

## **Le journalisme saisi par les Big Data ? Résistances épistémologiques, ruptures économiques et adaptations professionnelles**

*Are Big Data invading Journalism?*

*Epistemological Resistances, economic Breakings and professional adaptations*

*¿Está tocado por el Big Data el periodismo?*

*Resistencias epistemológicas, rupturas económicas y adaptaciones profesionales*

*Article inédit, mis en ligne le 15 novembre 2018.*

### **Alexandre JOUX**

*Alexandre Joux est maître de conférences en Sciences de l'information et de la communication, chercheur à l'IMSIC EA 7492 et directeur de l'École de journalisme et de communication d'Aix-Marseille (EJCAM). Ses recherches portent sur l'économie de l'information et des industries culturelles, en particulier dans les contextes de convergence et d'internationalisation, et sur les évolutions du journalisme dans son rapport aux environnements médiatiques et numériques.*

*alexandre.joux@univ-amu.fr*

### **Marc BASSONI**

*Marc Bassoni est maître de conférences (HDR) en Sciences de l'information et de la communication, chercheur à l'IMSIC EA 7492 et directeur des études à l'École de journalisme et de communication d'Aix-Marseille (EJCAM). Ses recherches portent sur l'économie des médias et l'évolution des pratiques éditoriales et journalistiques à l'aune de la révolution numérique.*

*marc.bassoni@univ-amu.fr*

### **Résumé**

La déferlante des données massives (*big data*) interroge le journalisme contemporain, ses pratiques, mais aussi les modèles économiques des médias et l'organisation des rédactions. Les enjeux sont protéiformes. Nous mettons tout d'abord en évidence la résistance épistémologique du métier face aux promesses d'un journalisme intégralement robotisé. En effet, le data journalisme se réfère toujours à un terrain identifié auquel les données doivent renvoyer à des fins de validation. Nous montrons ensuite que l'usage des données massives permet aux médias non seulement de personnaliser plus finement l'offre d'information, mais également de favoriser un processus d'innovation qui tend à hybrider de plus en plus les savoir-faire journalistiques avec des compétences techniques extérieures au métier. A défaut de disparaître, le métier mute en se recentrant sur certains de ses fondamentaux.

### **Mots clés**

Journalisme automatisé, data journalisme, *big data*, épistémologie, marchandisation de l'information.

## Abstract

Faced with the deluge of massive amounts of data (*big data*) today's journalism wonders about its practices, and also about the economic models of the media and the organization of editorial teams. The issues are multifaceted. At first we emphasize the epistemological reluctance of the profession to the promises of totally automated journalism. Actually data journalism is always linked with an identified ground to which data must refer in order to validate it. Then we show that the massive use of data enables the media not only to personalize their offer of information more precisely, but also to promote an innovation process which tends to combine the know-how of journalism with technical skills unrelated to the journalism professions. Even if the profession does not disappear, it undergoes an evolution by refocusing on some of its fundamental principles.

## Key words

Automated journalism, data journalism, big data, epistemology, information commodification.

## Resumen

La invasión de los largos datos (*big data*) no sólo hace reflexionar el periodismo contemporáneo, sobre sus prácticas, sino también sobre los modelos económicos de los medios y la organización de las redacciones. En primer lugar ponemos de relieve la resistencia epistemológica de la profesión frente a las promesas de un periodismo totalmente robotizado. En efecto, siempre se refiere el periodismo de datos a un terreno identificado al que los datos deben remitir como forma de validación. Demostramos luego que el uso de los datos masivos no sólo permite a los medios de comunicación personalizar con un análisis más fino la oferta de información, sino también fomentar el proceso de innovación que tiene por objeto hibridar cada vez más los conocimientos periodísticos con conocimientos especializados externos a la profesión. Si no desaparece, muta la profesión centrándose en algunos de sus fundamentos.

## Palabras claves

Periodismo automatizado, data periodismo, largos datos, epistemología, mercantilización de la información.

## Introduction

Les médias d'information sont parmi les premiers concernés par la déferlante nouvelle de données. Celle-ci interroge le journalisme professionnel, ses pratiques, mais aussi les modèles économiques des médias et l'organisation des rédactions. Il n'y a alors qu'un pas pour faire des *big data* le levier d'une transformation en profondeur du journalisme.

En l'absence de définition partagée de la notion, nous retiendrons ici par *big data* l'ensemble des données qui relèvent de la règle des trois V souvent invoquée (le volume, la variété, la vélocité dans la récolte), avec comme discours d'accompagnement l'idée que les *big data* autorisent une forme d'intelligence nouvelle ayant l'aura de la vérité et de l'objectivité (Boyd, Crawford, 2012). Ce discours atteste de la dimension prescriptive des *big data*, Dominique Cardon proposant à cet égard d'ajouter un quatrième V à la règle de trois, à savoir les valeurs véhiculées par les données (Cardon, 2015). La

confrontation des *big data* à d'autres méthodes de saisie du réel par les données est de ce point de vue révélatrice des enjeux et des limites de la notion, ce dont atteste le rapprochement trompeur, mais souvent opéré, entre data journalisme et *big data*.

Bien loin d'être une adaptation du journalisme aux *big data*, le data journalisme rappelle que le journalisme se réfère toujours à un terrain identifié auquel la donnée doit renvoyer *in fine* afin de pouvoir être vérifiée, trahissant de ce point de vue la résistance épistémologique du journalisme face aux promesses des *big data*. Cette résistance s'accompagne toutefois d'une adaptation des rédactions. Les algorithmes associés aux masses de données peuvent être utilisés autrement, parce qu'ils produisent des signaux faibles permettant d'explorer des sujets que la rédaction aurait pu manquer, ou encore parce qu'ils permettent d'automatiser le traitement de certaines informations, libérant ainsi les journalistes de tâches à faible valeur ajoutée.

Par ailleurs, les masses de données sont de plus en plus exploitées par certaines rédactions qui les utilisent comme des leviers puissants de marketing, rendant possible une personnalisation de la publicité, voire une personnalisation de l'offre d'information. Cette dernière possibilité vient transgresser les frontières professionnelles du journalisme qui séparent la rédaction des fonctions commerciales des journaux. La personnalisation de l'offre d'information peut ainsi conduire à un journalisme tiré par la demande qui vient mettre à mal le rôle d'intermédiation des médias d'information, tout en redéfinissant leurs arbitrages éditoriaux. Enfin, l'invocation des *big data* au sein des rédactions témoigne d'un impératif d'innovation qui fait émerger des pratiques aux frontières du journalisme professionnel. Si ces frontières mouvantes ont toujours été l'enjeu d'une adaptation négociée à un nouvel environnement économique, technique et social (Ruellan, 1992), elles sont aussi, pour le journalisme, un moyen de gérer le renouvellement de ses profils en accueillant des compétences nouvelles et, pour les firmes médiatiques, un moyen de repenser leur organisation.

Nous nous proposons ici de mettre en perspective ces enjeux dans une démarche d'exploration scientifique ancrée dans les SIC, en mobilisant une approche transdisciplinaire du journalisme, sociologique et économique. En termes méthodologiques, nous nous appuyons sur une revue de la littérature scientifique, avec une cartographie des premières recherches initiées sur le rapport du journalisme aux *big data*, principalement dans un contexte anglo-saxon (Lewis, Westlund, 2015). Des témoignages d'acteurs sont également mobilisés (entretien avec Nicolas Kayser-Bril réalisé en janvier 2017 ; atelier avec David Dieudonné, Directeur du *Google News Lab France*, en mars 2017).

## Résistances épistémologiques à l'automatisation du journalisme

S'interroger sur le rôle des *big data* et des algorithmes dans le champ journalistique conduit à s'intéresser au journalisme automatisé (ou « robot-journalisme »), à savoir la production d'articles par des logiciels à partir du traitement algorithmique de grandes masses de données. Le journalisme robotisé semble en effet s'imposer comme un modèle d'« objectivité » poussant à son paroxysme l'exigence de neutralité du journaliste dans l'observation du réel grâce à la suppression de l'intervention humaine. Souvent donné en exemple, le *Los Angeles Times* publie ainsi depuis 2014 des articles sur l'activité sismique en Californie grâce au logiciel-robot *Quakebot* qui exploite les données de l'*U.S. Geological Survey*. Si la perspective de rédactions sans journalistes inquiète, ce journalisme automatisé s'accompagne aussi de promesses. Parmi ces promesses, on notera la possibilité de multiplier les articles pour répondre à des demandes de niche et pour traiter de sujets que les rédactions n'avaient pas les moyens de couvrir jusqu'alors (voir Carlson, 2015). Cette prolifération des robots-journalistes doit dès lors libérer les journalistes des tâches répétitives pour se consacrer à des enquêtes ou des projets qu'une intelligence artificielle ne pourra pas assumer.

Ces promesses font toutefois l'objet de résistances de la part des journalistes que le chercheur américain Matt Carlson a recensées dans les débats qui ont accompagné le lancement de *Narrative Science* aux Etats-Unis, une entreprise qui propose du journalisme automatisé mais qui ne se revendique pas comme une entreprise de journalisme (Carlson, 2015). Parmi ces résistances, la première d'entre elles consiste à questionner le renversement des rapports entre l'offre et la

demande dans la production de l'information. Le journalisme se définirait d'abord par l'offre, par les choix de couverture qu'une rédaction opère, c'est-à-dire la sélection de l'information qui mérite d'être portée à la connaissance du public, ce que les anglo-saxons qualifient de « *newsworthiness* ». Cette sélection de l'information est au cœur de l'autorité journalistique qui assume ici son rôle de « *gate keeping* ». *A contrario*, les tenants des *big data* vont dénoncer les insuffisances de l'intermédiation journalistique, avec ses choix humains, auxquels les algorithmes opposent une connaissance plus fine des besoins des utilisateurs, la « pertinence » ou « *relevance* » (Gillespie, 2014) de leurs réponses se substituant ici aux logiques éditoriales. De ce point de vue, les *big data* constitueraient l'aboutissement d'un processus qui, depuis les années 1990, donne à la demande une plus grande importance dans les choix éditoriaux. Mais elles en transfigurent la signification : après l'impératif publicitaire et ses *metrics* (Ouakrat, 2012), la demande s'imposerait au nom de la *relevance* des informations sélectionnées, c'est-à-dire au nom d'une revendication épistémologique qui ferait des algorithmes le vecteur d'une nouvelle science des publics et de leurs attentes.

Les divergences portent également sur le statut des faits. Ces derniers, parce qu'ils s'inscrivent dans une réalité complexe, doivent de plus en plus être contextualisés et interprétés, ce qui interdit de produire un compte-rendu épuré du réel, faisant de l'objectivité une norme à vocation d'abord régulatrice (Schudson, 2001). A cette interprétation journalistique des faits, suspectée de laisser place au risque de biais idéologique, le journalisme automatisé de *Narrative Science* va opposer la possibilité de neutraliser toute forme de subjectivité dans le traitement de l'information grâce à l'application de processus normés. A l'évidence, des tensions épistémologiques se manifestent entre, d'une part, l'expertise journalistique et son rapport négocié au réel et, d'autre part, la prétention des algorithmes à faire émerger de manière normée des faits à partir du traitement de gigantesques bases de données.

Afin d'explorer ces tensions, nous nous proposons d'interroger ici le rapport du data journalisme aux données parce que cette nouvelle manière de faire du journalisme, qui émerge à la fin des années 2000, revendique justement la possibilité, à partir des données, de produire des contenus qui s'inscrivent au cœur du champ journalistique (Croissant, Touboul, 2013), sans lui être extérieur comme l'est le journalisme robotisé de *Narrative Science*. La démarche initiée par le data journalisme, ou journalisme de données, peut alors se comprendre comme une manière d'appriivoiser la profusion nouvelle de données.

Le data journalisme accorde une importance majeure aux données et à leur exploitation, prenant ainsi ses distances avec le « terrain », au moins en première intention. Il a témoigné, aux États-Unis, d'une évolution du journalisme qui, après avoir favorisé l'observation sur le terrain et les interviews dès le 19<sup>ème</sup> siècle, sur le modèle des sciences sociales, revient aux documents, donc aux données et à leurs archives, pour explorer des tendances sur le temps long (Anderson, 2015). Ce retour aux documents ne signifie pas pour autant une prise de distance à l'égard des canons scientifiques des sciences sociales, qui restent fortement mobilisés dans les discours des journalistes quand ces derniers cherchent à légitimer leur prétention à l'objectivité. Ainsi, l'application, dans les années 1970, des méthodes des sciences sociales au « journalisme de précision » (Meyer, 1973) imposera dans le champ journalistique le recours à l'informatique et aux statistiques comme moyen d'établir des faits. L'importance de l'informatique dans le journalisme se retrouve aujourd'hui dans le journalisme robotisé, dans le journalisme computationnel, lequel donne la priorité aux algorithmes sur les savoir-faire journalistiques, ou encore dans le data journalisme (Coddington, 2015). Ce dernier est à part ; il revendique en effet une dimension hybride ; il associe aux journalistes-rédacteurs des compétences venues du *hacking* (Dagiral, Parasié, 2011) et fait également appel aux destinataires de l'information (public) qui pourront ainsi contribuer à la qualification des données, le cas d'école restant en la matière le dispositif du *Guardian* pour les notes de frais des députés britanniques (Daniel, Flew, 2010). A cet égard, le statut de la donnée, ainsi que la manière dont celle-ci a été qualifiée, sont essentiels dans le data journalisme.

D'abord présentées comme des marqueurs de vérité parce qu'elles échapperaient à toute interprétation subjective, les données vont être ensuite repensées par le data journalisme pour mieux prendre en compte les conditions de leur établissement, leur structure, et donc aussi les possibilités de leur exploitation. Le data journalisme va alors s'engager dans une démarche constructiviste qui ne

prétend pas saisir le réel à l'état pur à travers la donnée pure (*raw data*). L'un des actes fondateurs du data journalisme est la déclaration d'Adrian Holovaty qui, dès 2006, exigeait une forme nouvelle de journalisme : « *Par exemple, supposons qu'un journal ait écrit un article sur un incendie local. [...] ce que je veux vraiment pouvoir faire, c'est explorer les faits bruts de cette histoire un par un, avec des couches d'attribution et une infrastructure permettant de comparer les détails de l'incendie avec ceux d'incendies précédents : date, heure, lieu, victimes, numéro de la caserne de pompiers, distance de la caserne, nom et nombre d'années d'expérience de chaque pompier présent sur les lieux, temps mis par les pompiers pour arriver sur place, et les incendies ultérieurs, le cas échéant.* » (Kayser-Bril et alii, 2013, p. 24). Or Nicolas Kayser-Bril, qui est à l'origine de l'édition française du *Guide du data journalisme* d'où la citation a été tirée, indique que cette vision d'Adrian Holovaty est pertinente uniquement parce qu'il s'agit dans son exemple d'une information hyper-locale. Elle ne nécessite pas de contextualisation particulière. Pour Nicolas Kayser Bril, la donnée seule n'est pas signifiante si elle est mise à disposition du public sans l'histoire qui lui donne sens.

Par ailleurs, la structure des bases de données n'est pas neutre. Là encore, les projets d'Adrian Holovaty témoignent des limites du premier data journalisme. Celui-ci a accordé une place très importante aux jeux de données pour les porter en l'état à la connaissance du public, des dispositifs visuels permettant ensuite d'opérer des tris dans ces données pour en faire émerger des informations signifiantes. Ainsi, Adrian Holovaty a-t-il lancé dès 2005 le site *chicagocrime.org* qui superpose sur une carte *Google Maps* les données en *open data* récupérées auprès du *Chicago Police Department*. Le site offre une cartographie interactive des crimes commis à Chicago. Ce site a révélé leur fréquence sur des secteurs urbains auparavant peu traités par la presse. Le data journalisme permet alors de corriger « l'inégale distribution des sources sur le territoire urbain » (Parasie, Dagiral, 2013, p. 56) tout en libérant en partie les journalistes de leur dépendance aux sources institutionnelles.

Il y a ici une croyance dans la force de la donnée brute qu'une enquête du *Los Angeles Times* va révéler comme trop naïve. En effet, le quotidien de la Cité des Anges s'interroge sur la pertinence des données mises à disposition par *everyblock.com*, un autre site d'Adrian Holovaty qui a étendu à l'ensemble du territoire américain le modèle imaginé avec *chicagocrime.org*. 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 (puisque'il s'agit de l'endroit où ils ont leurs bureaux !). Après enquête (Welsh, Smith, 2009), 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 sur *everyblock.com* et ce qui confirme la nécessité pour le data journaliste, après une phase d'exploitation des données, de retourner sur le terrain pour tester la pertinence des informations communiquées.

Cette approche s'oppose à la vision de Chris Anderson pour qui les *big data* changent en profondeur notre rapport à la vérité. Elles provoqueraient ainsi dans la science la même révolution copernicienne que celle opérée par *Google* dans la saisie statistique de la pertinence d'une page web, laquelle repose sur l'affectation de valeurs aux liens hypertextes (Cardon, 2013). Les *big data* et leur traitement mathématique rendraient caduques les anciennes méthodes scientifiques avec en ligne de mire la fin du dispositif empirique de vérification. D'où la sentence lapidaire de Chris Anderson : « *patabytes allow us to say : correlation is enough* » (Anderson, 2008). Comme dans la recherche en ligne, le critère d'efficacité du « *good enough* » l'emporte sur le test de réalité. La corrélation ne vaut pas preuve, elle « suffit ». Dès lors qu'elles échappent aux capacités de traitement humain et tendent vers l'exhaustivité, les masses de données rendraient obsolète le recours préalable au tri et au codage des informations au sein de bases structurées pour être interrogées. Or, justement, l'exemple des limites d'*everyblock.com* sur le positionnement interactif des crimes à Los Angeles montre la nécessité de maîtriser les modalités qui président à l'élaboration par le *LAPD* de la base de données des crimes à Los Angeles, sauf à énoncer des vérités qui n'en sont pas ! De ce point de vue, le data journalisme, parce qu'il maintient l'exigence de vérification, parce qu'il cherche donc à interpréter les données, et parce qu'il ne se résout pas au « *good enough* », relève d'une démarche épistémologique qui n'est pas celle des promoteurs des *big data*, au moins dans leurs promesses les plus radicales. Comme l'indique Dominique Cardon, les *big data* posent des problèmes opérationnels, plus que le « *good enough* » n'en résout, à savoir le problème des « corrélations sans causes » (Cardon, 2015, p.



21) et celui de la faiblesse du ratio « signal/bruit ». Le premier problème est un grand classique de l'analyse de données. Il ouvre la voie à une exubérance de micro-théories contingentes qui alimentent l'illusion de la fin des sciences de l'homme et de la société. Le second problème tient lui au fait que les données massives ainsi traitées offrent en fait une faible probabilité de présenter un vrai intérêt (signal) pour l'utilisateur.

Pour autant, il serait exagéré de considérer que les corrélations sans causes produites par des algorithmes opérant sur de grandes masses de données sont sans intérêt pour le data journalisme. Certes, le data journalisme favorise, comme le proclame Nicolas Kayser-Brill, les « bases de données structurées », celles qui sont questionnables parce qu'elles ont fait l'objet d'une construction maîtrisée par le journaliste, ou parce qu'elles sont suffisamment transparentes pour que le journaliste puisse exploiter les normes qui ont présidé à leur établissement. Il n'en demeure pas moins qu'il y a une utilité des données massives, celles qui excèdent par leur volume toute capacité d'appréhension et de tri, celles aussi pour lesquelles la transparence du tri algorithmique est inexistante. C'est ce que rappelle Eric Scherer quand il s'interroge sur l'utilisation des *big data* par les rédactions : « *Le tri algorithmique - qui gère les grands nombres et donc explique des choses qu'on ne peut pas voir - peut l'aider [le journaliste] à détecter des signaux faibles, à pointer vers des sujets où les rédactions restent frileuses* » (Scherer, 2015, p. 13).

### **Des contenus tirés « par la demande » qui émanent d'organisations médiatiques en voie de reconfiguration**

L'inscription du journalisme au cœur d'un écosystème informationnel dominé par les « data » ne provoque pas seulement un débat sur le rapport entre données et vérité journalistique. Les *big data* s'invitent également dans les logiques marketing des entreprises d'information. Leur premier effet structurel est d'accélérer l'assujettissement de la production éditoriale à la perception des attentes du public. Le phénomène est déjà ancien ; le poids croissant des régies publicitaires avait, dès la fin des années 1980, fait basculer le journalisme traditionnel, bâti jadis selon une logique de l'offre, dans une logique de la demande. Ce basculement est désormais accéléré et amplifié. La demande, extrêmement labile, est plus prégnante que jamais. De nouvelles pratiques journalistiques attestent ce changement de rythme, en même temps qu'elles mettent en relief ses enjeux éditoriaux et économiques.

Au début des années 1990, John Mc Manus prophétisait l'assomption du journalisme tiré par la demande, ou « *Market-Driven Journalism* » (Mc Manus, 1994). Force est aujourd'hui de constater que la pratique répandue du « *Google [ou Facebook]-journalisme* » donne un écho saisissant à cette anticipation. Certains médias - *BuzzFeed*, par exemple, ou bien le site pour adolescents *Melty* - procèdent à une veille permanente *via Google Trends* et les réseaux sociaux (*hashtags* les plus discutés) pour déterminer, en temps réel, les questions qui font le « *buzz* » et préoccupent leurs destinataires. La plateforme *CrowdTangle*, rachetée en 2016 par *Facebook*, favorise par exemple le suivi analytique des contenus diffusés sur ces réseaux. Des algorithmes prédictifs peuvent même renforcer ce travail de veille en soulignant les sujets et les angles de couverture qui pourraient devenir viraux et assurer ainsi l'audience la plus large possible. En fonction de tels signaux perçus, ces médias agencent et reconfigurent au fil de l'eau leur offre d'information.

Autres pratiques journalistiques notables, celle de « l'information-service » et celle du journalisme dit « de solutions », ou journalisme « d'impact » (Gyldensted, 2015). Pratique ancienne, constitutive même de l'information de proximité, « l'information-service » est aujourd'hui totalement renouvelée à l'aune de la massification des usages du numérique. En tant qu'information utile pour le quotidien des destinataires, issue d'une source identifiée (extérieure au média diffuseur) et transmise au public sans retouche éditoriale de la part des journalistes, « l'information-service » fait désormais l'objet d'interactions avec les médias, interactions qui favorisent l'enrichissement des bases de données commerciales de ces derniers. Ces interactions permettent d'affiner la connaissance des « destinataires-clients » (*via* leurs déclarations, le recueil de leurs intentions, l'historique de leurs comportements tracés). Elles permettent *in fine* de qualifier ces bases de données en vue d'une

exploitation commerciale par d'autres services des médias concernés. Cette transversalité entre l'information et la communication commerciale est d'autant plus efficace qu'elle repose, comme en matière de « *Google-journalisme* », sur une appréhension fine des attentes et des intentions du public. C'est un enjeu du même ordre qui est associé au journalisme dit « de solutions ». Ce type de pratique, qui connaît un fort engouement, favorise l'implication des destinataires de l'information au cœur même de la production journalistique. Ces destinataires sont non seulement des citoyens auxquels on vient proposer des « solutions » pour leur permettre d'affronter, à leur niveau, les grands défis de l'époque, mais également des pourvoyeurs de *data* monétisables au bénéfice d'activités diverses rassemblées sous une même marque-média (ou « marque-ombrelle »). Le journalisme dit « de solutions » renforce donc, au sein des médias, le tropisme en faveur du « marketing des traces » (Kessous, 2011) et de l'approfondissement de la relation de personnalisation autour de l'information produite.

L'accélération de ce basculement d'une logique éditoriale de l'offre vers une logique centrée désormais sur la demande a non seulement des effets sur les contenus produits, mais aussi sur l'économie des médias visés. Concernant la nature de l'information produite, l'effet collatéral est bien sûr un effet d'auto-renforcement. Une demande explicite suscite une offre qui, *a posteriori*, légitime et étaye cette demande au risque d'ailleurs d'éclipser d'autres sujets ou d'autres thématiques. La généralisation de telles pratiques « pro-demande » peut donc catalyser un appauvrissement qualitatif de cette partie de l'écosystème informationnel qui ne dépend que des recettes publicitaires. Cette dépendance exclusive oblige en effet les offreurs d'information à courtiser le même segment (large) du grand public et à proposer les mêmes sujets plébiscités, fédérateurs et non-clivants. On retrouve ici un résultat cohérent avec l'ancienne « loi de Steiner » (Bassoni, Joux, 2014, p. 86), loi selon laquelle la pluralité des offreurs ne rime pas forcément avec la diversité des contenus proposés au public.

Concernant l'économie des médias visés, ce basculement en faveur de la demande du public a des effets nombreux et structurants. Il catalyse en fait un reformatage en profondeur du modèle d'affaires de ces médias. Le premier effet à signaler a trait à la gestion de la manne publicitaire. Si les algorithmes permettent d'affiner la connaissance de l'audience, de ses caractéristiques et de ses attentes, ils renforcent par la même occasion la rentabilité liée à la monétisation de cette audience auprès des annonceurs. On appelle précisément publicité « programmatique » cette communication adaptée au public-cible et adossée à un traitement de données massives. Depuis 2016, celle-ci représente déjà plus de la moitié des investissements publicitaires en ligne en Europe (IAB, 2017). Les médias, légitimement soucieux de leur pérennité et de leur autonomie financière, doivent donc, plus que jamais, prendre en compte cette dimension économique cardinale de leur environnement. Cet enjeu éclaire d'ailleurs les ressorts de la bataille qui se joue actuellement entre les médias traditionnels et les GAFAs au sujet du partage du « gâteau » publicitaire. On le sait, ces médias se plaignent des modalités d'un partage qu'ils jugent déséquilibré (Guibert et *alii*, 2016, p. 162). Mais bien au-delà des revenus à rétrocéder, le sujet le plus sensible concerne la maîtrise pleine et entière des données-clients. Avec son format *Instant Articles*, Facebook – nonobstant ses dénégations – n'a pas tout à fait renoncé à exploiter pour son propre compte les données relatives aux médias partenaires de la plateforme. Les projets d'alliances dans le domaine de la publicité numérique qui fleurissent actuellement entre médias traditionnels (en France, *Gravity* et *Skyline*) doivent être appréciés à l'aune de ce bras de fer ; ils marquent, de la part des médias, une vraie volonté de s'autonomiser vis-à-vis des poids lourds de l'économie numérique.

Un second effet économique est à signaler ; il concerne la tendance à personnaliser de plus en plus finement l'offre d'information et à favoriser une forte discrimination tarifaire. En tenant compte des centres d'intérêt de tout un chacun, de ses échanges personnels et des interactions qu'il alimente *via* les réseaux sociaux, les algorithmes peuvent éditorialiser/personnaliser à l'infini l'information qui lui est dédiée. Comme l'a prétendu Mark Zuckerberg, le 6 novembre 2014, le but est « *de bâtir le parfait journal personnalisé de tout le monde* », c'est-à-dire un quotidien sur mesure dédié à chaque internaute. L'enjeu commercial est double ici : il s'agit tout à la fois de fidéliser celui-ci et de l'enfermer au sein d'un écosystème cohérent et de l'exposer à une publicité ciblée toujours plus efficace. En individualisant ainsi à l'extrême la relation-client, la personnalisation des contenus autorise une discrimination tarifaire poussée. Comme l'indique l'économiste Jean-Pascal Gayant, « ce

que nous connaissons déjà pour le transport aérien ou ferroviaire a vocation à se développer pour les moindres objets et services » (Gayant, 2015), y compris les services informationnels. Cette possibilité, si elle est exploitée, ouvre donc la voie à une grande pluralité de modèles d'affaires au sein de l'univers médiatique. Subséquemment, les journalistes et les rédactions devront faire montre d'une grande adaptabilité aux contextes divers au sein desquels ils seront amenés à travailler. Cette adaptabilité va bien sûr de pair avec la reconfiguration des organisations médiatiques.

En matière de reconfiguration organisationnelle, deux niveaux différents doivent être distingués. Le premier concerne la rédaction *stricto sensu* ; le second a trait à la firme médiatique *lato sensu*. Au niveau de chaque rédaction, l'adaptabilité va dépendre de la qualité du mélange des différentes « grammaires » professionnelles désormais mobilisées, ainsi que de la profondeur de l'acculturation des journalistes aux facettes marketing de leur activité. Au sein de chaque rédaction, l'éventail des compétences mobilisées va devoir s'ouvrir comme jamais (Desbordes, 2018, p. 124). A chaque compétence sont associés une « grammaire », des techniques et des savoir-faire. Les rédacteurs vont devoir interagir avec des spécialistes de design graphique (visualisation des données), des *community managers*, des *data miners*, des *data analysts*, des spécialistes d'ergonomie cognitive, ... et partager avec ces différents professionnels des connaissances, des diagnostics et des outils. L'adaptabilité sera donc affaire de polyvalence, de « langue commune » et de croisement de plus en plus marqué entre des univers techniques jadis dissociés. Bien au-delà de la simple problématique RH (gestion des compétences et des qualifications), cette nécessaire « hybridation » soulève la question de la formation au journalisme (qu'elle soit initiale ou continue). En France où les canons de la profession demeurent encore polarisés autour de référentiels foncièrement littéraires, une telle mutation ne va pas de soi et pose un défi quasi-culturel.

Autre défi de taille, celui de l'acculturation des journalistes aux facettes marketing de leur activité. Comme l'a bien montré Pauline Amiel dans son étude récente des journalistes « localiers » de la presse quotidienne régionale française, cette acculturation se heurte à un conflit d'identité. A l'heure de la dématérialisation de la presse locale et de la mutation de son modèle économique, ces « localiers » sont tiraillés entre une identité professionnelle traditionnelle (« fermée », c'est-à-dire bâtie en opposition avec les canons de la communication) et une identité moderne plus ouverte, plus mouvante, assise sur une « intégration du vocabulaire et des enjeux » (Amiel, 2017, p. 221) du marketing. Ces professionnels vivent souvent douloureusement ce qu'ils assimilent à une « dilution [...] de l'identité énonciative » (*id.*, p. 302) des médias d'information. Ce ressenti collectif peut à l'évidence freiner l'adaptation au nouveau contexte précédemment évoqué.

La reconfiguration organisationnelle ne s'arrête bien sûr pas aux portes de la rédaction ; elle impacte la firme médiatique *lato sensu*. Les médias, en tant qu'organisations économiques, seront soumis au décloisonnement interne. En mai 2014, le bilan « *Innovation Report* » qui tente de souligner les points faibles de la stratégie numérique du *New York Times* pointe, entre autres choses, la nécessité d'une coopération, ou d'un maillage plus étroit, entre la rédaction du journal, son service marketing et le service « expérience-client » (lequel rassemble, à l'époque, 30 designers numériques, 30 analystes de données, 120 chargés du développement-produit et 445 informaticiens ingénieurs ! (Fradin, 2016, p. 85)). En France, les expériences de « labs », mises en place récemment au sein de certains médias (*Le Parisien*, par exemple), participent de cette même logique de décloisonnement. Une dimension « *bottom-up* » est souvent mise en exergue, laquelle – en permettant une vraie implication des journalistes dans la recherche de solutions nouvelles – facilite l'acceptation, par ces derniers, du débordement des frontières traditionnelles de leur métier. L'enjeu de cette acculturation n'est pas mince : il s'agit de catalyser, au sein des rédactions, un processus permanent d'innovations (nouveaux formats éditoriaux, nouvelles interfaces, ...), seules susceptibles de concilier *in fine* les attentes labiles des consommateurs, par ailleurs « citoyens-internautes », avec les savoir-faire et les ambitions légitimes des professionnels de l'information ; comme si la révolution des « data » permettait aux médias de densifier les liens, jadis distendus, avec leurs publics.



## Conclusion

Dès lors que l'on s'interroge sur les impacts du traitement algorithmique des données massives sur l'univers du journalisme professionnel, force est de constater la polysémie qui auréole la notion de *big data*. De quelles données parle-t-on ? De celles qui traduisent un nouveau rapport au terrain d'investigation (les sources), ou bien de celles qui révèlent les caractéristiques comportementales des destinataires de l'information ? ... Ce distinguo est important, comme nous avons tenté d'ailleurs de l'établir. Cette polysémie est d'autant plus significative qu'elle ouvre la voie, du fait des différentes acceptions qu'elle véhicule, à des visions très contrastées de l'avenir du journalisme professionnel. Une vision ultra-techniciste, adossée à la « science » des *big data*, alimenterait pour sûr la prophétie de la disparition du métier ; à l'aune d'une telle vision, une production automatisée d'information, tirée en temps réel par la demande, se substituerait avantageusement à la médiation humaine constitutive de l'exercice de la profession. A cette vision peut être opposée la vision plus nuancée que nous formulons ici. Loin de « tuer » le métier, la déferlante des *big data* va sans doute, et de façon presque paradoxale, contribuer à sa mutation en renforçant deux de ses caractéristiques fondamentales, à savoir le tropisme en faveur du terrain ainsi que l'approfondissement des liens avec le public.

## Références bibliographiques

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.

Anderson Chris (2008), "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", *Wired*, June 23rd (consulté le 13 février 2018, <http://www.wired.com/2008/06/pb-theory/>).

Anderson Christopher W. (2015), "Between the unique and the pattern. Historical tensions in our understanding of quantitative journalism", *Digital journalism*, n°3, p. 349-363.

Bassoni Marc, Joux Alexandre (2014), *Introduction à l'économie des médias*, Paris : Armand Colin (Cursus).

Boyd Danah, Crawford Kate (2012) "Critical questions for big data", *Information, Communication & Society*, 15/5, p. 662-679.

Cardon Dominique (2013), « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux* 2013/1, n° 177 : 63-95

Cardon Dominique (2015), *A quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris : Seuil (La République des idées).

Carlson Matt (2015), "The Robotic Reporter", *Digital Journalism*, n°3, p. 416-431.

Coddington Mark (2015), "Clarifying journalism's Quantitative Turn", *Digital Journalism*, n°3, p. 331-348.

Croissant Valérie, Touboul Annelise (2013), « Le *datajournalism* par ses acteurs, autodéfinition d'une pratique émergente en France », *Recherches en communication*, 40, p. 133-149.

Dagiral Eric, Parasio Sylvain (2011), « Portrait du journaliste en programmeur : l'émergence d'une figure du journaliste "hacker" », *Les Cahiers du journalisme*, n° 22/23, p. 144-154.

- Daniel Anna, Flew Terry (2010), "The Guardian Reportage of the UK MP Expenses Scandal: a Case Study of Computational Journalism", *Communications Policy & Research Forum*, Sydney, November 15-16<sup>th</sup>.
- Desbordes Damien (2018), *Les robots vont-ils remplacer les journalistes ?*, Paris : Plein Jour.
- Fradin Selma (2016), *Les nouveaux business models des médias. Les trois piliers de la transformation*, Fyp (Entreprendre).
- Gayant Jean-Pascal (2015), « Le big data profite-t-il vraiment au consommateur ? », *Le Monde*, 17 janvier.
- Gillespie Tarleton (2014), "The relevance of Algorithms" (p. 167-194), in *Media Technologies: Essays on Communication, Materiality and Society* (Gillespie and alii dir.), Cambridge (MA): MIT Press.
- Guibert G r me, Rebillard Franck, Rochelandet Fabrice (2016), *M dias, culture et num rique. Approches socio conomiques*, Paris : Armand Colin (Cursus).
- Gyldensted Cathrine (2015), *From Mirrors to Movers: Five Elements of Positive Psychology in Constructive Journalism*, Seattle : CreateSpace Independent Publishing Platform.
- IAB (Europe) (2017), *European programmatic market sizing 2016*, september (consult  le 13 f vrier 2018, <https://www.iabeurope.eu/research-thought-leadership/programmatic/iab-europe-report-european-programmatic-market-sizing-2016/>).
- Kayser-Bril Nicolas, Gray Jonathan, Bounegru Liliana, Chambers Lucy (dir.) (2013), *Guide du datajournalisme : collecter, analyser et visualiser les donn es*, Paris : Eyrolles.
- Kessous Emmanuel (2011), « L' conomie de l'attention et le marketing des traces », *Communication au colloque ACFAS Web social, communaut s virtuelles et consommation*, UQAM, Sherbrooke, 11 mai.
- Lewis Seth C., Westlund Oscar (2015), "Big data and journalism. Epistemology, expertise, economics and ethics", *Digital journalism*, 3, p. 447-466.
- McManus John (1994), *Market-Driven Journalism: Let the Citizen Beware?*, London : Sage Publications.
- Meyer Philip (1973), *Precision journalism: a reporter's introduction to social science methods*, Bloomington, Indiana University Press.
- Ouakrat Alan (2012), « Le ciblage comportemental, une perte du contr le des  diteurs sur les donn es de l'audience », *Tic&soci t s* 6(1) : 33-55
- Parasie Sylvain, Dagiral Eric (2013), « Des journalistes enfin lib r s de leurs sources ? Promesses et "r alit s du journalisme de donn es" », *Sur le journalisme* [en ligne], Vol. 2, n 1, p. 52-63.
- Ruellan Denis (1992), « Le professionnalisme du flou », *R seaux*, Vol. 10, n  51, p. 25-37.
- Scherer Eric (2015) *Logiciels, algorithmes, robots : journalisme automatique*, m ta-media#8, Paris, France T l visions
- Schudson Michael (2001), "The objectivity norm in American journalism", *Journalism*, 2/2, p. 149-170.
- Welsh Ben, Smith Doug (2009), "Highest crime rate in L.A.? No, just an LAPD map glitch", *latimes.com*, April 5<sup>th</sup> (consult  le 8 f vrier 2018, <http://www.latimes.com/local/la-me-geocoding-errors5-2009apr05-story.html>).