

Date : 21/03/11

## Caroline Goulard : Data-expert



En complément de notre entretien-fleuve avec Nicolas Kayser-Bril (Owni) paru dans Chronic'art #70, en kiosque, voici la version intégrale de notre interview avec Caroline Goulard, data-journaliste et co-fondatrice de Dataveyes, une start-up de visualisations interactives de données, et membre d'Actuvisu, un site de visualisation de l'actualité via des infographies interactives et des données parlantes.

Chronic'art : Le journalisme tel que nous l'avons connu jusqu'à présent est-il mort ?

Caroline Goulard : Oui et non. Le journalisme n'est pas une discipline bien instituée, dont on pourrait tracer des contours propres et déterminer simplement ce qui en est et ce qui n'en est pas. Les pratiques, les normes et les modes y ont toujours été fluctuantes. Le journalisme, c'est le « professionnalisme du flou » pour reprendre le titre d'un ouvrage de Denis Ruellan (*Le Journalisme, ou le professionnalisme du flou*, aux **Presses Universitaires de Grenoble**, 2007). Ce qui est sûr c'est que, d'une part, le journalisme tel que nous l'avons connu jusqu'à présent ne répond plus toujours aux attentes de son public, et, d'autre part, qu'il est très mal adapté à l'économie numérique et aux logiques de fonctionnement du Web. Le journalisme doit faire face à de lourdes évolutions des pratiques de consommation de l'information : individualisation, personnalisation, fragmentation, désintérmédiation, etc. Les internautes veulent avoir le choix et garder la main sur ce qu'ils absorbent, ils ne supportent pas d'être inactifs face à un écran, ils veulent participer, jouer, partager, être amusés, se laisser surprendre, etc. Les journalistes doivent aussi se confronter à de nouvelles règles du jeu : sur le web, gérer l'information, c'est gérer la surabondance, l'immédiateté comme la mémoire infinie du numérique, l'absence de barrières à la propagation mais aussi la nature réticulaire de la Toile, la dynamique des liens, et l'économie de la participation. Parmi toutes

## Évaluation du site

Chronic'art est un webzine culturel francophone particulièrement riche en rubriques : événement, cinéma, musique, cyber, livres, sortir, le mag, portrait, forum, actu, et infos en tous genres. Un ton intelligent qui ne manque pas d'humour.

**Cible**  
Spécialisée

**Dynamisme\*** : 1

\* pages nouvelles en moyenne sur une semaine

ces tendances, il me semble que la plus fondamentale réside dans la richesse des données qui inondent désormais le Web : comment les faire parler ? Comment en extraire du sens ? Comment les représenter ? Comment leur donner une valeur éditoriale ? C'est, à mes yeux, le plus gros défi journalistique pour les prochaines années, car les données vont devenir une façon incontournable d'appréhender et de comprendre le monde. Le journalisme de l'ère analogique continuera, je pense, à exister, mais dans certains espaces et sur certains supports. Il ne peut plus suffire à répondre aux besoins d'information.

Quelle est votre définition du « data-journalisme » ?

Le journalisme de données consiste à exploiter des bases de données volumineuses pour en extraire de l'information et la présenter de façon engageante au public. Il invente une nouvelle grammaire de la représentation de l'information, qui s'adresse à l'intelligence visuelle du public plus qu'à son intelligence verbale. D'un côté, pour les journalistes, il s'agit d'une nouvelle façon de chercher de l'information, en étudiant des sources de données jusque là mal exploitées. Par exemple les épais rapports du FMI, de la Banque mondiale, de la CNCCFP, et., ou encore en collectant et analysant des données en provenances des réseaux sociaux pour décrypter des tendances, ou bien en collectant des données issues d'un vaste corpus de documents. De l'autre, pour le public, il s'agit d'un nouveau mode d'engagement dans l'information. Le journalisme de données propose d'expliquer ce que révèlent les données grâce à des applications graphiques et interactives. En manipulant lui même les données, en jouant des curseurs, des zooms ou des filtres, l'internaute s'approprie l'information et la mémorise bien mieux car il est placé dans une posture active. Concrètement, cela revient à faire communiquer des lourdes bases de données avec des interfaces d'analyse et d'exploration. Ainsi, le journaliste de données rend trois types de services à son public : il apporte de l'information jusqu'ici mal mise en valeur, celle issue des données, il lui donne des outils pour explorer cette information, et enfin, il lui présente des représentations pour dégager du sens de cette information. D'un point de vue plus organisationnel, cela suppose de faire travailler ensemble, du début jusqu'à la fin, des développeurs axés data, des journalistes, des statisticiens, et des designers de l'information. Il s'agit d'une façon de travailler qui ne peut pas être linéaire, mais qui doit être collective, gérée en « mode projet ». La démarche est la même que pour n'importe quel autre traitement journalistique : partir d'un sujet, d'un angle, respecter les mêmes procédures déontologique de vérification des sources et d'interrogation sur la nécessité de publication, etc. seulement la façon de travailler sur le sujet change : il s'agit d'abord de collecter les données, de les nettoyer, de les recouper et les vérifier, de les traiter, des les analyser, puis de les « scénariser », en construisant des applications interactives et graphiques qui racontent des histoires à partir des données tout en travaillant sur des expériences utilisateur innovantes.

Le data-journalisme peut-il, selon vous, relever le niveau et comment ?

Oui, car le journalisme de données est parfois mieux adapté pour traiter de certains sujets que d'autres formes journalistiques, et parce qu'il apporte de réelles innovations dans les pratiques journalistiques. Faire travailler ensemble développeurs, journalistes et graphistes n'a rien d'évident dans les rédactions, et pourtant c'est ainsi qu'il est possible de redonner de la valeur ajoutée aux contenus informationnels. Car le journalisme de données apporte une nouvelle

façon d'enquêter, une nouvelle façon d'engager le public dans l'information, une nouvelle façon de raconter le monde, une nouvelle expérience d'information propre à générer de la sérendipité, à susciter la curiosité, à marquer émotionnellement et intellectuellement le public. Finalement le journalisme de données amène à se poser les bonnes questions : comment répondre aux besoins de mon public ? et comment se donner les moyens sur le long terme d'appréhender la complexité du monde ? Le journaliste de données exerce ainsi une médiation visuelle, cinématique et esthétique entre des données complexes et peu parlantes en elles même, et la compréhension du public. Les journalistes peuvent ainsi retrouver un rôle de médiateur qu'ils avaient perdu depuis que sur Internet, les internautes peuvent directement accéder aux sources d'information et inversement. Lorsqu'il s'agit de données, un internaute moyen n'a pas les compétences pour les extraire et les analyser. Peut-être que cela viendra, mais pour le moment, les journalistes peuvent jouer ici un rôle important de décrypteur. Enfin, le journalisme de données réinstalle de la confiance avec le public, et ce point me semble important, en donnant accès aux données, il favorise une plus grande transparence dans l'exercice du journalisme. Ça me semble important aujourd'hui pour retrouver un rapport de confiance avec le public.

Pouvez-vous citer un exemple à suivre de média qui va de l'avant, dans ce sens ?

Les deux pionniers du journalisme de données sont le Guardian au Royaume Uni et le New York Times aux Etats-Unis. Le Guardian concentre ses activités de journalisme de données dans le DataBlog. Presque quotidiennement, le DataBlog met à disposition des internautes de nouvelles bases de données. L'internaute peut télécharger les bases de données sous forme de Google Document. Ces bases de données sont accompagnées d'un article, et souvent, d'un petit graphique interactif réalisé avec ManyEyes ou Timetric. Le DataBlog invite toujours ses internautes à s'emparer des données pour proposer leurs propres visualisations ou applications. Le Guardian réalise aussi de gros projets, mais de façon plus épisodique : des enquêtes menées grâce aux données - comme celle qu'il a menée sur les notes de frais des députés britanniques - ou des réalisations ambitieuses de visualisation de données - comme les cartes réalisées à partir des documents sur la guerre en Afghanistan et publiées par le site Wikileaks. Le New York Times publie lui aussi des travaux de visualisation de données conséquents. Ses infographies interactives sont toujours très ambitieuses d'un triple point de vue technique, graphique et journalistique. Elles nécessitent de mobiliser des équipes de journalistes, d'infographistes et de développeurs pour chaque réalisation. Le New York Times ne présente pas de données brutes, mais des applications pensées pour être au service des internautes : cartes explorables, timelines enrichies, interfaces de recherche dans les bases de données, visualisations innovantes, etc. Ces visualisations de données sont pensées comme un contenu rich media : de même qu'un diaporama photo ou un montage vidéo, elles sont un outil de traitement de l'information. Le Guardian et le New York Times ne considèrent pas les données seulement comme un mode de traitement de l'actualité parmi d'autres, les deux titres en ont fait un atout stratégique et une part importante de leur image de marque en ligne. Ainsi le Guardian a été le leader de la campagne pour l'ouverture des données publiques au Royaume-Uni. Le site d'information a exercé un important lobbying pour l'ouverture des données du cadastre en publiant tous les jours et pendant plusieurs semaines des articles sur l'enjeu de l'opendata pour l'innovation économique et sociale. La campagne « Free our data » a eu un fort écho, et elle a débouché sur la mise à disposition des données publiques par l'administration en charge du cadastre. Le New York Times s'est lui aussi positionné sur la mise à disposition



de données au public, mais en se centrant sur les problématiques propres au web de données. Le New York Times a réalisé plusieurs API, qui délivrent aux internautes des données issues du travail des journalistes ou du fonctionnement du site du New York Times. Les API facilitent le mashup des données et leur réutilisation pour créer des applications. De même, le New York Times a reversé au sein de la communauté du Web sémantique le thésaurus à partir duquel sont taggés tous les contenus du site et du journal depuis sa création. En se positionnant fortement sur les enjeux des données liées, le New York Times séduit la communauté des développeurs, et se donne les moyens de traiter des données de plus en plus complexes. Grâce à ses API, le New York Times n'est plus uniquement un producteur d'informations, il est devenu un producteur de données.

Propos recueillis par Cyril De Graeve