

École de Journalisme et de Communication d'Aix-Marseille

Le journalisme de données, une nouvelle forme de médiation de l'information ?

Mémoire de fin d'études dirigé par Mme Dominique Liautard

Xavier LACOMBE
Master CCN-J

« La quantité de données et d'informations disponibles, la complexité des problèmes à affronter, la digitalisation croissante des données et la diffusion des appareils mobiles encouragent le mélange des genres : analyse et narration se mêlent, donnant lieu à des expériences dynamiques et interactives de traitement des données et de l'information, qu'on peut difficilement ranger dans la case infographie »

Une autre façon de dire le monde, Courrier International n°1209

Remerciements

- Dominique Liautard, tuteur de mémoire, qui m'a aiguillé pour la méthodologie et le plan de mon mémoire.
- Nicolas Kayser-Bril, pionnier du journalisme de données en France et cofondateur de Journalism++, d'avoir pris le temps de répondre à mes nombreuses questions et de m'avoir guidé vers de nouvelles pistes de recherche.
- Nicolas Bocquet et Audrey Savournin, journalistes au journal La Provence, de m'avoir permis de travailler sur des projets de journalisme de données.

SOMMAIRE

Problématique : Le journalisme de données, une nouvelle forme de médiation de l'information ?

Introduction	6
I. Qu'est ce qui change ?	11
1. Le lecteur devient un collaborateur	11
a. Le crowdsourcing au Guardian	11
b. Le crowdsourcing à Owni	12
2. Des centres de données locaux mis à jour en temps réel	
a. Le Delaware Online, la base de données à son meilleur	13
b. Une base à jour pour de nombreuses utilisations	15
3. La tentation de l'information micro-locale	16
a. EveryBlock, l'information d'hyper proximité	16
b. Une évolution de la verticalité de l'information	18
4. Du consommateur d'information au conso-acteur	19
5. Un travail nouveau pour le journaliste qui suppose de nouvelles compétences	20
6. De nouvelles données	22
a. Le mouvement Open Data	22
b. Du chemin reste à faire	22
7. Une nouvelle visualisation	23
a. Visualiser pour mieux comprendre	23
b. La visualisation ne doit pas être systématique	26

II. Où en est-on aujourd'hui et quels sont les freins ?	28
1. Le journalisme de données : une pratique onéreuse	31
a. La tentation du crowdfunding	31
b. L'effet d'expérience	32
2. Un problème de formation et de compétences techniques	33
a. Une question de génération	33
b. Les écoles ne sont pas les seules à blâmer	34
3. Open data: l'enjeu de la libération des données	34
a. Des données libérées ?	34
b. Libérer les données, c'est bien. Les rendre exploitables, c'est mieux	35
4. Redonner de la valeur au travail journalistique	36
a. Mettre un terme au gaspillage	36
b. Les « Mediata Startups »	37
5. Les atouts séduction du journalisme de données	38
III. Conclusion	39
Bibliographie et webographie	43
Annexes	47

INTRODUCTION

En 2014, l'humanité produira 40 trilliards de données, un chiffre qui donne le vertige et que notre cerveau peine à représenter tant il est énorme. Un chiffre qui double tout les deux ans. Ainsi, au rythme actuel, les hommes produiront en 2020 autant de données qu'il existe d'étoiles dans l'Univers. Le terme anglo-saxon Big Data que l'on peut traduire littéralement par Grandes Données prend alors tout son sens. Nous évoluons bel et bien dans un monde dans lequel la production de données atteint un stade de profusion tel qu'il devient difficile de s'y retrouver.

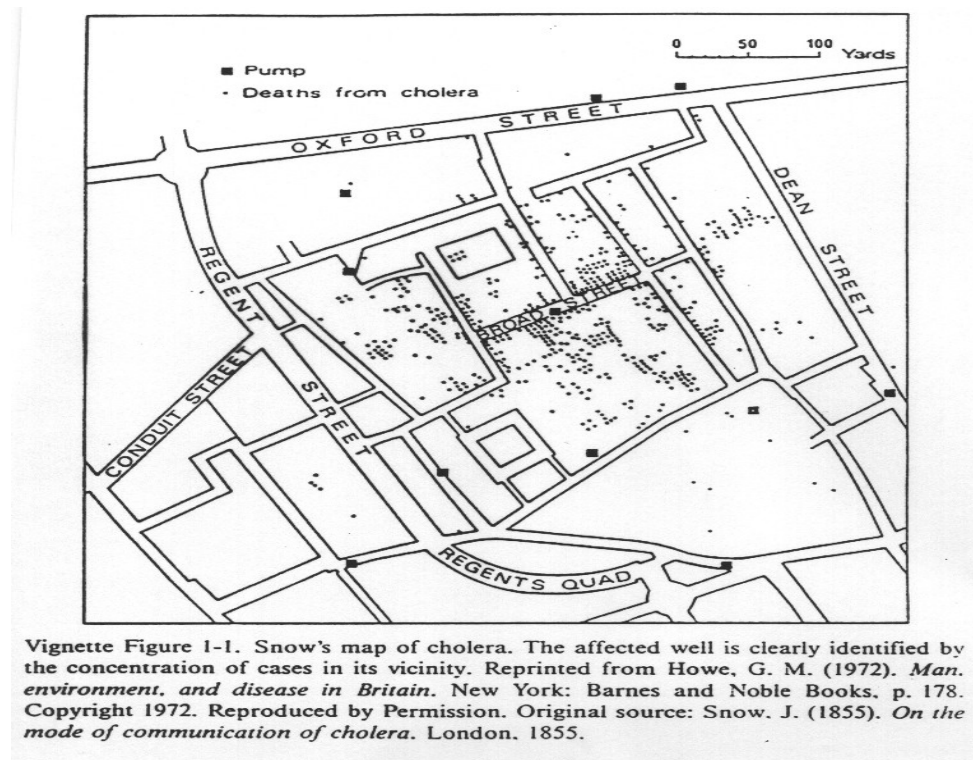
Pour caractériser ce déluge d'informations, le terme « infobésité » s'est banalisé dans les études sur les usages du web. Il désigne l'indénombrable masse de contenus textuels, sonores et visuels disponible sur le web. Il est issu d'une contraction des termes « information » et « obésité » qui traduit l'abondance d'information numérique présentée quotidiennement et sous toutes ses formes.

Cette quantité croissante de données ainsi que l'affirmation du numérique impactent de nombreux métiers dont le journalisme. Au point de donner naissance à une nouvelle branche baptisée data journalism ou journalisme de données (en français).

Celui-ci suppose l'étude de grandes bases de données. L'objectif, et c'est l'une des difficultés de cette nouvelle pratique, est d'établir des relations entre plusieurs données. Pour qu'une information complexe soit accessible, les liens détectés entre les données doivent être facilement compréhensibles par le public. C'est là qu'intervient la mise en image des données ou data visualisation, une forme nouvelle de restitution du travail journalistique à destination du public. Le journalisme de données ne devient réellement tel qu'à partir du moment où ceux qui le pratiquent ont la capacité de scénariser les données et les relations plus ou moins évidentes qu'elles entretiennent entre elles.

La data visualisation ou représentation graphique des données permet de répondre à un besoin de compréhension. Il s'agit de rendre intelligible une multiplicité de données.

Ce que l'on pourrait croire être une évolution logique des pratiques pour s'adapter à la profusion informationnelle d'aujourd'hui n'est pourtant pas nouvelle. Déjà en 1854, soit il y a 160 ans, l'épidémiologiste John Snow avait établi une carte des cas de choléra à Londres.



Depuis, les outils d'exploration des données et les formes de représentation visuelle se sont multipliés.

Il revient tout d'abord d'expliciter ce qu'on entend par donnée, un terme qui derrière sa prosaïque familiarité revêt plusieurs sens. Alain Joannès dans son ouvrage *Data Journalism. Bases de données et visualisation de l'information* donne quatre acceptations. La première définit une donnée comme "quelque chose que chacun est censé avoir en mémoire ou présent à l'esprit."

La seconde acceptation est la définition qu'en donne l'Association Française de Normalisation (AFNOR) "un fait, une notion ou une instruction représenté sous forme conventionnelle convenant à la communication, l'interprétation

ou au traitement par des moyens humains ou automatiques". La troisième manière de définir une donnée est celle de "brique élémentaire d'information". Enfin, la quatrième acceptation définit la donnée comme "l'élément abstrait nécessaire à un raisonnement mathématique ou philosophique".

A partir de ces différentes acceptations, il est possible d'esquisser un début de compréhension pour le terme "donnée" au sens journalistique. Les données sont des parcelles d'information qui supposent d'être traitées pour produire des informations. Il s'agit de croiser les données, de trouver des liens entre elles pour produire une information.

Les informations personnelles fournies par une personne lorsqu'elle remplit un formulaire d'identification sont un bon exemple pour expliquer ce raisonnement. La plupart du temps, il s'agit de micro-informations: civilité, âge, profession, lieu de résidence, état matrimonial,... Plus ces informations sont nombreuses, plus elles permettent de dresser un portrait précis de l'individu. Cette collecte de données permet de signaler des individus mais aussi de décrire des comportements. Ainsi explique Alain Joannès : *"Les grandes réussites technologiques et commerciales du web, comme Google ou Facebook, fonctionnent à la manière de puissants aspirateurs de données comportementales. Transformées en informations, ces données génèrent une matière première très recherchée par le marketing et la publicité. D'où la puissance économique de Google et de Facebook qui collecte ces données auprès de centaines de millions de personnes à travers le monde ».*

Un journalisme de précision

Les données parce qu'elles sont des émanations chiffrées de la réalité font du journalisme de données un journalisme de précision. Une expression inventée par Philip Meyer, dans les années 70's dans son ouvrage éponyme *The New Precision Journalism*, pour qualifier les nouvelles techniques de recherche d'informations utilisées par les journalistes. Elles-mêmes inspirées des méthodes de recherche issues des sciences sociales et comportementales. Voici un exemple d'Alain Joannès : *"Dix millimètres de pluie est une donnée. Elle se distingue de la bribe d'information par une quantification. Facile à réaliser*

avec n'importe quel appareil numérique, la représentation photographique de cette donnée quasiment palpable peut générer une visualisation beaucoup plus expressive qu'une succession ou un tableau de chiffres. Il suffit de photographier le même pluviomètre du même point de vue tous les jours à la même heure pendant trente jours dans le but de réaliser un diaporama ; le fait de limiter au minimum la durée d'exposition de chacune des trente images produit une visualisation spectaculaire de la pluviométrie pendant un mois quelconque à un endroit donné".

C'est là qu'intervient le rôle crucial du journaliste, à savoir celui de trouver des relations entre les données pour créer du sens.

Alain Joannès explique : « la détection des relations de causes à effets est au cœur de l'investigation, qu'elle soit scientifique, policière ou journalistique. Les grandes découvertes et les belles enquêtes reposent sur la curiosité et le discernement appliqués aux liens de causalité entre des phénomènes apparemment étrangers l'un à l'autre ».

$$A + B = C$$

La donnée C existe parce qu'elle est la conséquence de la donnée A et de la donnée B. Deux exemples pour illustrer cette équation.

- En 1989, Bill Dedman reçoit le prix Pulitzer suite à son investigation *The Colour of Money*. Une étude qui démontre, à partir de chiffres collectés et croisés sur des cartes, que les Noirs obtiennent moins de prêts bancaires que les Blancs à Atlanta.

- Autre cas, la aussi récompensé par un prix Pulitzer, l'enquête de Stephen Doig. Alors journaliste au quotidien local *The Miami Herald*, il enquête sur l'ouragan Andrew qui, en août 1992, a laissé des millions de maisons en ruines dans le sud-est des États-Unis et notamment en Floride. Dans un État pourtant habitué aux catastrophes naturelles, il était légitime de se penser à l'abri de telles destructions. Stephen Doig se pose la question suivante : pourquoi l'ouragan a-t-il causé autant de dégâts ? Il collecte, aidé par ses confrères, des données sur les maisons détruites : leur localisation, leur année de

construction et leur valeur immobilière. Son enquête lui permet de découvrir que les maisons les plus récentes ont été les plus touchées par l'ouragan. Dans l'article « *What went wrong* », il décrit un système de corruption à l'échelle locale articulé entre les promoteurs immobiliers soucieux de réduire leurs coûts de construction et des représentants politiques qui ont assoupli la réglementation immobilière. Parmi d'autres modifications, les promoteurs sont ainsi autorisés à utiliser des matériaux moins chers ce qui leur permet d'accroître leur rentabilité.

La relation entre données peut aussi être trouvée entre deux facteurs qui ne dépendent pas l'un de l'autre mais qui peuvent se comparer. Les données pluviométriques de juin 2012 ne sont pas à l'origine des données pluviométriques de juin 2013 mais comparer ces données peut permettre d'établir une aggravation des phénomènes météorologiques.

C'est au journaliste que revient ce rôle de creuser les données qui sont accessibles pour créer des relations de causalité ou de comparaison.

I. QU'EST CE QUI CHANGE?

Une nouvelle pratique s'accompagne forcément de son lot de changements. Ceux-ci interviennent tout au long de la chaîne de l'information et touchent à la fois le travail du journaliste mais également le récepteur.

Le journalisme de données s'appuie sur le triptyque Visualisation-Interaction-Participation. Les deux volets visualisation et interaction sont traités plus en avant dans le document.

1. Le lecteur devient un collaborateur

a. Le crowdsourcing au Guardian

Le volet participation fait référence au crowdsourcing. Le journalisme de données met en jeu des quantités astronomiques de données. Lorsque le traitement informatique automatisé de ces bases est impossible, il n'y a d'autre solution que de s'y mettre à plusieurs. La rédaction d'un journal par manque d'effectif et de temps peut dans certains cas ne pas suffire. Les journalistes peuvent choisir de faire appel au public, on parle alors de crowdsourcing. Wikipédia définit ce terme comme *« l'utilisation de la créativité, de l'intelligence et du savoir-faire d'un grand nombre de personnes, en sous-traitance, pour réaliser certaines tâches traditionnellement effectuées par un employé ou un entrepreneur »*.

De nombreux exemples existent. L'un des plus célèbres est le scandale des notes de frais des parlementaires britanniques révélé par The Guardian. Comme l'explique Simon Rogers, alors responsable du département data du quotidien : *« Une chose dont The Guardian peut se vanter, c'est d'avoir beaucoup de lecteurs, de paires d'yeux. Si nous avons besoin de leur avis, nous pouvons compter sur eux. C'est ce que nous avons fait avec les notes de frais des parlementaires britanniques. Nous avons 450 000 documents et très peu de temps pour faire quoi que ce soit. Alors quelle meilleure solution que de proposer cette tâche à notre lectorat »*

Et le public a répondu présent, chaque lecteur s'est vu confier la mission d'analyse d'une ou de plusieurs notes de frais afin d'en relever les anomalies, les excès potentiels. Cette mobilisation a permis de révéler que plus de la moitié des parlementaires britanniques (390 sur 646) ont commis des irrégularités dans leurs notes de frais. Certains élus n'ont pas hésité à demander le remboursement de téléviseurs à écran plat, d'un abri flottant pour canards, de nourriture pour animaux,... mais aussi pour certains comme Alistair Darling et Geoff Hoon, alors ministre des Transports, le remboursement de dizaines de milliers de livres de frais concernant leurs résidences secondaires tout en habitant des logements de fonction à Londres.

Une manière efficace et à moindre coût d'analyser une grande quantité de données. Évidemment, tous les sujets n'engendrent pas une telle mobilisation. Simon Rogers a son idée sur la raison qui a entraîné cette mobilisation du public. Selon lui, avec un sujet comme les notes de frais, on touche à un problème qui atteint « *...personnellement les gens. Ceux-ci sont donc plus enclins à y consacrer du temps. Une bonne partie de notre crowdsourcing se fait grâce à l'aide de quelques acharnés. Avec les frais des députés, nous avons généré un énorme trafic au début qui s'est progressivement éteint ensuite. Mais il reste des gens qui parcourent compulsivement chaque page à la recherche d'histoires et d'anomalies : un de nos lecteurs a consulté 30 000 pages.* »

b. Le crowdsourcing à Owni

En France, le site Owni s'est essayé à la géolocalisation des bureaux de vote en recourant au crowdsourcing. Pour exploiter un document et construire une application à partir de celui-ci, il faut que les données qu'il contient soient lisibles par un ordinateur. Or, pour son projet Owni s'est heurté à des documents qui ne permettaient pas de traitement automatique par ordinateur.

Ainsi, les phrases du type "Le 14^e bureau se trouve à côté de l'église, place Michelet" sont incompréhensibles pour un ordinateur. En revanche, "numéro: 14 ; adresse: place Michelet" permet un traitement automatisé par ordinateur. Dans le cas présent, 92 % des préfectures ont envoyé les fichiers

comportant les adresses des bureaux de vote. 63 % des fichiers ont été traités automatiquement mais les 37 % restants ont dû être traités à la main.

Pour ce faire, Owni a demandé l'aide de ses lecteurs pour plusieurs tâches :

- Appeler les préfectures qui n'ont pas envoyé de fichiers ou se présenter directement à l'accueil.
- Transformer les PDF illisibles en données exploitables. Pour cela, Owni a développé une application qui permet aux lecteurs de saisir manuellement les informations contenues dans les fichiers inexploitablement informatiquement.

Au total, 300 internautes ont contribué au projet en saisissant pas moins de 13 000 adresses de bureaux de vote à la main.

2. Des centres de données locaux mis à jour en temps réel

Le journalisme de données participe à la création de bases de données qui sont actualisables en temps réel. Il s'agit là d'une différence fondamentale avec les autres formes de journalisme qui se périment très vite. Un article de presse écrite peut très rapidement se trouver dépassé, idem pour la radio et la télévision. Disposer d'une base de données régulièrement mise à jour permet de tirer des enseignements sur la durée.

a. Le Delaware Online, la base de données à son meilleur

Le Delaware Online, un quotidien américain publie une cartographie interactive de la criminalité de l'état du Delaware actualisée au fur et à mesure que de nouveaux crimes interviennent.

La page d'accueil de l'application (ci-dessous) se présente sous la forme d'une carte qui référence les crimes signalés à la police dans l'État du Delaware. Une recherche plus précise est possible en rentrant directement une adresse ou le nom d'une ville dans la zone de recherche.

La gradation de couleur tend vers le sombre plus le nombre de crimes est élevé.

L'application qui constitue un mini site au sein du site du Delaware Online remplit un rôle d'archive puisqu'il est possible d'accéder aux données détaillées depuis 2007.



Capture d'écran tiré du site Delaware online (<http://php.delawareonline.com/news/evergreen/crime/>)

En cliquant sur les zones colorées, on accède aux détails des crimes commis (cf ci-dessous) :

Number of serious crimes last month in grid 076326

Total: 13

Arson: 0

Drugs: 2

Property: 10

Violent: 0

Weapons: 0

Other: 1

View recent crime reports

Recent incidents for grid 076326

Crime Date	Agency	Description	Type	Group
3-15-14	NEW CASTLE COUNTY PD	Offensive Touching/Other Assaults/Non-Aggravated	Assault	OTHER
3-10-14	NEW CASTLE COUNTY PD	Offensive Touching/Other Assaults/Non-Aggravated	Assault	OTHER
3-10-14	NEW CASTLE COUNTY PD	Offensive Touching/Other Assaults/Non-Aggravated	Assault	OTHER
3-10-14	NEW CASTLE COUNTY PD	Contributing to the Delinquency of a Minor/Endangering the Welfare of a Child	Domestic	OTHER
3-05-14	NEW CASTLE	Larceny/From Building	Theft	PROPERTY

EJCAM 2013-2014

L'image de gauche montre le nombre de crimes commis le mois précédent. Sur un total de 13 crimes, 2 sont liés à la drogue, 10 à des violations de propriété et un est placé dans une catégorie non détaillée. L'image de droite présente le descriptif précis de chaque crime. On trouve la date de l'incident, le commissariat qui s'est occupé de l'incident, une brève description de celui-ci, le type d'incident et la catégorie dans laquelle il est enregistré au sein de l'application.

Avec toutes les données accumulées, l'application du Delaware Online permet de générer le bilan par année des crimes (cf capture d'écran ci-dessous).

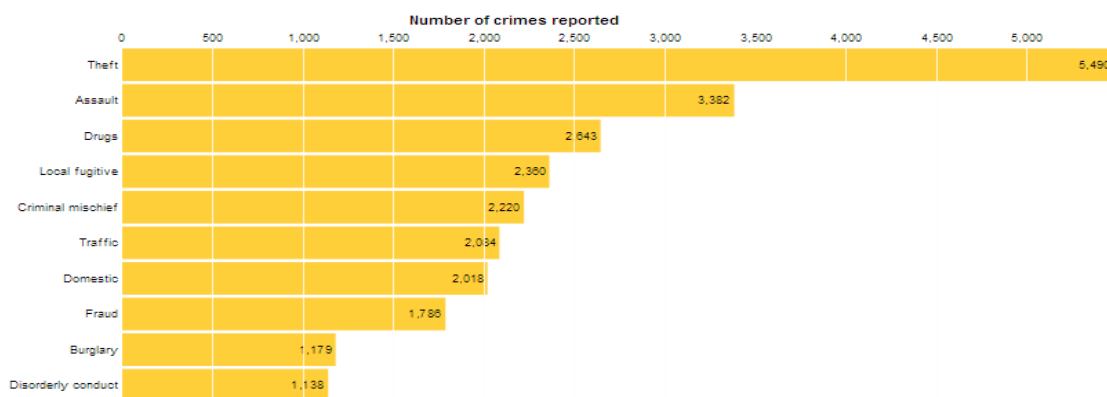
Top crimes by year

[View crime map](#)

Toggle through the years to see how the top ten crimes have changed in Delaware.

2014 ▼

ABOUT: All law enforcement agencies in Delaware report crimes to the Delaware Criminal Justice Information System, which provides a weekly report to The News Journal. To protect victims' privacy, the agency does not provide information about rape. Also, crime locations are based on a statewide grid system, rather than exact addresses.



b. Une base à jour pour de nombreuses utilisations

De nombreuses informations peuvent être tirées de ces bases de données/archives. On peut, par exemple, comparer l'évolution de la criminalité entre quartier, sur une période de plusieurs mois ou années.

Selon l'utilisateur, la base de données peut être utilisée différemment :

- Les lecteurs peuvent consulter les informations relatives à leur quartier, déterminer le niveau de quiétude d'un quartier en vu d'un déménagement,...

- Les journalistes peuvent s'appuyer sur ces informations pour mesurer l'efficacité, l'impact des mesures prises par l'État et/ou l'action de la police. Si la criminalité baisse dans une zone peut-être y a-t-il un lien avec la décision de l'État d'augmenter les effectifs policiers. A l'inverse, si l'application montre que le nombre de cambriolages a augmenté, peut-être faut-il creuser la piste d'une baisse d'effectif policier dans la zone.

D'une manière générale, une base de données actualisée est très intéressante puisqu'elle peut servir de point de départ à un grand nombre de sujets. Il s'agit d'une ressource très précieuse pour une rédaction mais aussi pour son lectorat.

3. La tentation de l'information micro-locale

On aborde d'ailleurs ici une tendance de fond du journalisme de données. Une caractéristique qui pourrait bien jouer un rôle déterminant dans la popularisation de cette pratique : l'information micro-locale. Il s'agit de l'information la plus à même d'intéresser un maximum d'individus.

a. EveryBlock, l'information d'hyper proximité

Adrian Holovaty, pionnier de l'information micro-locale, utilise d'ailleurs la formule suivante pour résumer l'intérêt du public : *« Longtemps, j'ai cru que le plus important dans le journalisme c'était de veiller à ce que les gens soient mieux informés. Mais être mieux informé n'est pas une fin en soi : le commun des mortels ne se réveille pas le matin en se disant : j'ai envie d'être mieux informé aujourd'hui »*. En revanche, la possibilité d'être informé de ce qui se passe au coin de sa rue intéresse tout le monde. Des éléments qui ne seraient pas mentionnés ailleurs constituent ici le cœur de l'information, la motivation qui pousse les visiteurs à consulter le site.

Voici l'idée qui a traversé l'esprit d'Adrien Holovaty pour son projet *EveryBlock*. Le site agrège des données hyper locales de 16 villes américaines. Ces données permettent aux internautes de suivre l'actualité de leur quartier et d'être en contact avec leurs voisins. L'objectif avoué est de créer un lieu d'information et d'échange entre voisins autour de sujets qui les intéressent tout particulièrement. Adrian Holovaty explique : *« Le site est simple à utiliser.*

Il suffit de créer un compte et de choisir les endroits sur lesquels vous désirez être informés, par exemple, le quartier où vous habitez et travaillez. Mise à jour en permanence, votre page d'accueil EveryBlock personnalisée vous montre ce qui s'est passé à proximité des lieux que vous avez indiqués et ce dont parlent les gens. »

The screenshot shows the EveryBlock New York City website. At the top, the logo "EveryBlock New York City" is on the left, and the tagline "A news feed for your block." is on the right. Below the logo is a search bar with the text "Enter your address, find news nearby" and a subtext "We gather all sorts of local info daily so you can follow news near you." The search bar has a placeholder "Address, ZIP or neighborhood" and a "Search New York City" button. Below the search bar, it says "New York City only, for now. Examples: 175 Fifth Ave., Manhattan, 10001, or Williamsburg". To the right of the search bar is a section "Explore the city" with a map icon, listing "5 Boroughs", "202 Neighborhoods", "199 ZIP codes", and "9,350 Streets". Further right is a "NEW HERE?" section with a question mark icon and a link to "READ OUR FAQ". Below the search bar is a "FEATURED NEIGHBORHOOD" section with the text "Or, explore other neighborhoods..." and a map of New York City. The featured neighborhood is "Clinton Hill". Below the map, there are two tabs: "Recent articles" and "Public records". The "Recent articles" tab is selected, showing a list of articles. The first article is titled "The Week in Crime: Bites, Bones and Stolen Phones" and describes a robbery. The second article is titled "Development Watch: 154 1/2 Washington Avenue" and describes a new building. The third article is titled "CLINTON HILL: One suspect collared in Pratt student beating" and describes an assault. Below the articles, it says "115 more articles in Clinton Hill in the last 30 days." To the right of the featured neighborhood section is a "LOCAL MEDIA" section with the text "Which neighborhoods are in the news?". It features a map of New York City with red circles indicating the number of articles for each neighborhood. The largest circle is in the center, with the number "61". Other circles are scattered across the city, with numbers like "2", "3", "13", "8", and "2". Below the map, it says "3,977 more articles in the last 30 days."

Capture d'écran du site EveryBlock

Le site fonctionne sur le même principe que Facebook à la différence qu'il est totalement orienté sur l'information de proximité.

Il est possible d'obtenir des informations sur presque tout: fête de quartier, crime, brocante, circulation, travaux, cours de langue, etc. Tous les centres d'intérêt sont couverts sur le site. La diversité des publications dépend du nombre d'individus qui y contribuent et de leurs centres d'intérêt. La base de données est donc potentiellement illimitée.

Capture partielle récapitulative des événements de la journée du 5 avril 2014:

SAT 05 APR		9am to 6pm	50 W Washington St.	March Against Child Abuse I will be leading the 2nd annual March Against Child Abuse for Chicago this year. The march will be held April 5th 2014 Nationwide, our planned location will be at the Daley Center. I would love for you support in this event. Please share this even Join the Facebook event and also like and follow...
		11:30 a.m.	LUX BAR	Birthdays in January, February, March and April - Brunch Celebration ! Chicago Downtown Girls
		3 p.m.	Harold Washington Library Center (400 S. State Street)	Pre-Teen Book Club Join us to discuss Who's Haunting the White House? by Jeff Belanger, while hanging out with friends, playing games and snacking at our pre-teen book club meeting. For children ages 9 to 12 years old. This program takes place in the Children's Program Room of the Thomas Hughes Children's Library.
		10:30 a.m.	Harold Washington Library Center (400 S. State Street)	Saturday Lapsit Series This 25-minute weekly session for children 6 to 35 months old emphasizes adult-child interaction and presents books, fingerplays and songs for babies, toddlers and adults to share. For additional information, please call (312) 747-4200. This program takes place in the Children's Program Room of t...
		9:30 a.m.	Harold Washington Library Center (400 S. State Street)	Tax Assistance Program Ladder Up provides free electronic preparation and e-filing of tax returns for individuals, total income up to \$25,000/year, and for households, total income up to \$50,000/year. Located on the 5th floor-North, this program operates on a first come, first served basis. For more information, contact ...
		1 p.m.	Harold Washington Library Center (400 S. State Street)	Film Screening/ Q&A: Jean Cocteau's Orpheus A screening of cinematic poet Jean Cocteau's Orpheus, starring Jean Marais and Maria Casares. The film will be followed by a Q&A led by Charles Coleman, Film Program Director of the Facets Cinematheque. France 1950, 95 min. In French with English subtitles.

(source EveryBlock Chicago)

b. Une évolution de la verticalité de l'information

EveryBlock illustre à la perfection les changements induits par l'apparition du web collaboratif ou web 2.0 où chacun devient créateur d'information. Le projet rend évident l'évolution de la **verticalité de l'information**. Auparavant, le journaliste avait un rôle de fournisseur d'information incontestable, de prescripteur. Désormais avec l'apparition de ce genre d'outils, la collaboration journaliste/lecteur est de mise. Un journaliste qui se rend sur un site comme EveryBlock peut y trouver une mine de sujets potentiels.



Voici un autre exemple, expliqué par Caroline Goulard, une des pionnières du journalisme de données en France : *« Le site du quotidien régional Cincinnati Enquirer a commencé en 2007 à poster des bases de données : statistiques des agressions sexuelles, plaintes pour tabagisme, gains aux casinos de l'Ohio River, prix de l'immobilier, boîtes de croquettes pour chien rappelées par leur fabricant, etc. Le jour de son ouverture au public, le Data Center du Cincinnati Enquirer a enregistré plus de 67 000 pages vues, soit deux fois plus de trafic que les galeries photos les plus populaires du site. Sur ces bases de données, le nombre moyen de pages vues par visite avoisinait 11,5 suggérant une audience impliquée et intéressée par les contenus qu'elle y trouvait. Six mois après son lancement en décembre 2006, le Data Universe du site de l'Asbury Park Press comptabilisait plus de 40 millions de pages vues. La base de données la plus populaire, concernant les salaires des employés fédéraux, a été vue 4 millions de fois en moins de trois mois. »*

4. Du consommateur d'information au conso-acteur

Bien souvent, les journalistes de données facilitent l'accès aux données qu'ils ont exploités, cela pour plusieurs raisons :

- Le public peut vérifier, constater par lui même que les journalistes n'ont pas tirés de conclusions erronées à partir des données à leur disposition.
- Le public qui explore les bases de données exploitées par les journalistes peuvent signaler de nouveaux indices, de nouvelles pistes à explorer,... Et comme le remarque Steve Doig, professeur à l'Arizona State University et pionnier du journalisme de données, *«...un lecteur impliqué qui s'intéresse à nos données est un lecteur qui revient »*.

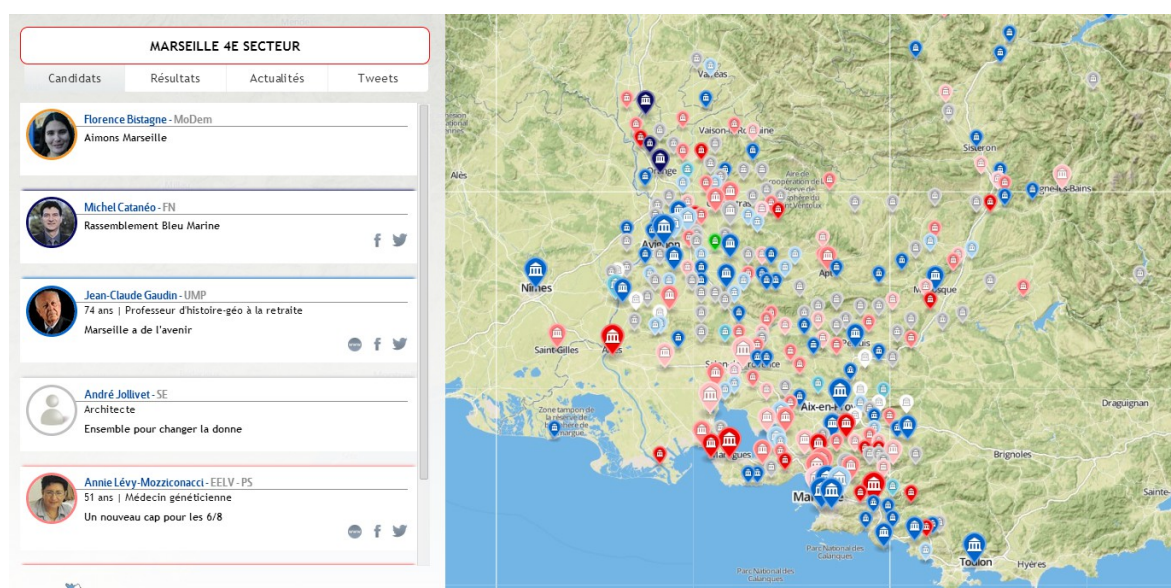
On le voit le journaliste de données par la publication des ressources qu'il a utilisé ouvre la voie à de nouveaux comportements de la part du lecteur. Celui-ci devient parfois un contributeur voir un collaborateur.

5. Un travail nouveau pour le journaliste qui suppose de nouvelles compétences

Le journalisme de données suppose de repenser la méthode d'investigation journalistique. Il ne s'agit plus ici de traquer le scoop ou la petite phrase assassine mais bien de se plonger dans un océan de données qui, au premier abord, n'a rien d'engageant. Voici un exemple basique de journalisme de données.

J'ai eu l'occasion au sein du quotidien La Provence de m'essayer au journalisme de données.

La rédaction web du journal a réalisé à l'occasion des élections municipales 2014, une carte qui fournit des renseignements par ville et par arrondissement, sur les candidats (âge, actualité du candidat, compte Twitter, Facebook, site internet,...).



Carte des Municipales, La Provence

Pour générer la carte, une journaliste de La Provence s'est attelée à un exercice de recherche et de compilation de données pour l'ensemble des 602 candidats que comportent la base. Les informations pour chaque rubrique n'étant pas forcément disponibles systématiquement, la case comporte la mention « NULL », à chaque fois que l'information n'a pas été trouvé. J'ai travaillé à partir d'une exportation des données de la base qui se présente sous la forme d'une feuille de calcul (cf ci-dessous).

fx	name																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
1	name	title	vid	nid	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	field_personnalit	
2	Aix-en-Provence	Joissains Masini	2687844	2677198	Mayrse	Avocate	1657380	Bruno Genzana	Bruno Genzana et Maryse Joissains, un leader de c	@marysejoissain	https://www.face	http://www.mary	1942-08-15T00:0				
3	Aix-en-Provence	Rouvier	2687854	2677208	Catherine	Universitaire	1746464		Catherine Rouvier ; le 20/12/2013 13	Photo Sophie S	@Rouvier2014	https://www.face	NULL	NULL	NULL	NULL	
4	Aix-en-Provence	Genzana	2687857	2677211	Bruno	Consultant	1767469	Bruno Genzana	Bruno Genzana, Bruno Genzana, Bruno Genzana	Photo serge mer	@brunogenzana	https://www.face	http://www.genz	NULL	NULL	NULL	
5	Aix-en-Provence	De Peretti	2687859	2677213	François-Xavier	Professeur de ph	49124	François-Xavier c	François-Xavier de Peretti et Brigit	© Bruno Soullard	UnnairepourAix	http://facebook.c	http://www.unmg	NULL	NULL	NULL	
6	Aix-en-Provence	Baldo	2687861	2677215	Edouard	Avocat	1631516	Hier, Édouard Bc Hier, Édouard Baldo a publieque	Photo s	mercier	NULL	https://www.face	http://www.edou	NULL	NULL	NULL	
7	Aix-en-Provence	Hamy	2687863	2677217	François	Ingénieur qualité	1703058	François Hamy (	François Hamy (EELV, et prochain	Photos Sophie S	@FrancoisHamy	https://www.face	NULL	NULL	NULL	NULL	
8	Aix-en-Provence	Mesliand	2687868	2677222	Anne	Universitaire	1746465		Anne Mesliand ; le 20/12/2013 13	Photo Pierre-Ant	aixagauche	https://www.face	http://www.aixag	NULL	NULL	NULL	
9	Aix-en-Provence	Keita	2687870	2677224	Jean-Louis	Avocat	1746466		Jean-Louis Keita ; le 20/12/2013 1	Photo Jean-Eric	NULL	https://fr-fr.faceb	NULL	NULL	NULL	NULL	
10	Aix-en-Provence	Jennane	2687872	2677226	Najia	Surveillante dans	1746467		Najia Jennane ; le 20/12/2013 13	Photo Serge ME	NULL	https://www.face	NULL	NULL	NULL	NULL	
11	Allauch	Robineau-Chailan	2719918	2709272	Monique	Chef d'entreprise	1762620		Monique Robineau-Chailan ; le 16/	Photo Patrick N	https://twitter.co	https://www.face	http://www.robin	NULL	NULL	NULL	
12	Allauch	du Crest	2719922	2709276	Claude	Maître d'oeuvre	1762621		Claude du Crest ; le 16/01/2014 11	Photo DR	@Claude_du_Cr	https://www.face	http://parti-radica	NULL	NULL	NULL	
13	Allauch	Gonzalez	2719927	2709281	José	Retraité	1779917		José Gonzalez ; le 10/02/2014 13	05	NULL	NULL	NULL	NULL	NULL	NULL	
14	Allauch	Desblancs	2719931	2709285	Lucie	Expert-comptabl	1762627		Desblancs ; le 16/01/2014 18	39	Photo DR	@allauch2014	https://www.face	http://www.desbl	NULL	NULL	
15	Allauch	Povinelli	2725034	2714388	Roland	Retraité	1762619		Roland Povinelli ; le 16/01/2014 16	Photo Cynil Solli	@rolandpovinelli	https://www.face	http://www.rolan	NULL	NULL	NULL	
16	Alleins	Grange	2687908	2677262	Philippe	Ingénieur en gén	1746475		Philippe Grange ; le 20/12/2013 14	35	NULL	NULL	NULL	NULL	NULL	NULL	
17	Alleins	Juigny	2736016	2725370	Daniel	Coordinateur indi	1770290		Daniel Juigny ; le 27/01/2014 13	57	NULL	NULL	NULL	NULL	NULL	NULL	
18	Althen-des-Paluc	Terrisse	2730109	2719463	Michel	Retraité	1767439		Michel Terrisse ; Michel Terrisse		NULL	NULL	NULL	NULL	NULL	NULL	
19	Althen-des-Paluc	Stanzione	2730111	2719465	Lucien	Retraité	1767440		Lucien Stanzioni	Lucien Stanzione	NULL	NULL	NULL	NULL	NULL	NULL	
20	Annot	Ballester	2709636	2698990	Jean	Directeur de soc	1757712		Jean Ballester ; le 09/01/2014 16	1	Photo Jessica C	NULL	NULL	NULL	NULL	NULL	
21	Annot	Rigault	2709662	2699016	Philippe	Gérant de socié	1757732		Philippe Rigault ; Philippe Rigault	Photo DR	NULL	NULL	NULL	NULL	NULL	NULL	
22	Ansois	de Sabran Ponte	2716940	2706294	Géraud	Retraité	1761186		Géraud de Sabran Ponteves Ansois ; le 14/01/201	1	NULL	NULL	NULL	NULL	NULL	NULL	
23	Ansois	de Sabran Ponte	2717007	2706361	Géraud	Retraité	1761219		Géraud de Sabran Ponteves ; le 14/01/2014 18	04	NULL	NULL	NULL	NULL	NULL	NULL	
24	Apt	Curel	2715341	2704695	Olivier	Agriculteur	1760417		Olivier Curel ; le 13/01/2014 17	57	NULL	NULL	NULL	NULL	NULL	NULL	
25	Apt	Santoni	2715344	2704698	Dominique	Chef d'entreprise	1760419		dominique santoni ; le 13/01/2014 18	02	@aptvillevante	http://facebook.c	http://www.domi	NULL	NULL	NULL	
26	Apt	Tantini	2733057	2722411	Valère	Retraité	1768831	Valère Tantini a	Valère Tantini a	Valère Tantini a	Photo L PP	NULL	NULL	NULL	NULL	NULL	
27	Aries	Schiavetti	2711042	2700396	Hené	Ancien agent ter	1745602	Cynil Juglaret (à	Cynil Juglaret (à droite) a demand	Photo V F	NULL	NULL	NULL	NULL	NULL	NULL	
28	Aries	Juglaret	2711043	2700397	Cynil	Cadre territorial	1745609	Cynil Juglaret (à	Cynil Juglaret (à droite) a demand	Photos V F	NULL	NULL	NULL	NULL	NULL	NULL	

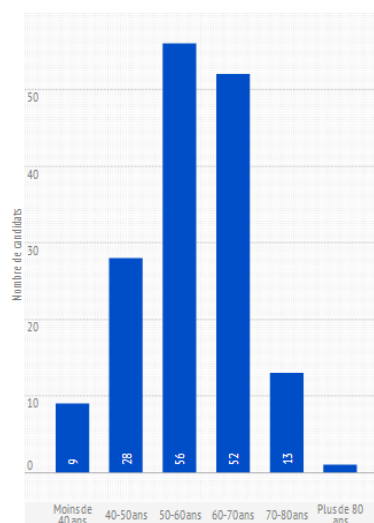
L'idée de départ était de croiser les données relatives à l'âge et à la présence web des candidats et d'en tirer des conclusions.

Il a d'abord fallu « nettoyer » les colonnes pour ne garder que les cases qui contiennent des informations exploitables c'est à dire toutes les cases qui ne portent pas la mention « NULL ». J'ai effectué cette opération pour les colonnes Âge, Twitter, Facebook et site internet.

Une fois cette tâche effectuée, j'ai utilisé un outil de représentation visuelle « Infogr.am » qui permet facilement et rapidement de réaliser des infographies de qualité (cf. le résultat ci-dessous).

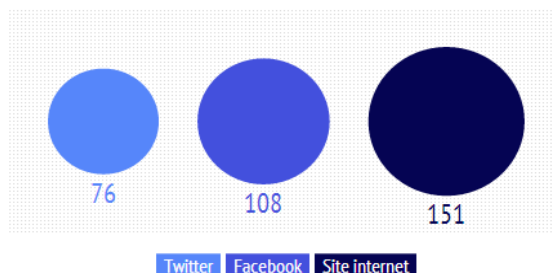
J'ai ensuite contacté Pierre Guillou, spécialiste de la communication politique numérique et fondateur du site élus 2.0, pour qu'il commente mes pistes. A partir de ses propos et des graphiques tirés de la base de données, j'ai pu construire mon article.

Répartition de la présence web des candidats par tranche d'âge



Etude réalisée, au 11 février 2014, sur 159 candidats pour lesquels nous disposions de données (âge et présence web). Est considéré comme présent sur le web tout candidat disposant au moins d'un compte Twitter ou Facebook ou d'un site internet.

Une présence timide sur le web pour les candidats aux municipales



Au 11 février 2014, notre base de données recensait 602 candidats dont 183 étaient présents sur internet à notre connaissance. Ils disposaient d'un compte Twitter et/ou d'un compte Facebook et/ou d'un site internet.

6. De nouvelles données

a. Le mouvement Open Data

La vague Open Data a été lancée en 2009 par les États-Unis qui sont les premiers à libérer leurs données via le portail data.gov. Le Royaume-Uni suit, en janvier 2011, avec Data.gov.uk puis la France quelques mois plus tard, mars 2011, avec data.gouv.fr

Pour les collectivités et les organismes publics, l'Open Data consiste à publier sur des plateformes ouvertes des informations statistiques, cartographiques, des horaires, des données économiques et financières sur les territoires.

La mise à disposition des données publiques est d'ailleurs une obligation légale. Un cadre juridique définit les informations qui peuvent être rendues publiques. Ainsi, les données sensibles et à caractère personnel sont exclues. L'Open Data permet de constituer une source d'information fiable pour les chercheurs, les journalistes, les étudiants, les citoyens...

Ce mouvement d'ouverture est représentatif d'un monde toujours plus connecté. Henri Verdier, Directeur d'Etalab, mission chargée de l'ouverture des données publiques, explique: *« Le mouvement Open Data est un mouvement de fond à la croisée de plusieurs histoires: la transparence, l'implication de la société civile dans l'action publique, les nouvelles stratégies d'innovation après la révolution numérique ».*

b. Du chemin reste à faire

En France, par exemple, l'open data est encore en phase de mise en place, d'amélioration. Au mois de janvier 2014, on estime à 350 000 le nombre de fichiers disponibles sur data.gouv.fr. Une proportion encore faible. Pour que la libération des données porte toutes ses promesses, la masse de données disponibles doit être critique.

Pour y parvenir, il faut faciliter les démarches de diffusion des administrations et permettre au public d'atteindre facilement ce qu'il cherche.

Ces deux points sont primordiaux car ils sont le carburant du journalisme de données.

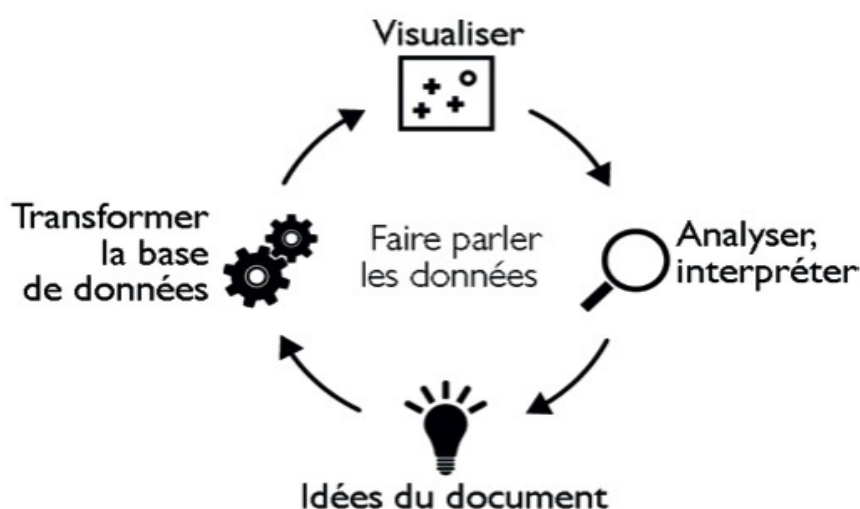
Outre ce frein technique à la diffusion, un autre facteur entre en ligne de compte. La réticence des administrations françaises traditionnellement habituées à une culture du secret autour de leurs données. Trop de données sont également enregistrées et conservées dans des formats inexploitable (cf deuxième partie). Ces différentes contraintes ont provoqué l'apparition des « people person ». Des spécialistes de la récupération de données qui ont pour tâche principale de harceler les organismes, entreprises, administrations aux données difficiles d'accès.

7. Une nouvelle visualisation

a. Visualiser pour mieux comprendre

Dans certaines situations, les données sont plus explicites que les mots ou les photos. De plus en plus d'outils comme Google Fusion Tables, DataWrapper, ManyEyes ou Infogr.am très facilement utilisables permettent à n'importe qui de générer des représentations visuelles qualitatives.

Les données numériques sont composées de bits et d'octets illisibles à l'état brut. Pour rendre ces données compréhensibles, il faut les visualiser.

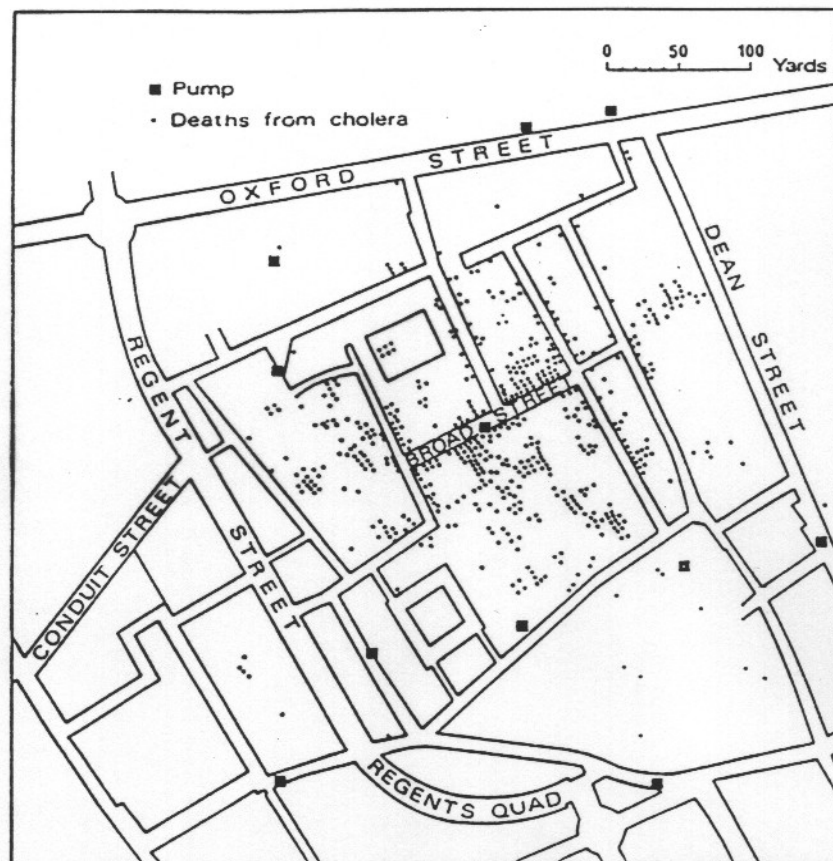


Faire parler les données, Gregor Aisch

La représentation visuelle appliquée à l'information ne constitue pas une mode passagère, loin de là. Au contraire, ce sont les graphiques plats et peu compréhensibles que l'on retrouve à foison dans la presse écrite qui semblent menacés par les possibilités offertes par des outils de visualisation toujours plus nombreux.

A l'époque actuelle du Big Data roi, les visualisations prennent toujours plus d'importance car elles sont la meilleure solution pour traduire de manière compréhensible la complexité des données. Autrement dit, « plus l'ensemble des données est important et articulé, plus la synthèse par l'image sera efficace... »

Deux exemples pour illustrer ce propos. Le premier est la fameuse carte du Docteur John Snow qui représente l'épidémie de choléra de 1854 à Londres.

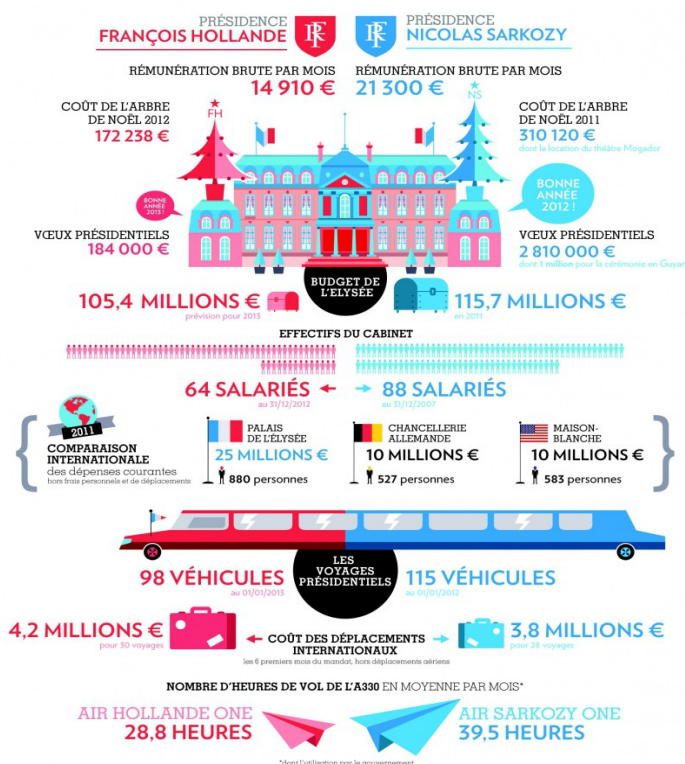


Vignette Figure 1-1. Snow's map of cholera. The affected well is clearly identified by the concentration of cases in its vicinity. Reprinted from Howe, G. M. (1972). *Man, environment, and disease in Britain*. New York: Barnes and Noble Books, p. 178. Copyright 1972. Reproduced by Permission. Original source: Snow, J. (1855). *On the mode of communication of cholera*. London, 1855.

Cette représentation de presque 200 ans demeure d'une efficacité redoutable. La théorie de base du docteur Snow était de dire que la probabilité de mourir du choléra dépendait de la distance entre le lieu de vie des victimes et une source d'eau contaminée par la bactérie. Il apparaît très clairement sur le schéma que plus on se rapproche de Broad Street, plus le nombre de décès augmente. La source d'eau contaminée devait donc se trouver à cet endroit.

Il faut également prendre en compte le fait que le cerveau humain dispose de nombreux capteurs destinés à la perception visuelle.

Paris-Match du 30
mai au 5 juin 2013



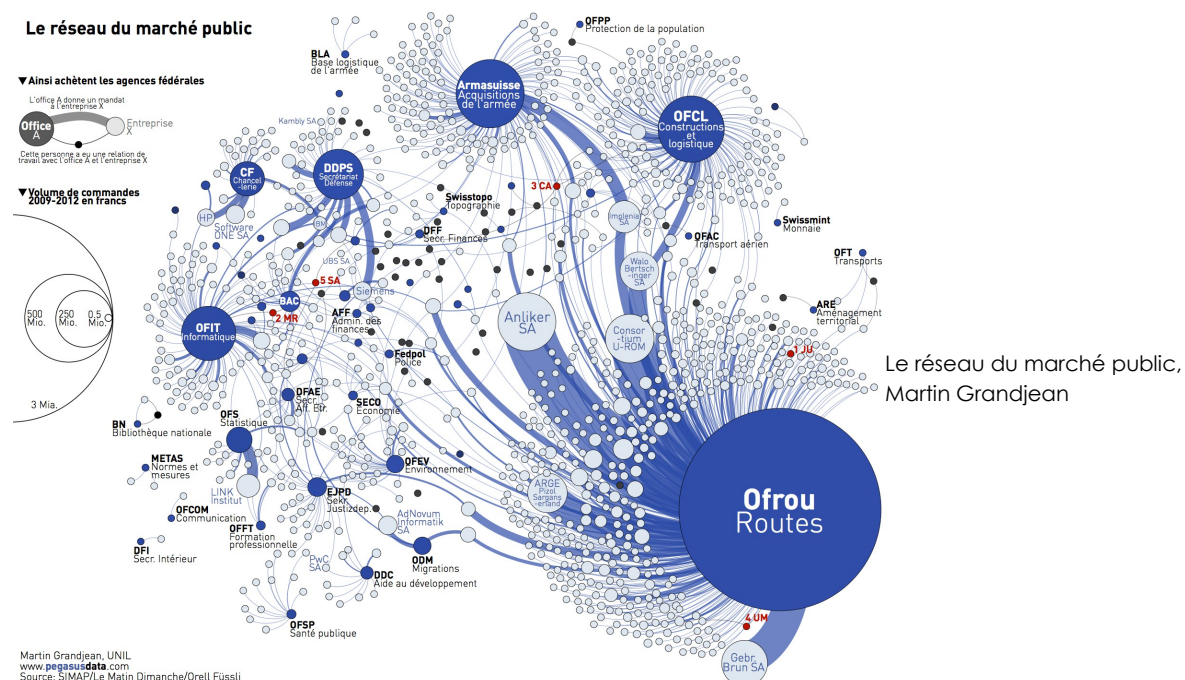
Cette visualisation de données tirée de l'hebdomadaire Paris Match compare le fonctionnement de l'Élysée sous la présidence Hollande à la présidence Sarkozy. Il s'agit d'un bon exemple puisque les avantages de la représentation graphique sont évidents.

La couleur tout d'abord permet de faire la distinction entre les dépenses de M. Hollande et M. Sarkozy. Puis, c'est la variation de la taille des objets qui permet de bien se représenter la différence.

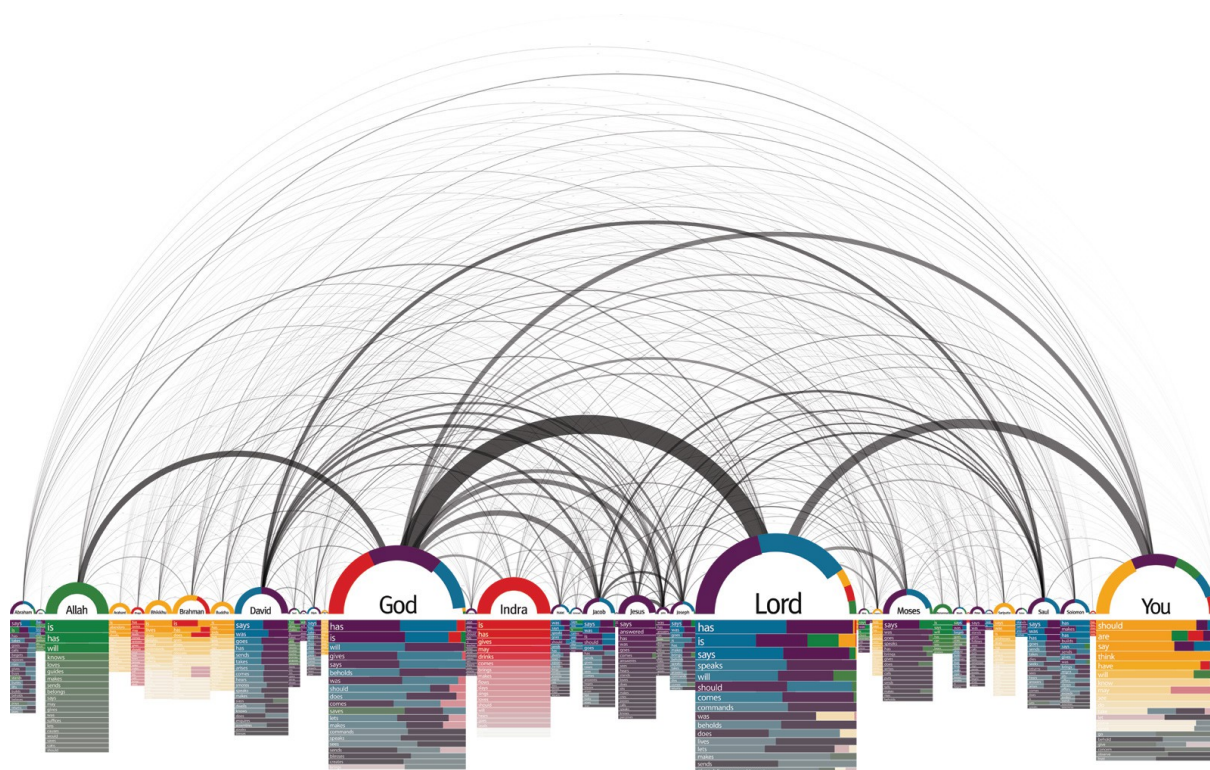
b. La visualisation ne doit pas être systématique

Néanmoins, des défauts parasitent encore la représentation de données. Il arrive que certaines data visualisations privilégient l'esthétique à la compréhension. Une complexité gratuite qui dessert le sujet traité. Les visualisations de données bénéficient d'un engouement qui ne se dément pas. Cela peut constituer un problème puisqu'on a tendance à recourir à la visualisation de données pour tout et n'importe quoi y compris lorsque cela ne s'impose pas. Une sorte de faisons une infographie quelque soit le sujet. Il est ainsi très facile de tomber sur des visualisations boursouflées et totalement incompréhensibles. Un contre sens complet puisque les visualisations sont là pour transformer une densité de données complexes en une représentation immédiatement compréhensible : la bien nommée simplicité visuelle.

Voici quelques mauvais exemples :



La visualisation ci-dessus représente la structure du marché public suisse. Elle bénéficie d'un joli traitement visuel mais est extrêmement fouillée et peu compréhensible au premier regard alors que c'est justement sa finalité.



Visualisation sur l'occurrence des mots dans les différentes religions.

Même reproche que pour la visualisation précédente. Ici, la représentation doit faire ressortir la récurrence de certains mots clés dans différentes religions. En l'occurrence, l'enchevêtrement de demi-cercles et la quantité de couleurs rendent peu évidente la compréhension.

Toutes les transformations abordées dans cette partie supposent de nouvelles pratiques et de nouvelles connaissances pour en tirer profit. C'est donc tout naturellement que le journalisme a évolué. On parle bien d'une évolution car le journalisme de données n'est pas une révolution. Il s'agit d'une adaptation des pratiques journalistiques à la société d'aujourd'hui. Alain Joannès évoque « *une manière de chercher, de trouver et de montrer de l'information avec les outils et pour les audiences de notre époque* ».

II. OÙ EN EST-ON AUJOURD'HUI ET QUELS SONT LES FREINS ?

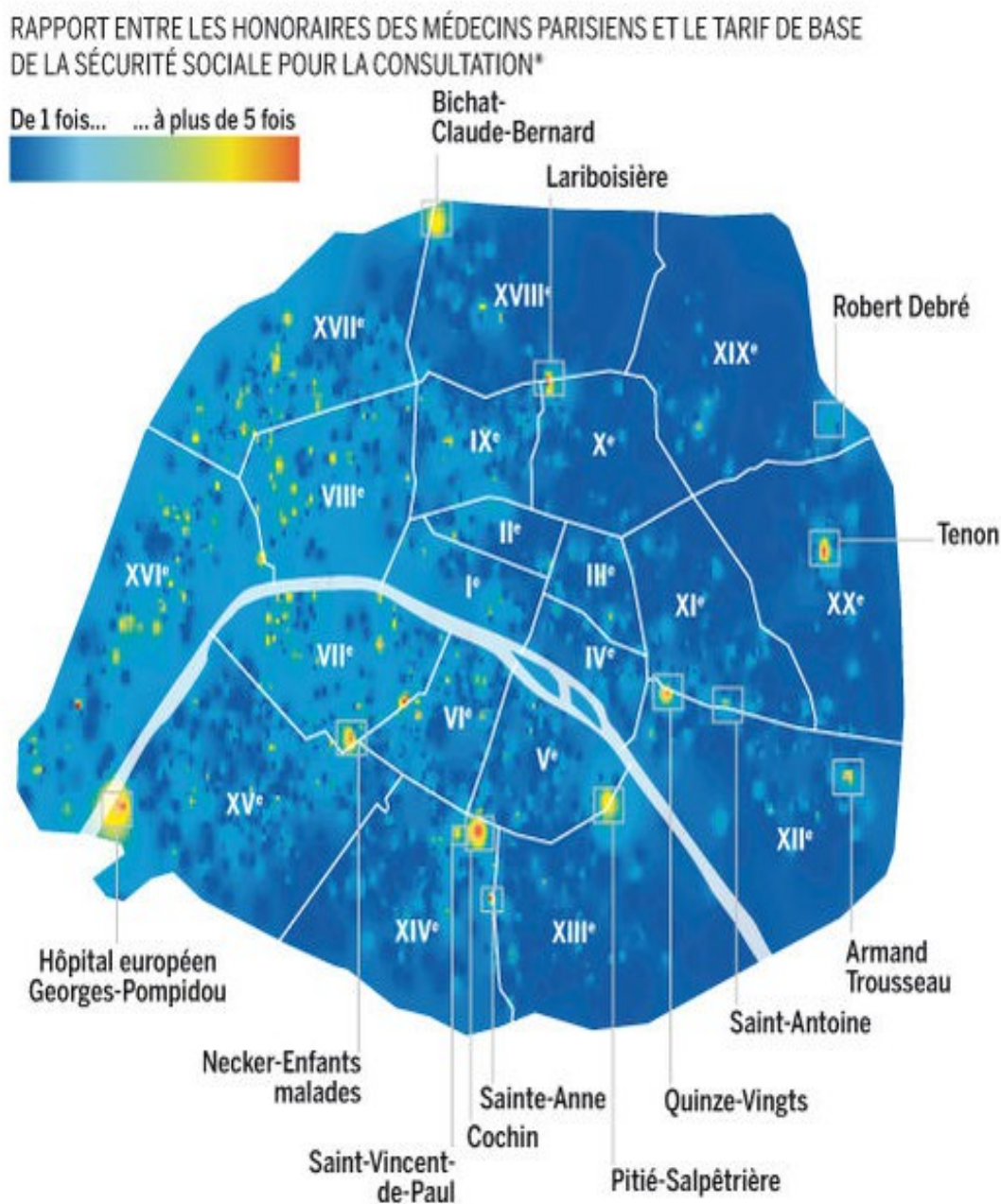
Cette pratique neuve suppose de nouvelles compétences et il s'agit là d'un des freins majeurs au développement du journalisme de données. Un temps d'adaptation semble nécessaire avant qu'il puisse espérer s'imposer dans les rédactions du monde entier. La pratique doit encore faire ses preuves, se faire connaître du grand public et c'est la multiplication de « coups » marquants qui permettront de la pérenniser.

A l'heure actuelle, les rédactions n'ont pas encore effectuée une mue numérique complète et les incursions dans la sphère du journalisme de données sont encore timides particulièrement en France. Néanmoins, le journalisme de données se développe de plus en plus dans de grands médias hexagonaux.

Le Monde s'est fait remarquer en 2012 en publiant une enquête sur les dépassements d'honoraires médicaux. Le quotidien s'est appuyé sur les données publiques d'Ameli-direct, le site de l'assurance-maladie, pour mesurer le tarif demandé en moyenne aux patients, médecins généralistes et spécialistes.

L'enquête a permis de montrer que Paris est la ville où les dépassements d'honoraires sont les plus élevés en moyenne et les montants très supérieurs au tarif de base de la Sécurité sociale (Cf. première figure ci-dessous). La seconde visualisation de données (cf. seconde figure ci-dessous) permet de mettre en perspective les honoraires parisiens en les comparant avec ceux pratiqués dans les plus grandes villes de France. En s'appuyant sur ces graphiques, les journalistes du Monde ont produit une analyse chiffrée, poussée et détaillée.

Tarifs médicaux : enquête sur les dépassements d'honoraires, Le Monde, avril 2012



Le Monde, Tarifs médicaux : enquête sur les dépassements d'honoraires

A Paris, des dépassements plus fréquents et plus élevés

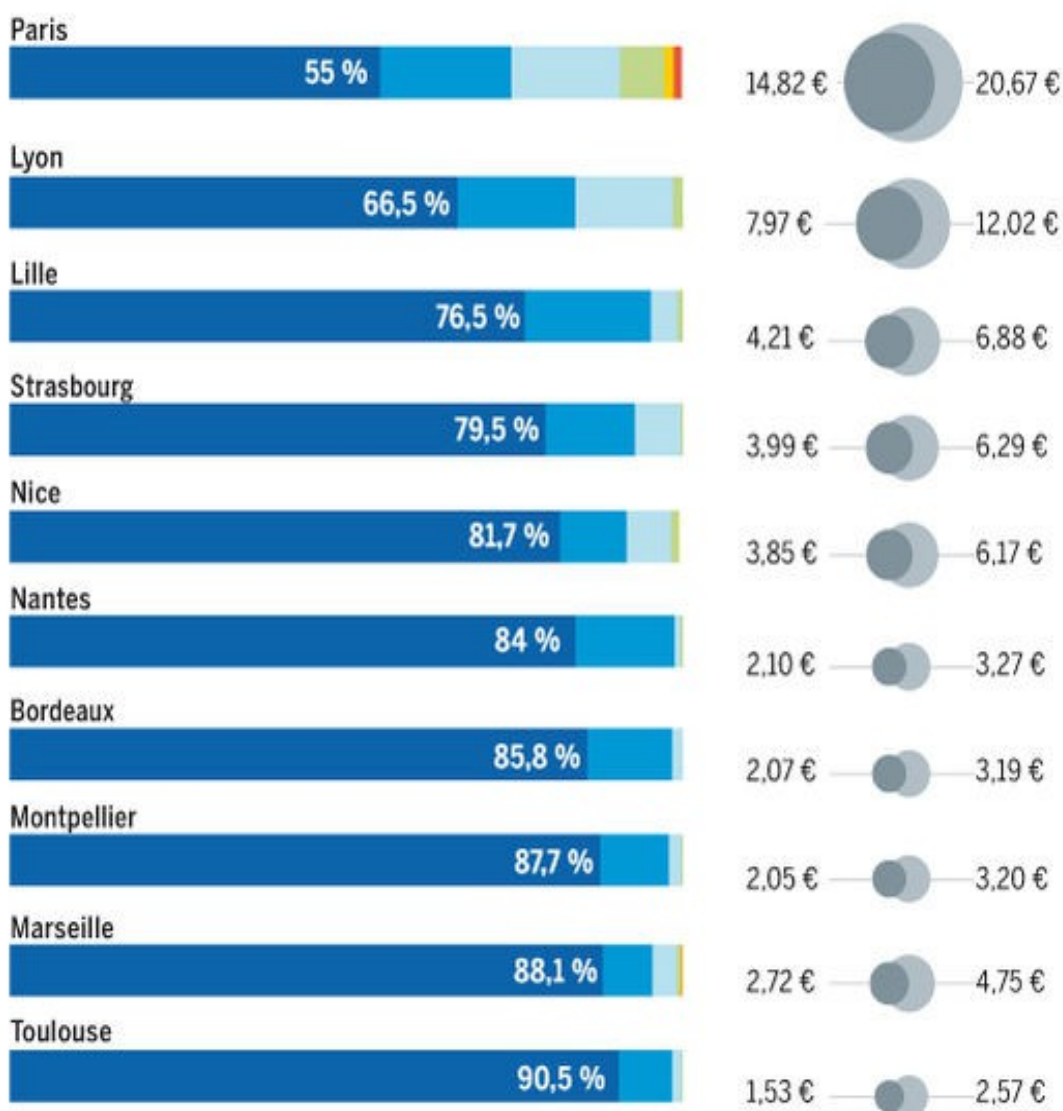
RÉPARTITION DES MÉDECINS SELON LEURS HONORAIRES DE CONSULTATION COMPARÉS AU TARIF DE BASE DE LA SÉCURITÉ SOCIALE* EN 2011, en %

Généralement au tarif

1 à 2 fois 2 à 3 fois 3 à 4 fois 4 à 5 fois + de 5 fois

DÉPASSEMENT MOYEN DU TARIF DE BASE DE LA CONSULTATION

● Ensemble des médecins (généralistes et spécialistes)
● Médecins spécialistes



* Selon les consultations, la base de remboursement de la Sécurité sociale s'élève de 23 € (consultation simple) à 69 € (avis de consultant)

SOURCES : DONNÉES AMELI-DIRECT.FR, LE MONDE

Les rédactions françaises préfèrent encore sous-traitées leurs productions à des agences indépendantes telles que WeDoData, Dataveyes ou Journalism++.

L'existence de ces pure players spécialistes de la visualisation de données est révélatrice.

Elle montre d'une part l'envie des médias de se mettre au journalisme de données. Mais révèle aussi le manque de moyens, de compétences et de savoir-faire en interne des médias français.

Les pays anglo-saxons bien que plus avancés se heurtent eux aussi à des freins.

1. Le journalisme de données : une pratique onéreuse

Les journalistes de données ont des compétences rares donc chères. Le journaliste de données possède un profil hybride qui n'existe presque pas à l'heure actuelle dans les rédactions. De ce fait, ces dernières sont contraintes d'employer au moins deux personnes :

- un journaliste porté sur les chiffres et familier de l'exploration de données.
- un développeur (ou une agence) pour mettre en forme, proposer une visualisation dynamique et interactive.

En plus de mobiliser une équipe qualifiée, les projets de journalisme de données nécessitent du temps et se déroulent souvent sur plusieurs semaines. Une autre source de coûts élevés.

a. La tentation du crowdfunding

L'organisme, à but non lucratif (la mention a son importance), Pro Publica, fondé en 2008, se compose d'une rédaction d'une trentaine de journalistes. Il réalise des enquêtes financées par crowdfunding (appels à financement auprès du grand public) et par mécénat.

Parmi les nombreuses réalisations de Pro Publica, on trouve bon nombre de projets de journalisme de données. L'un des plus connus *Dollars for Doctors. How industry money reaches physicians* explique comment les labos paient les docteurs pour promouvoir leurs médicaments.

Le travail d'agrégation de données a été réalisé par Pro Publica qui les a ensuite mises à disposition gratuitement de l'ensemble des rédactions américaines. Chaque média a ainsi pu bénéficier de données sur sa région pour écrire des histoires.

Ces données librement accessibles ne signifient pas qu'elles ont été produites sans frais. Les journalistes de Pro Publica ont récolté les fonds pour réaliser ce travail de compilation et de mise à disposition gratuite des données. Une autre enquête de Pro Publica sur les hôpitaux de la Nouvelle-Orléans récompensée par le prix Pulitzer en 2010 a nécessité plusieurs mois de travail et pas moins de 300 000 €. Une somme que serait incapable de sortir n'importe quel média traditionnel pour un seul projet.

b. L'effet d'expérience

Le journalisme de données a donc un coût important mais comme l'explique Simon Rogers, responsable du service data journalism au Guardian : « *les coûts de production se réduisent du fait de l'expérience acquise et de la réutilisation et/ou réactualisation de bases de données* ». Apprendre cette forme de journalisme, former et accumuler des bases de données prend du temps et a un coût. Une fois cette étape franchie et à condition que les bases de données soient à jour, le retour sur investissement est très intéressant car les ressources créées sont réutilisables.

En prenant ses distances avec les contenus écrits classiques facilement duplicables, le journalisme de données crée un contenu à forte valeur ajoutée.

On peut penser qu'il est plus facile de monétiser ces contenus différenciant, nouveaux et originaux pour le lecteur qui sera sans doute plus apte à délier les cordons de sa bourse.

2. Un problème de formation et de compétences techniques

Le journalisme de données exige une véritable culture informatique et des chiffres. Des compétences neuves qui ne sont pas l'apanage de la majorité des journalistes qui ont davantage une culture littéraire. Les journalistes ne sont pas familiers des bases de données, de leur nettoyage et interprétation, ni des nouveaux outils (Storify, Prezi, ScrollKit...) dans leur ensemble.

Un journaliste de données sait exploiter une base de données, jongler avec les chiffres, coder pour développer l'application nécessaire à la restitution de son enquête. Ceci suppose des connaissances techniques spécifiques dont ne sont pas dotés la plupart des journalistes actuels.

Le contexte économique actuel ne permet pas non plus d'avancée rapide dans ce sens comme l'explique Estelle Prusker-Deneuville, doctorante à l'université Panthéon-Assas : *« Former plusieurs générations de journalistes représente un lourd investissement pour des groupes de Presse déjà fragilisés par la baisse des ventes papier, le recul des recettes publicitaires et la mise en place de stratégies digitales -web, mobile, tablette- coûteuses. »*

a. Une question de génération ?

Les rédactions en place ne sont pas formées au journalisme de données, c'est un fait. La transition se fera sans doute avec l'arrivée des nouvelles générations de journalistes, les « digital natives », pour lesquelles l'informatique et internet n'ont pas de secret. Pourtant, là encore, les écoles de journalisme ont pris du retard et commencent seulement à intégrer des enseignements spécifiques. Un instantané de la réalité actuelle sans trop la caricaturer ressemblerait à ça : la plupart des étudiants journalistes ne savent pas coder et ont des compétences limitées en mathématiques, tandis que la majorité des étudiants en informatique n'ont pas l'art de la narration indispensable pour ce métier. Et c'est justement à cette intersection que se situe le journaliste de données.

b. Les écoles ne sont pas les seules à blâmer.

Nicolas Kayser-bril, pionnier du journalisme de données en France et fondateur de Journalism++, déplore le manque de vision des médias français : *« Les compétences sont difficiles à trouver. Mais le problème des médias français reste qu'aucun d'entre eux n'a de stratégie sur le sujet. Plusieurs ont embauché des compétences intéressantes (Le Monde avec Alexandre Léchenet, Le Temps avec Jean Abbiateci). Mais aucun n'a constitué d'équipes, avec des processus clairs, des budgets propres... Ce qui permet de ne pas se baser uniquement sur des individus et de gagner en productivité avec l'expérience. De telles équipes existent chez Zeit online, The Times (Londres), NPR, LA Times, Financial Times et d'autres. »*

Le lent essor du journalisme de données paie la défiance originelle suscitée par internet. Les journaux n'ont pas suffisamment anticipé l'ampleur qu'allait prendre les nouvelles technologies. Ce postulat se retrouve aujourd'hui dans les rédactions avec des journalistes obnubilés par le papier (Une de Libération, édition week-end des 8 et 9 février, « Nous sommes un journal ») et de l'autre des informaticiens qui ne souhaitent pas se frotter au journalisme.

3. Open data: l'enjeu de la libération des données

Journalisme de données et Open Data sont étroitement liés. L'expression Open Data est formée à partir des mots « Open » qui signifie « ouvert » et « Data » qui renvoie à l'idée d'une donnée numérique d'origine privée ou publique. Ainsi, l'Open Data est le mouvement d'ouverture des données qui consiste à les mettre à disposition de tous facilement et gratuitement.

a. Des données libérées ?

On parle aussi de données libérées c'est à dire de données auxquelles on a enlevé tout obstacle qui aurait pu empêcher la circulation. Ces obstacles peuvent être de différentes natures :

- économiques, la gratuité d'accès à ces données est une condition indispensable.
- juridiques, les données doivent pouvoir être utilisées sans porter préjudice à quiconque.
- techniques, les données ne doivent pas être protégées par des mots de passe, par des DRM ou autres verrous techniques qui empêcheraient de les réutiliser pour un usage personnel.

La publication de données publiques est une obligation légale mais certaines administrations refusent par méconnaissance de la loi de diffuser leurs informations alors même que la loi du 17 juillet 1978 les y contraints. Celle-ci reconnaît *« à toute personne le droit d'obtenir communication des documents détenus dans le cadre de sa mission de service public par une administration, quels que soient leur forme ou leur support. »*

b. Libérer les données, c'est bien. Les rendre exploitables, c'est mieux

On touche ici au(x) format(s) des données. Un écueil majeur puisque de nombreux jeux de données mis en ligne sont tout simplement inexploitables en raison de leur format.

Avant de mettre les données en ligne, il faut les convertir dans un format récupérable. Trop de données sont perdues au milieu de fichiers PDF, dispersées dans des masses de texte, scannées en format image,... Ainsi, lorsque l'on souhaite exploiter, croiser des données, on se retrouve bien vite face à des formats différents et incompatibles entre eux. A ce titre, l'expérience de Nicolas Kayser-Bril est révélatrice : *« Je travaille en ce moment sur un projet agrégeant des données départementales. Ce qui implique d'appeler chaque préfecture. A quelques exceptions près, les fonctionnaires transmettent les fichiers demandés. La qualité des documents, en revanche, laisse à désirer. 20% des fichiers sont au format Excel. 10% au format Word. Le reste, du PDF. En enlevant les tableurs faits avec les pieds et en rajoutant les PDF convertis depuis Excel, environ 25% des données sont structurées, c'est-à-dire lisibles par un ordinateur ».*

Contacté sur ce même sujet, Nicolas Kayser-Bril précise *« Les portails open data présentent encore souvent des données dans des formats qui ne sont pas ouverts, ce qui fait que ce n'est plus de l'open data ».*

On le voit, la libération des données doit être accompagnée par une harmonisation des formats pour les rendre exploitables.

La libération des données concernent aussi les données privées d'entreprises par exemple. Or, comme l'explique Nicolas Kayser-Bril : *« les données de nombreuses entreprises ont une valeur stratégique, elles ne peuvent pas les communiquer. C'est au journaliste de trouver un moyen de collecter des informations équivalentes. »* La réticence des entreprises à transmettre certaines données est compréhensible. Elles doivent protéger leurs actifs, tout ne peut se retrouver en ligne et librement consultable.

4. Redonner de la valeur au travail journalistique.

a. Mettre un terme au gaspillage

La numérisation a permis de mettre au jour le gâchis réalisé avec les données journalistiques. Les données produites chaque jour sont utilisées une ou deux fois avant d'être oubliées: *« ...Chaque service mouline des chiffres qu'il a vérifié, ou peut traduire certains faits en données qui seraient précieuses si elles étaient centralisées et réutilisables à l'envi. Combien d'articles sur la crise du logement où les chiffres restent à plat sur un carnet de notes au lieu d'aller rejoindre une base de données ? »* explique Sophian Fanen, journaliste à Libération.

L'apport du journalisme de données est de ne pas figer les choses mais au contraire de les faire vivre.

Le témoignage de Frédéric Waringuez, rédacteur en chef adjoint au journal L'Equipe, est particulièrement éclairant. Il évoque la déperdition d'informations de son quotidien : *« Depuis la naissance de notre site internet en 2000, on rentre dans une base de données centralisée des palmarès, des stats sur des rencontres dans tous les sports... Mais ces données sont peu ou pas exploitées. »* Une nouvelle fois, c'est le manque de compétences en interne qui est pointé du doigt : *« On manque de compétences techniques, de développeurs et de statisticiens pour faire de cette matière quelque chose de sexy pour le lecteur. Aujourd'hui, on doit donc travailler avec des*

boîtes extérieures, qui nous fournissent des stats de football, alors qu'on en produit nous-mêmes énormément. Mais on n'est pas capable de les mettre en forme. La première étape était de réaliser la valeur de ce gisement, ce qui est fait. »

b. Les « Mediata Startups »

Actuellement, aucun média ou section consacré au journalisme de données n'entend être profitable mais de nouvelles possibilités s'ouvrent avec lui. Le journalisme de données est un très bon moyen d'ajouter de la valeur à l'information, que ce soit en collectant des informations inexistantes, en les analysant ou en les mettant en forme.

On touche ici au concept de « Mediata Startups ». Historiquement, les médias et les entreprises de données étaient séparés.

Le concept de « Mediata Startups » est de vendre des informations B2B, soit des données exclusives, soit des analyses basées sur des données. Les articles ne sont là que dans une fonction commerciale, pour attirer des clients vers les services payants et asseoir la marque. Bloomberg est le meilleur exemple de ce mode de fonctionnement, même si ce n'est pas une startup.

On s'oriente ici vers un modèle de pure player hybride. D'un côté, une vocation information et de l'autre, une partie entreprise B2B dont est tirée le gros des revenus. Ce concept propose aux entreprises de presse de rompre avec le modèle économique publicitaire traditionnel.

Rompre avec le modèle publicitaire, c'est aussi tourner le dos à la tyrannie du nombre de pages vues qui pousse souvent à un nivellement vers le bas qualitatif.

Voilà l'une des (r)évolutions que peut provoquer le journalisme de données. Sortir du modèle centré sur l'article pour aller vers un modèle où la valeur tient dans des bases de données exploitées de diverses manières, l'une d'elles étant la rédaction d'articles.

5. Les atouts séduction du journalisme de données

La pérennisation du journalisme de données passe évidemment par la réception qu'en fait le public. Voici, synthétisés les principaux atouts de la discipline pour les lecteurs :

- Il permet de mieux traiter les informations qui s'appuient sur d'importants volumes de données. Les grands nombres parlent peu à l'esprit du public. Passé un certain stade de grandeur, le cerveau peine à donner une réalité mentale claire. Des représentations graphiques, interactives qui plus est, sont beaucoup plus parlantes. Si l'on prend l'exemple de la dette publique française, il est plus évocateur de visualiser son évolution graphiquement que de dire qu'elle a doublé par rapport à il y a 10 ans.
- Une image vaut mille mots. Les représentations visuelles de données permettent d'ôter à un texte une accumulation de termes compliqués ou d'informations complexes et de les transposer en représentations facilement compréhensibles. Autrement dit, un raisonnement complexe peut parfois être traduit en une représentation visuelle simple.
- Contrairement à un article classique, le journalisme de données propose des contenus interactifs. Le lecteur peut ainsi jouer avec les données, les manipuler, se les approprier. Une posture active qui fait la différence comme l'explique Caroline Goulard: *« Plongé dans une posture active, l'utilisateur appréhende et mémorise plus facilement l'information : il peut rechercher dans une base, zoomer sur une carte, personnaliser une infographie, paramétrer les données à comparer, apporter ses propres données, etc. »*

III. CONCLUSION

Le journalisme de données part d'un postulat : l'abondance de données constitue un matériau riche source d'un nombre de sujets potentiels important. A l'heure d'internet et de la rapidité, l'information se périme très rapidement, il faut être le premier à la sortir. Un lecteur qui a déjà pris connaissance d'une information a peu de chance d'être intéressé pour la lire à nouveau à un autre endroit même si l'article est plus complet. Dans cette course de vitesse, le premier perdant est le rédacteur qui a moins de temps pour croiser ses sources, pour vérifier l'information.

Les sites proposent souvent des informations identiques. Cela n'est pas un problème pour la presse écrite puisque le lecteur achète le plus souvent un seul journal. Sur internet le problème est différent car tout est accessible au même endroit, il suffit de taper ses termes de recherche pour obtenir une multitude de résultats. Ainsi une information trouvée sur un site peut être trouvée ailleurs. On a donc une perte de la valeur intrinsèque de l'information. Les lecteurs sont moins enclins à payer pour de l'information basique, ils recherchent de la viande autour de l'os informationnel.

Les lecteurs tout comme les annonceurs cherchent la rareté. Or, celle-ci ne se trouve plus dans l'accès à l'audience (ce pourquoi les annonceurs allaient vers les médias), ni comme on vient de le voir dans l'accès à l'information (ce pourquoi les consommateurs se tournaient vers les médias).

Désormais, la rareté réside dans la confiance accordée à l'information. Par ailleurs, à l'heure numérique, énormément de contenus sont numérisés ou en passe de l'être.

Cette abondance de données constitue le terreau idéal pour de nouvelles pratiques.

Le journalisme de données s'inscrit dans ces multiples changements. Il s'agit d'une forme de journalisme plus technique au sens des compétences nécessaires pour le pratiquer. Il suppose un profil hybride pour être capable de transposer visuellement les données mais aussi des compétences

journalistiques traditionnelles teintées d'un attrait pour les chiffres. Presque un profil scientifique permettant de comprendre, d'exploiter et de tirer au mieux profit des bases de données.

La constitution et la disponibilité de données regroupées en bases est un autre facteur important. A ce titre, le mouvement de libération des données, l'Open Data, est primordial.

Il est toujours possible de se fournir en données autrement mais il est certain que le développement de l'Open Data multiplie les pistes d'exploration possibles du journaliste de données.

En France, le développement du journalisme de données paie le rythme peu soutenu du mouvement d'ouverture des données. Quoi de plus logique serait on tenté de dire lorsqu'on demande à l'État et à ses représentants en région, habitués à une culture du secret, d'ouvrir massivement leur banque de données. Une vraie remise en question s'impose, les structures concernées doivent revoir leur principe de fonctionnement, leur rapport à l'information ainsi que leur stratégie de communication...

A tel point qu'une activité à part entière est apparue : les « people person ». Ces individus sont chargés de « harceler » les autorités pour récupérer des données exploitables. La notion de données exploitables est importante puisqu'il est encore très commun d'obtenir des données qui ne permettent pas d'être traitées automatiquement par ordinateur (fichiers scannés, photos, données noyées dans du texte...).

Le journalisme de données ne révolutionne pas le journalisme en soi, il s'agit simplement d'une adaptation de la pratique au monde actuel. Il tire profit des possibilités offertes par les nouvelles technologies. Les évolutions qu'il entraîne ne sont pas anodines, toute la chaîne de l'information s'en trouve secouée. Les rédactions doivent acquérir de nouveaux profils : développeur, graphiste, webdesigner, statisticien... Et dans l'idéal un journaliste de données touche à tout en mesure d'occuper ces différents postes. Des profils qui font cruellement défaut à l'heure actuelle.

Côté lecteur, de nouvelles possibilités s'ouvrent. Parmi les principales figure l'émergence croissante de l'information micro-locale.

On atteint un niveau de granularité de l'information tel que les informations obtenues sont de l'ordre de ce qui se passe au coin de la rue, littéralement. L'un des exemples les plus fameux est Every Block d'Adrian Holovaty. Il s'agit d'une sorte de Facebook au regard de sa dimension sociale et participative (l'aspect Big Brother et fichage du profil des individus en moins) entièrement porté sur l'information de proximité. L'emblème même du web collaboratif.

Le lecteur au même titre que le journaliste est acteur de l'information.

Cette dimension participative est prépondérante dans le journalisme de données puisque le lecteur peut-être mis à contribution lorsque le journaliste se retrouve face à de grandes quantités de données à analyser, sans possibilité de le faire automatiquement par ordinateur. L'un des plus illustres exemples est celui des notes de frais des parlementaires britanniques. Dans cette affaire, les lecteurs du journal The Guardian ont épluché les dizaines de milliers de notes de frais récupérées par le journal puis ont rapportés les anomalies décelées dans une interface dévouée développée spécialement par le Guardian. Les différents abus révélés par cette enquête ont fait grand bruit outre-manche.

Le lecteur devient acteur de l'information pour une autre raison, il interagit avec le contenu final proposé par le journaliste. Il peut consulter la base de données utilisée par le journaliste afin d'y trouver des informations spécifiques à ses centres d'intérêts. Il peut jouer avec les paramètres de la représentation visuelle interactive. Les possibilités d'interaction sont nombreuses. Cette posture active du lecteur est un atout non négligeable car l'information se trouve mieux assimilée et consommée de manière presque ludique. Elle est qualitative et apporte une vraie valeur ajoutée par rapport aux formes de restitution habituelle.

Les bases de données générées par les journalistes ont un autre atout, celui de la permanence. En effet, si elles sont mises à jour régulièrement, elles peuvent être réutilisées dès que l'occasion se présente. Un point intéressant qui permet de réaliser des économies. En effet, la matière créée pour un article peut-être réutilisée pour d'autres.

La confiance que l'on peut accorder aux chiffres ne doit pas être aveugle, ils sont aussi subjectifs que les mots. Il ne faut donc pas leur vouer une totale confiance sous prétexte qu'il s'agit de chiffres auxquels on peut prêter un caractère irréfutable.

L'exploration puis l'exploitation d'une base de données répond aux mêmes critères et exigences qu'une enquête journalistique classique. Les fondamentaux sont les mêmes. Il reste indispensable de vérifier et de recouper les données.

Le journalisme de données est un moyen de proposer une information enrichie. Le contenu proposé à l'internaute sera différent sur la manière dont il a été traité et sur l'expérience qu'il offre au lecteur. Le journalisme de données est un moyen de redonner de la valeur au contenu. Le lecteur n'est plus prêt à payer pour une information qu'il peut trouver ailleurs, il paie pour un contenu et une expérience exclusifs.

Le modèle économique des journaux sur internet peut s'en trouver modifié. La modification du modèle économique induite par le journalisme de données peut même s'avérer encore plus profonde. Il pourrait dans certains cas bouleverser la finalité même d'un journal. Sur le modèle des « Mediata Startups » dont l'objectif est de vendre des informations B2B, des données exclusives ou encore des analyses basées sur des données. Dans ce modèle, les articles n'ont qu'une visée commerciale pour attirer les clients vers les services payants et asseoir la marque.

BIBLIOGRAPHIE ET WEBOGRAPHIE

Sources transversales à l'ensemble du mémoire :

- Guide du Data Journalisme, Collecter, analyser et visualiser les données. Sous la direction de Jonathan Gray, Liliana Bounegru et Lucy Chambers et de Nicolas Kayser-Bril pour la version française. Editions Eyrolles, août 2013
- Alain Joannès, Data journalism, Bases de données et visualisation de l'information Editions CFPJ, décembre 2010.

Introduction

- Une autre façon de dire le monde, Courrier International n°1209, du 2 au 8 janvier 2014
- John Snow, Carte du Choléra, internet
- Mael Inizan, Chicago : le parrain du data journalisme, Owni, juin 2011
- Bill Dedman, The Color of Money, powerreporting.com/color
- Stephen Doig, Wikipedia

I. Qu'est ce qui change ?

1. Le lecteur devient dans certains cas, un collaborateur

- Entretien avec Caroline Goulard, horizonsmediatiqueslucilejeanniard.wordpress.com, paru en mars 2012.
- Crowdsourcing, Wikipedia
- Le crowdsourcing de données au Guardian datablog, Guide du datajournalism, éditions Eyrolles, ouvrage collaboratif supervisé par Nicolas Kayser-Bril
- Nicolas Kayser-Bril, Datajournalisme : ensemble, géolocalisons les bureaux de vote, Owni, paru en février 2010.
- Hubert Guillaud, Journalisme de données : data as storytelling. Site Internet actu, paru en juillet 2010

2. Des centres de données locaux mis à jour en temps réel

- Jim Meek et Patrick Sweet, Delaware Crime Map, site Delaware online.

- Top crimes by year, site Delaware online

3. La tentation de l'information micro-locale

- Sandrine Proton, 4 questions à Adrian Holovaty, fondateur du site EveryBlock. World association of newspaper and news publishers, paru en juin 2011.
- <http://www.everyblock.com/>
- Caroline Goulard, Le journalisme de données, site INA Global, la revue des industries créatives et des médias, paru en août 2010

4. Du consommateur d'information au conso-acteur

- Steve Doig, Ouverture des données, Guide du datajournalisme, p.178

5. Un travail nouveau pour le journaliste qui suppose de nouvelles compétences

- Estelle Prusker-Deneuville, Les enjeux du data-journalisme pour la Presse : Partie II, site Culture-RP, paru en octobre 2013
- Carte des Municipales, site du journal La Provence
- Xavier Lacombe, graphiques tirés de l'article Municipales en Provence : une campagne 2.0 ?, site du journal La Provence, paru en mars 2014

6. De nouvelles données

- Eric Scherer, Données publiques et journalisme : une mine de richesses !, Owni, paru en juillet 2010
- Dominique Cardon, Zoomer ou dézoomer ? Les enjeux politiques des données ouvertes, Owni, paru en février 2011.
- Yann Bresson, L'Open Data à la loupe, Youtube, publiée en décembre 2011
- Léa Lejeune, Interview d'Henri Verdier « La France est le premier pays à aller au bout de la logique de l'open data », Libération, paru en janvier 2014
- Gabrielle, Le Data journalisme pour les nuls, 123opendata, paru en juin 2012
- Mael Inizan, interview de Sylvain Parasie, Chicago de la prohibition au data journalisme, Siliconmaniacs, paru en mars 2011

- Alain Joannès, Data journalism Bases de données et visualisation de l'information, Editions CFPJ

7. Une nouvelle visualisation

- Gregor Aisch, Faire parler les données : une visualisation, Guide du datajournalisme, p. 165
- Une autre façon de dire le monde, Courrier International n°1209, du 2 au 8 janvier 2014
- John Snow, Carte du Choléra, internet
- Élysée, la vie de château ?, représentation visuelle, Paris-Match, du 30 mai au 5 juin 2013
- Martin Grandjean, le réseau du marché public, site pegasusdata.com, paru en novembre 2012
- Occurrence des mots selon les religions, site electronlibre.info, paru en juin 2012
- Hubert Guillaud, Journaliste de données : data as storytelling, Internet actu, juillet 2010

II. Ou en est on aujourd'hui et quels sont les freins ?

- Jean-Baptiste Chastand, Laetitia Clavreul et Alexandre Léchenet, Tarifs médicaux : enquêtes sur les dépassements d'honoraires, Le Monde, paru en avril 2012

1. Le journalisme de données : une pratique onéreuse

- Pierre Haski, Prix Pulitzer pour Pro Publica, un site à but non lucratif, Rue 89, paru en avril 2010
- Benoit Raphaël, Pro Publica : Journalisme à (très) haut coût, financé par des dons, Owni, paru en novembre 2010
- Dollars for doctors. How industry money reaches physicians, Pro Publica
- Sheri Fink, The Deadly Choices at Memorial, Pro Publica, paru en août 2009
- Estelle Prusker-Deneuville, Les enjeux du data-journalisme pour la presse : Part II, site culture-rp, octobre 2013

2. Un problème de formation et de compétences techniques

- Estelle Prusker-Deneuville, Les enjeux du data-journalisme pour la presse : Part II, site culture-rp, octobre 2013
- Tow center for digital journalism, Columbia Journalism School

3. Open data: l'enjeu de la libération des données

- Q.R., Open Data : définition, enjeux, perspectives, site le cube vert
- Tatiana, Les données ouvertes, futur eldorado du journalisme, entretien avec Valérie Peugeot, La voix du dodo
- Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal, Legifrance
- Divers auteurs, Data journalism : pourquoi les journalistes français ne s'y mettent pas ?, databasejournalism.wordpress.com, mars 2010
- Nicolas Kayser-Bril, « On l'a pas en format ordinateur », Owni, décembre 2009
- Estelle Prusker-Deneuville, Les enjeux du data-journalisme pour la presse : Part II, site culture-rp, octobre 2013

4. Redonner de la valeur au travail journalistique.

- Sophian Fanen, Le journalisme les doigts dans les données, Libération Ecrans, janvier 2012
- Sophian Fanen, Les journalistes à data sur les données, Libération Médias, janvier 2012
- Rafat Ali, Here come the Mediata Startups, LinkedIn, décembre 2013

5. Les atouts séduction du journalisme de données

- Caroline Goulard, Le journalisme de données, site INA Global, la revue des industries créatives et des médias, paru en août 2010

Conclusion

- Nicolas Kayser-Bril, Le Datajournalisme : vecteur de sens et de profits, Owni, avril 2011
- Hubert Guillaud, Journalisme de données : data as storytelling. Site Internet actu, paru en juillet 2010

ANNEXES

Entretien avec Nicolas Kayser-Bril réalisé par mail :

Le journalisme de données coûte plus cher que les autres formes de journalisme mais sur le long terme, puisqu'il est possible de réutiliser la base de données actualisée, cela devient intéressant (profitable?).

Tout dépend du modèle d'affaire du média en question. Je ne connais pas de média ou de section consacré au journalisme de données qui prétende être profitable. Mais le NYT semble pouvoir monétiser ses app, par exemple, en insistant sur l'engagement des utilisateurs.

Il y a donc un frein à l'entrée (se doter d'une équipe de data journalistes, constituer ses bases de données).

Oui, trouver les bonnes compétences est difficile.

Selon toi, qu'est ce qu'il faut faire ressortir concernant le modèle économique du journalisme de données?

Le modèle le plus intéressant est celui des "Mediata Startups". Lis ce post de Rafat Ali <https://www.linkedin.com/today/post/article/20131202172401-40578-here-come-the-mediata-startups>

Quels sont les freins à une pratique massive dans les rédactions?

Le manque de compétences, principalement, et un mode d'organisation archaïque, centré sur la production d'articles ou de clips audio et vidéo.

Est ce que le journalisme de données peut être une solution pour redonner de la valeur au journalisme.

Tout à fait. C'est un très bon moyen d'ajouter de la valeur à l'information, que ce soit en collectant des informations inexistantes, en les analysant ou en les mettant en forme.

Les gens ne veulent plus payer pour de l'information duplicable, des sujets repris presque à l'identique sur la plupart des sites, de l'enrobage de dépêches AFP. Est ce que le journalisme de données peut redonner de la valeur, remonétiser l'information ?

Encore une fois, c'est une question de modèle d'affaire. Il ne faut pas analyser de la même manière une société qui vend l'attention de ses utilisateurs à des annonceurs et une société qui vend du contenu. Les deux modèles d'affaire n'ont rien à voir.

Qu'en est il de la libération des données, l'Open Data, c'est un préalable indispensable.

Pas vraiment. On peut collecter et analyser des données qui ne sont pas de l'open data.

Est ce qu'il est facile d'accéder aux données que l'on veut ?

Non, les données de nombreuses entreprises ont une valeur stratégique, elles ne peuvent pas les communiquer. C'est au journaliste de trouver un moyen de collecter des informations équivalentes.

Est ce qu'elles sont disponibles dans des formats exploitables ?

Ça dépend. Les portails open data présentent encore souvent des données dans des formats qui ne sont pas ouverts (ce qui fait que ce n'est plus de l'open data).

Il existe apparemment des people person dont la tâche est de « harceler » les institutions et autres détenteurs de données.

Je ne savais pas. As-tu une source?

Il y a des réticences de certains qui ne souhaitent pas divulguer leurs données ?

Oui. Parfois à bon escient (une entreprise qui protège un actif) ou à mauvais escient (une administration qui refuse de respecter la loi du 17 juillet 1978)

Qu'est ce qui change par rapport au journalisme dit traditionnel ?

Énormément de choses changent. La plus importante, c'est la sortie du modèle centré sur l'article, pour aller vers un modèle où la valeur tient dans une base de donnée complète, qui va être exploitée de diverses manière (l'une d'entre elles étant d'écrire des articles). Ce post de Holovaty de 2006 est toujours d'actualité <http://www.holovaty.com/writing/fundamental-change/>

La verticalité de l'information est moins marquée, on est moins dans un schéma rigide de l'information qui est délivré par le journaliste au lecteur. Si on reprend l'exemple d'Every Block, je pense que les journalistes peuvent s'en servir pour récolter des informations fournis par les membres. J'ai relevé une participation du lecteur au début et à la fin de la chaîne. Au début, c'est le fameux exemple des lecteurs du Guardian qui ont aidé à analyser les notes de frais des parlementaires britanniques. A la fin de la chaîne, c'est des graphiques modulables, le lecteur qui peut se balader et régler les paramètres de l'infographie.

Tu as tout compris.

J'ai l'impression que le data journalism est en progression, c'est évident, mais globalement les journaux français externalisent pas mal. Je me trompe peut être mais l'existence de société comme la tienne ou WeDoData semble le révéler. Pourquoi les médias font appel à des prestataires extérieurs alors qu'ils pourraient le traiter en interne ? Tu as du travailler avec différents médias, comment tu juges leur degré de réceptivité et leur volonté de s'y mettre ? Est ce qu'il y a un problème de formation, de compétences au sein des rédactions ?

Les compétences sont difficiles à trouver. Mais le problème des médias français reste qu'aucun d'entre eux n'a de stratégie sur le sujet. Plusieurs ont embauché des compétences intéressantes (Le Monde avec Alexandre Léchenet, Le Temps avec Jean Abbiateci). Mais aucun n'a constitué d'équipe, avec des processus clairs, des budgets propres... Ce qui permet de ne pas se baser uniquement sur des individus et de gagner en productivité avec l'expérience. De telles équipes existent chez Zeit online, The Times (Londres), NPR, LA Times, Financial Times et d'autres.

Dernièrement, Le Monde a décidé d'intégrer et de mettre en avant l'un de ses blogs « Les décodeurs ». Désormais, tu les retrouves régulièrement en Une du site. C'est surtout du fact checking de la parole politique et notamment des chiffres qui parsèment les discours des politiques. A ce titre, les décodeurs qui vérifient la véracité de ces statistiques, de la parole politique, c'est du journalisme de données ? Si oui, c'est un petit événement qu'un journal comme Le Monde prenne ce virage.

Je ne trouve pas que les décodeurs est un virage en termes de journalisme de données. Pour moi, la véritable nouveauté du journalisme de donnée tient dans le mode d'organisation: Des développeurs, des chefs de projet et des journalistes qui créent et font vivre des produits éditoriaux. Au Monde, ce sont des journalistes qui écrivent des articles. Des articles avec des charts, qui sont de meilleure qualité que le reste, mais ça n'est pas un virage ni une stratégie nouvelle.

Enfin, au niveau de la réception du public. Il est peut être plus facile pour certaines histoires de les raconter visuellement plutôt qu'avec de longs textes, la force du visuel sur les mots. A notre époque hyper connectée, je pense qu'on a développé une sorte d'intelligence et de mémoire visuelle. C'est un atout pour séduire le public, c'est le genre de contenu qu'il recherche ?

La communication par les images n'a rien de nouvelle. C'était nouveau entre 1850 et 1930. Tu devrais lire à ce propos ce qu'a fait Otto Neurath http://en.wikipedia.org/wiki/Otto_Neurath

Un site comme OWNI pour lequel tu as travaillé, est ce que ce n'est pas le cas typique du journal trop en avance sur son public ?

Non, l'échec d'Owni est surtout financier. Ça n'aurait pas plus marché aujourd'hui.

A l'époque du lancement d'Owni, le journalisme de données en étaient à ses balbutiements en France, peu voyait le véritable potentiel de celui-ci et il devait être complètement inconnu du grand public. Aujourd'hui, j'ai l'impression qu'on commence à prendre un peu la mesure des possibilités qu'il offre. Tu en penses quoi ?

Je n'ai rien vu qui permette de penser que le regard du public sur les productions de data journalism a évolué. Le succès d'un projet tient, comme avant, à la qualité de l'histoire qu'il raconte et à son accessibilité (mise en forme).