obligation pour toute parution scientifique et ne peut être que bénéfique à un média qui chercherait à gagner en crédibilité et "facteur confiance". C'est d'ailleurs, pour les gens que je connais, une des raisons du succès d'OWNI.

**nkb: Le datajournalisme, c'est, en grande partie, la systématisation de l'utilisation des méthodes des sciences sociales par les journalistes. Et ce, depuis les premiers travaux de Phil Meyer en 1967.**

- Je vous entend parler de l'info en terme de consommation et de de la confiance en terme de marché et ça me fait un poil peur. Je vais peut être passer pour le gauchiste de base ou le gentil petit naïf mais non, tout ne se résume pas à l'économie en ce bas monde. Et ce malgré l'intrusion de celle ci dans un nombre croissant d'activité et surtout de mentalité.

Subordonner son action à l'économie pousse tôt ou tard à rechercher le profit plus que la qualité du service, à chercher une situation de monopole, etc. Attention je ne dit pas qu'il en faut pas penser à l'économie : il faut bien vivre, mais à mon sens à devrait être secondaire à la qualité de l'info.

Et j'ai bien compris que justement cet article dit que les médias vont désormais devoir miser sur la fiabilité. Sauf que vous subordonné cette recherche à la recherche du profit, ce qui introduit de nombreux biais. Par exemple ça n'enraye pas les dynamique grégaires des métiers médiatiques qui se sentent obligés de parler de la même chose, des "scoop", des "buzzs" et créent ainsi une information stéréotypée et limitée, une forme de censure par l'ignorance et non l'interdiction, de nombreux sujets.

**nkb: D'accord encore une fois. Mais à une époque où les modèles d'affaires s'écroulent, le seul moyen de ne plus penser à l'argent, c'est encore d'être profitable. C'est pour ça qu'il faut s'interroger sur là où se trouve la valeur, et comment la transformer en salaires durables. Une fois que l'on a réussi ce tour de force, on pourra se payer le luxe de ne plus y penser. En attendant, les industriels et les gouvernements sont ceux qui financent le plus avidement les médias – et ça, ça me fait peur.**

- La relation de confiance et la recherche de la confiance peut avoir des effets pervers comme le montre très bien l'influence des agences de notation. Un des principaux biais viens de la recherche de standards d'évaluation pour rendre comparable les éléments d'un même groupe (ce groupe étant bien entendu une construction sociale et non une réalité intrinsèque). Si la confiance se base sur de tels indicateurs et si la confiance devient la source de valeur, alors l'objectif ne va plus être la confiance elle même mais l'indicateur.

Par exemple l'évaluation des chercheurs selon leur taux de publication et leur taux de citation provoque des effets pervers catastrophiques en favorisant une diffusion prosélyte au service de l'ambition et de la carrière du chercheur plutôt que de la qualité et de la pertinence de sa recherche.

je conseille cet excellent article (<http://www.contretemps.eu/interventions/petits-conseils-enseignants-chercheurs-qui-voudront-reussir-leur-evaluation>) à ceux que ce dernier point intéresse, à titre d'exemple.

**nkb: Oui, les chiffres de la délinquance sont un autre exemple canonique de ce grand n'importe quoi. C'est là où le datajournalisme s'insère dans un mouvement plus global vers plus de culture numérique (numeracy) en complément de l'alphabétisation (literacy). Il faudrait apprendre à penser de manière probabiliste, par exemple, à avoir un regard critique sur les chiffres etc. Mais on en est loin.**

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

**XAVIER C.**

le 18 avril 2011 - 15:40 &bullet; SIGNALER UN ABUS - PERMALINK



*"En attendant, les industriels et les gouvernements sont ceux qui financent le plus avidement les médias – et ça, ça me fait peur."*

*Justement c'est un point sur lequel il est impossible de les concurrencer, leurs moyens économiques sont trop importants. Alors je veux pas dire qu'il faut pas se poser la question, et bien sûr il faut arriver à une situation assez stable pour assurer la pérennité du média. Mais je suis pas sûr que rechercher à tout pris le profit soit la meilleure voie. Je conçoit tout à fait que ce point de vue est celui de quelqu'un qui n'est pas directement confronté à ces problèmes. Mais quand je lis*

*"C'est pour ça qu'il faut s'interroger sur là où se trouve la valeur, et comment la transformer en salaires durables. Une fois que l'on a réussi ce tour de force, on pourra se payer le luxe de ne plus y penser."*

*Je ne peux pas m'empêcher d'être inquiet. Il s'agit effectivement d'un "tour de force" qui peut détourner la pratique, l'objectif final par les sacrifices qu'il peut imposer. Après je le répète, je parle un peu dans le vent dans le sens où c'est facile de dire ça de mon côté de l'écran et du clavier ^^;*

*Mais il me semble qu'il y a de nombreuses étapes intermédiaires entre "avoir pour fin ultime le profit" (Murdoch TM) et "se payer le luxe de ne plus y penser".*

*D'ailleurs j'avoue pas avoir très bien compris comment OWNI fonctionne, va falloir que j'aille regarder ça après les partiels (il me semble avoir vu passer un ou deux articles sur le sujet). Mais dans l'ensemble j'aime beaucoup le résultat : bravo, merci et continuez ;)*

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

#### 4 pings

Ma leçon de journalisme par Bob Woodward | Monjournalisme.fr le 20 avril 2011 - 0:26

*[...] ou au contraire de ne jurer que par les bases de données, outil parfois présenté comme le seul capable de donner du sens. Ainsi, Woodward explique bien que, quel que soit le « buzz » dont cet [...]*

Les enjeux des Data » Datajournalism – présentation le 26 avril 2011 - 14:38

*[...] NKB sur Owni : LE DATAJOURNALISME: VECTEUR DE SENS ET DE PROFITS [...]*

Extrait de Veille Open Data – Avril | Libertic le 17 novembre 2011 - 18:01

*[...] Le datajournalism et le standard rNews [...]*

Introduction au data-journalisme>Uppercult | Cultures et combats le 27 février 2012 - 16:32

*[...] Pour aller plus loin : Le data-journalisme : vecteur de sens et de profits Nicolas Kayser-Bril <http://owni.fr/2011/04/17/e-datajournalisme-vecteur-de-sens-et-de-profits/> [...]*