

# Data journalisme

## Un métier d'avenir

Martine Otter

*De nouveaux métiers apparaissent avec le numérique, tandis que d'autres disparaissent ou sont profondément transformés. Cet article aborde l'évolution du métier de journaliste à l'ère d'Internet et du Big data.*

*Le data journalisme ou journalisme de données est-il un nouveau métier ? Une façon habile d'utiliser les nouvelles technologies pour retenir l'attention des lecteurs ?*

*Afin de mieux comprendre ce qu'il en était, j'ai suivi le MOOC « Doing journalism with data : first steps, skills and tools » proposé sur la plateforme Canvas ([www.canvas.net](http://www.canvas.net)). Les vidéos étaient en anglais, sous-titrées en français et accompagnées de nombreux liens et exercices pratiques de traitement de données. J'ai ainsi pu m'initier aux méthodes et outils du data journalisme.*

## Le métier de journaliste : de la littérature au traitement de données

Comme nous l'indiquions dans la Lettre n° 73, pour pouvoir parler d'un métier, il faut que plusieurs conditions soient réunies :

- la reconnaissance sociale, qui peut s'exprimer par l'existence d'une organisation représentative, telle que l'ordre des médecins ou l'ordre des avocats ;
- la reconnaissance économique sur le marché du travail, matérialisée par des offres d'emploi et des grilles de rémunération, car le métier est d'abord un travail dont on peut tirer un revenu pour gagner sa vie (gagne-pain) ;
- l'existence de formations, diplômantes ou non, permettant d'acquérir les compétences nécessaires à l'exercice du métier ;
- éventuellement, des règles morales spécifiques, formalisées dans un code de déontologie.

En ce sens, le journalisme, en tant que métier, n'est apparu qu'au XIX<sup>e</sup> siècle. Ce n'était, auparavant, qu'une activité littéraire, souvent annexe, de quelques écrivains. On est alors passé progressivement du journalisme d'opinion au journalisme d'information, avec la création de l'agence Havas, première agence de presse en 1832.

Dans les années quatre-vingt, le journaliste utilisait un carnet, un magnétophone, quelquefois une caméra et avait recours aux archives de bibliothèque pour rédiger ses articles.

Le journalisme Web est apparu dès 1995 avec le développement de la presse en ligne mais n'a vraiment été perçu en France comme un métier spécifique qu'à partir de 2007, à l'occasion des élections présidentielles.

20minutes.fr avait à cette occasion calculé et publié des audiences et parts d'audience des vidéos politiques sur Internet en extrayant des données journalistiques sur les plateformes de diffusion de vidéos numériques DailyMotion et Youtube<sup>12</sup>.

Un métier c'est une activité associée à un statut. Le statut du journaliste est matérialisé en France par la possession d'une carte de presse, carte d'identité professionnelle dont l'attribution est définie par l'article L. 7111-6 du code du travail<sup>13</sup>. Cette carte permet au journaliste de prouver son activité et de bénéficier de différents avantages, y compris celui d'un abattement d'impôts spécifique. Les journalistes en ligne peuvent, tout comme les journalistes classiques et photojournalistes, bénéficier d'une carte de presse, à condition de remplir les conditions requises, dont celle d'une activité rémunérée régulière, exercée à titre principal. Mais cela n'est pas forcément simple, la collaboration à un journal gratuit n'ouvre pas forcément droit à une carte de presse, en particulier lorsque le journal intègre une proportion de publicité importante. Cette obtention devient d'autant plus délicate, qu'aujourd'hui les journalistes salariés permanents sont de plus en plus rares, une grande proportion de journalistes travaillant pour la presse et les sites d'information en ligne étant indépendants.

Le métier de journaliste se renouvelle ainsi avec les nouveaux médias en ligne : de nouvelles plateformes amateurs sont venues concurrencer les médias historiques. En retour, les colonnes de ces derniers se sont ouvertes aux lecteurs, sous le contrôle d'un community manager.

<sup>12</sup><http://videolab.lhivic.org/?cat=7>

Voir la timeline « french media and Data Journalism » sur <http://www.timetoast.com/timelines/french-media-and-data-journalism>

<sup>13</sup><http://vosdroits.service-public.fr/professionnels-entreprises/F22389.xhtml>

Le journalisme, confronté à la crise économique, se caractérise sur le Web par des exigences de rentabilité croissantes, associées à la mesure des taux de clics. Les blogs ont permis à chacun de se croire journaliste : rien ne semble plus facile que d'exprimer une opinion ou de rapporter des faits en ligne.

Que reste-t-il au journaliste professionnel s'il veut dépasser le simple rôle de relais des agences de presse ? Faire comprendre les faits reste son véritable métier... et, pour ce faire, de nouvelles voies sont apparues avec le développement de l'open data et du big data, ces masses de données à analyser et à traduire en informations compréhensibles.

Il apparaît que les pages de données et les infographies génèrent du trafic sur les sites des médias en ligne. À titre d'exemple on peut citer les pages data du Texas Tribune<sup>14</sup> qui représentent entre 45 et 55 % du trafic du site Web.

## Le data journalisme

### Raconter des histoires

Les Anglo-Saxons définissent le data journalisme comme la présentation des faits de la meilleure façon possible en utilisant des chiffres. Notons que la traduction de l'expression anglaise « story telling » en français par « présentation des faits » nous fait passer d'un monde d'interprétation, celui où l'on raconte des histoires à ses lecteurs, à un monde apparemment plus rationnel, celui des faits. Mais ne nous trompons pas : même si le journaliste s'efforce de recueillir et de présenter différents points de vue, rapporter un fait suppose de faire des choix dans les éléments de réalité observés et dans la présentation du contexte. Mathématiques, graphiques et codes ne sont qu'au service de l'histoire à raconter.

### Visualiser les données

Le travail du data journaliste consiste à transformer les données en faits. Il s'agit de décrypter les faits derrière les informations via des analyses.

On assimile souvent data journalisme et visualisation des données. La visualisation des données est en effet une technique privilégiée du data journalisme, mais ce n'est pas la seule. L'interprétation des données peut aboutir à de superbes graphiques mais aussi à de longs articles bien documentés.

Elle n'est d'ailleurs pas apparue avec l'Internet : on cite la carte de Londres établie en 1850 par John Snow qui y localisait les cas de choléra et y a noté la proximité de certains points d'eau.

Traditionnellement certains domaines tels que le sport, la finance, les crimes ou les résultats d'élection avaient, bien avant Internet, une tradition de communication de chiffres. Au temps du papier, les journaux de la presse financière, tels le Wall Street Journal, présentaient des analyses de données mais restaient dépendants des statisticiens.

### Les précurseurs

L'apparition du data journalisme s'explique par deux grands changements :

- d'une part, la possibilité pour le journaliste de traiter directement de grands volumes de données avec des outils de type tableur ;
- d'autre part, l'accès aux grandes masses de données ouvertes de l'open data.

Parmi les précurseurs du data journalisme on citera Philip Meyer, Adrian Holovaty et Clarence Jones.

- Clarence Jones du Miami Herald utilisait déjà l'informatique en 1969 pour repérer des structures dans le système de justice pénale. Les données étaient saisies sur cartes perforées et le programme d'analyse écrit en Cobol. L'objectif initial de détecter la corruption dans le système ne fut pas atteint mais les données collectées révélèrent entre autres que la plupart des arrestations ne débouchaient jamais sur des peines de prison.
- Philip Meyer a publié en 1973 « precision journalisme »<sup>15</sup>, ouvrage où, en réaction à la complexité croissante de l'information et à des interrogations sur le caractère utopiste de la notion d'objectivité, il défendait l'introduction des méthodes des sciences sociales et comportementales dans le journalisme et expliquait aux futurs journalistes l'exploitation des statistiques et les techniques d'analyse de données.
- Adrian Holovaty<sup>16</sup> est le prototype parfait du créateur polyvalent, développeur, journaliste, community manager et musicien. Il exposait dans un article paru en 2006<sup>17</sup> les principes de structuration des données sur les sites des journaux en ligne.

En France, le site OWNI (owni.fr), lancé le 6 avril 2009, a été à l'avant-garde du data journalisme en contribuant à l'analyse des documents collectés par Wikileaks.

<sup>14</sup><http://www.texastribune.org/library/data/>

<sup>15</sup><http://www.unc.edu/~pmeyer/book/>

<sup>16</sup><https://www.linkedin.com/in/adrianholovaty>

<http://www.holovaty.com/>

<sup>17</sup><http://www.holovaty.com/writing/fundamental-change/>

OWNI a malheureusement fermé en 2012, mais ses créateurs continuent à contribuer à la diffusion des méthodes du data journalisme et à sa professionnalisation : Nicolas Kayser-Brill est en particulier l'un des conférenciers intervenant sur le MOOC « Doing journalisme with data » où il expose différentes techniques de nettoyage des données. L'agence WeDoData ([wedodata.fr](http://wedodata.fr)), également française, a créé des infographies pour Radio France et Libération.

## Quelles compétences ? Quels outils ?

Les compétences du data journaliste couvrent à la fois la programmation, le design et le journalisme, sans oublier un minimum de connaissance en matière de statistiques. Le data journaliste doit maîtriser au moins l'un de ces domaines. Ceci explique qu'il s'agisse très souvent d'un travail d'équipe, même si un travail individuel est possible pour un professionnel polyvalent.

### Comment transformer les chiffres en histoire ?

La méthode du journaliste tient en quelques lettres qui résument le fameux hexamètre de Quintilien QQOQCCP (pour « Qui ? Quoi ? Où ? Quand ? Comment ? Combien ? Pourquoi ? ») ou « cinq W », pour « Who, What, Where, When, Why ? » :

- Who ? D'où viennent les données ? Justifier de leur origine est essentiel pour établir la confiance.
- What ? Quelle est la question posée ? que s'agit-il de montrer, de dire ?
- When ? De quand datent les données ?
- Where ? Peut-on présenter les données de façon localisée ?
- Why ? Peut-on établir des corrélations entre les séries de données ? Comment éviter des conclusions hâtives sur les relations de cause à effet.

### La démarche

Le processus métier du data journaliste débute, suivant le cas, par une question qui nécessite des données, ou par des données qui doivent être questionnées.

#### Obtenir les données

C'est la première étape.

Les méthodes pour trouver des données sont multiples : on peut poser des alertes sur les sites officiels tels que les sites gouvernementaux ou ceux de l'Union européenne qui publient régulièrement des données statistiques ou utiliser leurs fils RSS. À défaut, le site [ChangeDetection.com](http://ChangeDetection.com) permet par exemple d'être informé de tout changement survenu sur un site particulier.

Les organisations statistiques nationales sont une source importante de données ainsi que les organisations internationales telles que l'ONU ou la Banque mondiale, les ONG sur des sujets plus spécialisés, les entreprises au travers de leurs rapports annuels, les syndicats et organisations professionnelles. Les initiatives open data représentent une part croissante de ces sources.

Le data journaliste se doit de maîtriser les moteurs de recherche en exploitant leurs fonctionnalités avancées<sup>18</sup>. Des outils de « Web scraping » ou « Extraction de contenu » permettent de récupérer directement le contenu de tables ou de listes présentes sur des pages Web, sans avoir à écrire une seule ligne de code. Google Drive<sup>19</sup> ainsi qu'OutWit<sup>20</sup> Hub, tous deux gratuits, offrent cette fonctionnalité.

Lorsque les informations ne sont pas en ligne, le data journaliste pourra en demander la communication auprès des organismes qui les détiennent. Dans beaucoup de pays, des lois encadrent la communication des informations. Il faut savoir frapper à la bonne porte. Le site <http://www.wobbing.eu> fournit des informations sur la législation applicable en la matière dans les différents pays de l'Union européenne. En France, par exemple, il est fait référence à la loi 78-753 du 17 juillet 1978 qui définit la liberté d'accès aux documents administratifs, ainsi qu'à la loi 79-18 du 3 janvier 1979 qui régle l'accès aux archives et la loi 2000-321 du 12 avril 2000 qui définit les droits des citoyens dans leurs relations avec les administrations. On notera ainsi que « Les budgets et les comptes des autorités administratives mentionnées à l'article 1er et dotées de la personnalité morale sont communicables à toute personne qui en fait la demande ». On citera également la plateforme <http://www.asktheeu.org/> qui peut être utilisée pour demander des informations détenues par la Commission Européenne.

Lorsque des données ne sont pas disponibles sur un sujet particulier, l'appel au Crowdsourcing peut être une solution, moyennant quelques précautions sur les méthodes de collecte et de vérification.

#### Les nettoyer

Les données récupérées nécessitent la plupart du temps un nettoyage et différentes transformations, afin d'harmoniser les différents codages et formats de données numériques ou de dates, éliminer les doublons ou les données aberrantes, recadrer les champs, corriger les noms de lieu... Cela demande un peu de doigté. Il existe heureusement divers outils qui peuvent faciliter ce travail, tel qu'OpenRefine, un logiciel open source très pratique, ou plus simplement Excel.

<sup>18</sup> Voir : [http://www.google.com/advanced\\_search](http://www.google.com/advanced_search) et [http://www.googleguide.com/advanced\\_operators\\_reference.htm](http://www.googleguide.com/advanced_operators_reference.htm)

<sup>19</sup><http://schoolofdata.org/handbook/recipes/liberating-html-tables/>  
<sup>20</sup><https://www.outwit.com/products/hub/>

Il est également possible de faire appel au Crowdsourcing pour le nettoyage des données, à condition de fournir aux utilisateurs une interface suffisamment simple<sup>21</sup>.

### **Combiner les données pour les faire parler**

Le traitement statistique des données n'est pas forcément complexe mais si l'on veut établir des comparaisons entre différentes séries de données mieux vaut savoir ce que sont une moyenne, une médiane et savoir calculer un pourcentage d'évolution. Un outil comme Excel fournit toutes les fonctions statistiques nécessaires, encore faut-il en comprendre la signification. Le data journaliste devra apprendre à les manipuler et à combiner les données, en utilisant par exemple les tableaux croisés dynamiques. C'est à ce prix qu'il pourra faire mettre en évidence les faits cachés derrière les données.

### **Communiquer le résultat**

C'est l'étape finale de la mise en forme, celle de la datavisualisation. La visualisation ou infographie est une représentation graphique qui se doit d'être évidente. C'est à la fois un outil d'analyse de communication et de compréhension. C'est un moyen unique d'extraire du sens et de voir les structures et les tendances dans un ensemble de données.

Comment choisir le type de représentation à utiliser ? Avant de se lancer dans l'utilisation de son outil graphique préféré, que ce soit Adobe Illustrator, D3, Data Wrapper, Google Fusion Tables ou tableau ou un autre, le plus simple est de prendre un crayon, un papier et son cerveau. L'esthétique n'est pas tout et il ne suffit pas d'être un bon designer pour créer une bonne visualisation. La représentation doit être à la fois esthétique, fonctionnelle et pertinente. Belle pour attirer le regard du lecteur, fonctionnelle pour être comprise sans avoir besoin de lire les chiffres et sans effort d'interprétation et pertinente dans la réponse qu'elle apporte aux questions posées dans l'article. Beaucoup d'infographies sont belles et fonctionnelles mais peu pertinentes. Dans sa conférence sur les principes de la datavisualisation,

Alberto Cairo cite le cas d'une superbe visualisation des langues utilisées en Europe sur Twitter<sup>22</sup> : la visualisation est effectivement très esthétique mais n'apporte pas vraiment d'information. On y voit seulement que c'est la langue du pays qui est le plus communément utilisée : rien de vraiment très surprenant ! Albert Cairo indique dans ce cas qu'il aurait été plus pertinent de montrer les exceptions (les régions qui émettent des tweets dans d'autres langues que la langue du pays) plutôt que les grandes tendances trop évidentes.

Le choix de la représentation graphique à utiliser dépendra de la question posée : suivant que l'on souhaite comparer des données de différentes catégories ou différentes périodes, faire apparaître des relations entre plusieurs variables ou montrer la répartition géographique de certains phénomènes, on utilisera tel ou tel type de graphique ou de représentation cartographique<sup>23</sup> : courbe, histogramme, nuage de points... Il existe sur le Web de nombreux sites spécialisés dans la publication d'infographies où puiser l'inspiration. La plupart sont en anglais : FlowingData<sup>24</sup>, Information aesthetics<sup>25</sup>, Cool Infographics<sup>26</sup>. La représentation graphique n'exclut pas, bien évidemment, le développement d'explications textuelles qui doivent préciser l'origine des données et la méthodologie employée et permettent de souligner tel ou tel aspect des données présentées. Il ne faut d'ailleurs pas oublier que les textes explicatifs sont indispensables à la lecture par les personnes malvoyantes ou non-voyantes.

## **Conclusion**

Les grands titres américains ont pour la plupart mis en place des équipes de rédaction spécialisées dans le data journalisme. Cela reste encore rare en France, même si les journaux en ligne, tels que le monde.fr multiplient les infographies.

Mise en cause par la montée des nouveaux médias, la confiance dans la presse traditionnelle s'est émoussée. Le data journalisme en s'appuyant sur des données vérifiées et vérifiables peut contribuer à retrouver la confiance des lecteurs, en jouant un rôle de médiateur entre des données pléthoriques et le public qui n'a ni le temps ni les moyens de les analyser. ▲

***[martine.otter@adeli.org](mailto:martine.otter@adeli.org)***

<sup>21</sup> Voir par exemple la plateforme [crowdcrafting.org](http://crowdcrafting.org)

<sup>22</sup><http://flowingdata.com/2011/10/27/language-communities-of-twitter/>

<sup>23</sup>[http://extremepresentation.typepad.com/blog/files/choosing\\_a\\_good\\_chart.pdf](http://extremepresentation.typepad.com/blog/files/choosing_a_good_chart.pdf)

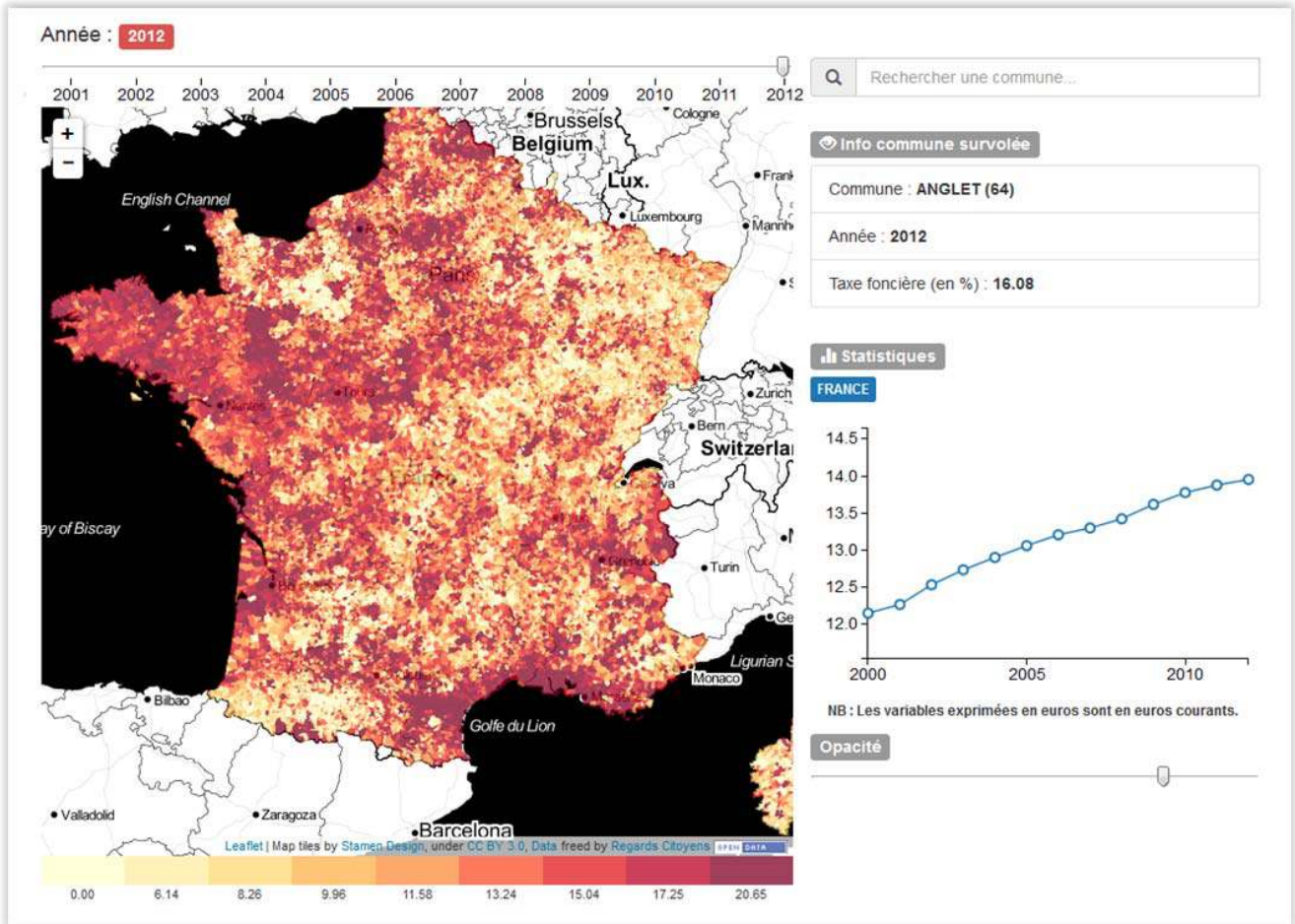
<sup>24</sup> <http://Flowingdata.com>

<sup>25</sup><http://infosthetics.com/>

<sup>26</sup><http://www.coolinfographics.com/>

## Quelques exemples

Sur le site [www.nosfinanceslocales.fr](http://www.nosfinanceslocales.fr) une infographie interactive présentant la répartition de la taxe foncière.



Sur [www.nosdeputes.fr](http://www.nosdeputes.fr)

