



HAL
open science

Le journalisme au défi des big data. Mort annoncée ou mutation salvatrice ?

Alexandre Joux, Marc Bassoni

► **To cite this version:**

Alexandre Joux, Marc Bassoni. Le journalisme au défi des big data. Mort annoncée ou mutation salvatrice?. 2019. hal-02160315

HAL Id: hal-02160315

<https://hal.science/hal-02160315>

Submitted on 19 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le journalisme au défi des big data. Mort annoncée ou mutation salvatrice ?

fr.ejo.ch/deontologie-qualite/journalisme-defi-big-data-mort-annoncee-mutation-salvatrice-datajournalisme-
everyblock
13 juin 2019 • À la une , Déontologie et qualité • by Marc Bassoni | Alexandre
Joux



Les données massives sont capables de mettre en valeur le travail de terrain du journaliste. (Crédit photo: Wikipédia)

La déferlante des données massives interroge le journalisme contemporain, ses pratiques, mais aussi les modèles économiques des médias et l'organisation des rédactions. Au-delà des résistances épistémologiques que cela suscite, les chercheurs de l'IMSIC Marc Bassoni et Alexandre Joux montrent comment les *big data* peuvent aussi contribuer à recentrer la pratique autour de ses fondamentaux, que sont notamment le terrain et le lien avec le public.

Quand l'expression *big data* (données massives) est associée au mot journalisme, il y a fort à parier qu'un tel rapprochement déclenche, en ces temps de grands bouleversements techniques, un quasi-réflexe de représentation du journalisme automatisé (ou « robot-journalisme »). Ce dernier se caractérise par une production de contenus opérée par des logiciels à partir d'un traitement algorithmique de grandes masses de données.

Présenté comme un modèle d'« objectivité », le « robot » évince ainsi doublement la médiation humaine, d'une part en s'arrogeant la fonction exclusive de « *gate-keeper* » et, d'autre part, en renouvelant le rapport au réel (neutralisation supposée de toute forme de subjectivité dans le traitement de l'information).

A l'aune de cette représentation, les *big data* semblent sonner le glas du journalisme traditionnel ; ce dernier s'inscrivant ainsi sur la (déjà) longue liste des pratiques professionnelles condamnées à la disparition à brève échéance.

A rebours de cette prophétie, nous voudrions esquisser une voie alternative. Dès lors que les algorithmes associés aux masses de données peuvent être utilisés autrement, ces derniers peuvent au contraire catalyser une mutation salvatrice du métier en renforçant, d'une façon quasi-paradoxe, deux de ses caractéristiques fondamentales, à savoir la valorisation du travail de terrain et l'approfondissement des liens avec le public destinataire de l'information. Ce sont ces deux aspects de la question que nous allons ici rapidement passer en revue.

Valoriser le travail de terrain

Loin de toute prétention des algorithmes à faire émerger de manière normée des faits et à toujours dire le « vrai » à leur propos, il est en effet possible d'envisager un usage professionnel des *big data* compatible avec les fondamentaux du métier, et en particulier dans ses rapports au terrain. Le recours aux données massives n'est pas un obstacle au travail d'investigation classique ; il en est plutôt le nouveau marchepied.

Deux pistes peuvent être soulignées. La première a trait à ce qu'Eric Scherer appelle le « tri algorithmique »^[1]. En brassant de grandes quantités de données, les algorithmes peuvent mettre au jour des choses que l'œil humain ne saurait appréhender. Partant, les algorithmes diffusent les fameux « signaux faibles » qui peuvent suggérer à des rédactions de creuser des sujets singuliers ou émergents et d'arpenter de nouveaux terrains d'investigation.

Toutes proportions gardées, c'est la même démarche qu'empruntent les sciences de la santé. Les algorithmes « moulinent » de grandes quantités d'informations-patients anonymisées et tentent de mettre en exergue des corrélations inattendues, lesquelles permettront ensuite de forger des hypothèses alternatives et d'orienter la recherche médicale sur des voies thérapeutiques nouvelles.

Renforcer les exigences de vérification

La seconde piste qui reconduit au terrain est celle de la « structuration » des bases de données utilisées. Nicolas Kayser-Bril et ses collègues promoteurs du « datajournalisme » insistent beaucoup sur ce point^[2]. En effet, structurer une base de données n'est pas une opération neutre. De nombreux biais peuvent l'affecter.

Dans une étude publiée récemment, nous soulignons ainsi les déboires qu'a connus le site *everyblock.com* lancé dans les années 2000 par Adrian Holovaty[3]. Pionnier du datajournalisme, ce site superpose sur une carte *Google Maps* des villes américaines, en quasi temps réel, les données sur la criminalité fournies en *open data* par les services de police. Sauf que cette géographie dynamique de la criminalité trahit parfois la réalité du terrain.

Comme nous le rappelons, « sur *everyblock.com*, le *City Hall* de Los Angeles apparaît comme l'endroit le plus criminogène de la ville, alors que les journalistes ne ressentent pas cette réalité sur ce terrain bien connu d'eux (puisqu'il s'agit de l'endroit où ils ont leurs bureaux !). Après enquête (...), ils découvriront ainsi que le *Los Angeles Police Department* (LAPD), dont *everyblock.com* exploite les données, assigne par défaut le code postal du *City Hall* à tous les crimes dont la géolocalisation est problématique, ce qui provoque ce biais majeur ».

Ce cas d'école souligne pour les journalistes l'importance de la maîtrise des bases de données exploitées afin, *via* un retour sur le terrain, de tester la pertinence et la robustesse des informations issues du traitement automatique. Dès lors, loin de menacer l'investigation journalistique, les *big data* peuvent au contraire renforcer les exigences de vérification et de contextualisation des informations, deux éléments « canoniques » du métier.

Consolider les liens tissés avec le public

Au-delà du terrain, les *big data* s'invitent également dans les logiques marketing des entreprises d'information. Leur premier effet est bien sûr d'accélérer l'assujettissement de la production éditoriale à la perception des attentes du public. La plupart des médias procèdent à une veille permanente *via Google Trends* et les réseaux sociaux (*hashtags* les plus discutés) pour identifier les questions qui semblent fortement concerner leurs propres destinataires.

Des algorithmes prédictifs peuvent même renforcer cette veille en soulignant les sujets et les angles de couverture qui pourraient devenir viraux et garantir ainsi l'audience la plus large possible. En fonction de telles indications, les médias agencent et reconfigurent en temps réel leur offre de contenus.

L'enjeu plus fondamental est de personnaliser toujours plus finement l'information produite. S'opère ici un « effet de boucle » : les outils prédictifs permettent de mieux cibler l'offre ; celle-ci suscite alors un intérêt plus grand qui favorise les interactions avec le public ; en retour, les médias concernés affinent leur connaissance de leurs segments d'audience et se mettent ainsi en capacité de personnaliser leur production, c'est-à-dire de la « démassifier ».

A cette fin, la pratique croissante de « l'information-service » et le développement du journalisme dit « de solutions » – ou journalisme « d'impact », pour reprendre l'expression de Cathrine Gyldensted^[4] – s'avèrent des plus précieux.

En tant qu'information utile pour la vie quotidienne, issue d'une source identifiée (extérieure au média diffuseur) et transmise au public sans retouche éditoriale, « l'information-service » catalyse désormais des interactions entre public et médias, interactions qui favorisent l'enrichissement des bases de données commerciales de ces derniers. Typiquement, il peut s'agir d'une information qui émane d'une collectivité locale ou d'un organisme public et qui concerne, par exemple, un projet d'amélioration relatif à la fourniture d'un service d'intérêt général ; il peut également s'agir d'une initiative prise par une association, une entreprise ou une ONG et qui concerne toute une communauté.

Ces informations enclenchent des interactions qui permettent d'affiner la connaissance des « destinataires-clients » (*via* leurs traces, qu'il s'agisse de leurs déclarations, de leurs choix, de leurs goûts). Elles autorisent la fourniture de contenus « sur mesure » et permettent *in fine* de « qualifier » ces bases de données en vue d'une monétisation par d'autres services des médias concernés.

Favoriser l'implication des destinataires de l'info

C'est un enjeu du même ordre qui est associé au journalisme dit « de solutions ». Celui-ci favorise l'implication des destinataires de l'information au cœur même de la production journalistique. Ces derniers sont non seulement des citoyens auxquels on vient proposer des « solutions » pour affronter les problèmes du présent, mais également des pourvoyeurs de *data* monétisables au bénéfice d'autres activités commerciales rassemblées sous une même « marque-ombrelle ».

Évidemment, selon les différents types de médias et leurs « missions » respectives (depuis l'adoption, en France, du « Plan d'action pour la croissance et la transformation des entreprises » [PACTE, 2019], la notion de « mission » dévolue à l'entreprise a été élargie au-delà de la maximisation des intérêts des seuls actionnaires), le côté « citoyen », ou le côté « client », du destinataire de l'information sera plus ou moins privilégié.

Cette problématique « marketing » permet de bien appréhender les enjeux relatifs à la gestion de la manne publicitaire comme à la fidélisation des abonnés. En affinant la connaissance des attentes du public, les algorithmes renforcent la rentabilité liée à la monétisation de l'audience auprès des annonceurs et ils permettent de pousser vers les abonnés les articles qui sauront le mieux retenir leur attention, cette stratégie étant notamment déployée par le prestigieux *New York Times*.

Cette exigence de personnalisation explique d'ailleurs la bataille qui se joue actuellement entre ces médias et les GAFAs au sujet de l'accès aux données des lecteurs. En effet, au-delà des revenus publicitaires et de leur partage, le sujet le plus sensible concerne la maîtrise pleine et entière des données-clients, gage d'une vraie autonomie vis-à-vis des poids lourds de l'économie numérique. Si les *big data* offrent potentiellement aux médias des outils inégalés de réponse aux attentes labiles de leurs destinataires, ces mêmes *big data* pourraient également permettre de court-circuiter ces médias en termes de retombées économiques ; étrange paradoxe lié à la structuration oligopolistique de l'économie numérique contemporaine.

Acculturation des professionnels et des données

Pour que puisse s'opérer cette mutation du métier de journaliste, deux conditions doivent *in fine* être posées. La première concerne l'acculturation des professionnels de l'information aux facettes marketing de leur activité. Cette acculturation ne va pas encore de soi comme l'a montré, par exemple, Pauline Amiel dans son étude des journalistes « localiers » de la presse quotidienne régionale française[5].

Des freins qui relèvent de la conception de « l'identité professionnelle » sont toujours actifs. En effet, les journalistes français entretiennent traditionnellement une représentation de leur activité dépourvue de tout ancrage commercial ou communicationnel. À l'aune d'une telle image, toujours prégnante, l'activité journalistique s'inscrit donc dans une économie de l'offre supposée protégée de tout tropisme pro-marketing. Il va sans dire que l'intrusion récente des nouveaux médias et des nouvelles formes de concurrence ont vivement, et douloureusement, percuté un tel stéréotype ...

La seconde condition concerne l'acculturation aux données et à leur traitement. Elle a trait au décloisonnement interne des entreprises médiatiques. Par décloisonnement, il faut entendre un maillage plus étroit entre la rédaction proprement dite et les nouvelles compétences qui peuvent l'enrichir (design graphique, *data mining*, ergonomie cognitive, *community management*, ...).

Des obstacles sont là encore toujours prégnants, héritages d'organisations traditionnelles dites « en silos ». On le voit bien, la prospective autour des *big data* remet en pleine lumière des questions très classiques, et somme toute anciennes, de transformation des métiers et d'organisation productive.

Cet article est publié sous licence Creative Commons (CC BY-ND 4.0). Il peut être republié à condition que l'emplacement original (fr.ejo.ch) et l'auteur soient cités, mais le contenu ne peut pas être modifié.

Références

[1] Scherer Eric (2015) *Logiciels, algorithmes, robots : journalisme automatique, méta-media#8,*

Paris, France Télévisions.

[2] Kayser-Bril Nicolas, Gray Jonathan, Bounegru Liliana, Chambers Lucy (dir.) (2013), *Guide du datajournalisme : collecter, analyser et visualiser les données*, Paris : Eyrolles.

[3] Joux Alexandre, Bassoni Marc (2018), « Le journalisme saisi par les Big Data ? Résistances épistémologiques, ruptures économiques et adaptations professionnelles », *Les Enjeux de l'information et de la communication*, n° 19/2 : 125-134.

[4] Gyldensted Cathrine (2015), *From Mirrors to Movers: Five Elements of Positive Psychology in Constructive Journalism*, Seattle : CreateSpace Independent Publishing Platform.

[5] Amiel Pauline (2017), *L'identité professionnelle des localiers à l'heure des mutations économiques et de la dématérialisation de la presse locale*, Thèse de Doctorat de l'Université de Toulouse-Paul Sabatier, Toulouse, 24 novembre.