

Comment devenir un data-journaliste ?

Le journalisme de données est énorme. Je n'emploie pas « énorme » en guise d'expression à la mode - bien qu'il le soit devenu ces derniers mois - mais « énorme » pour connoter son immensité. Il représente la convergence d'un certain nombre de domaines qui sont significatifs à part entière : investigation, statistique, conception et programmation. L'idée de combiner ces compétences pour raconter des histoires importantes est séduisante, mais aussi intimidante. Qui peut faire tout ce cela? En réalité presque personne ne fait tout cela, mais il y a assez de parties différentes dans ce puzzle pour que les gens s'y impliquent facilement et débiter de là. Selon moi, ces parties consistent en quatre points :

1. La découverte de données

La découverte de données peut impliquer bon nombre de choses : une connaissance experte d'un sujet, des contacts, la capacité à utiliser un ordinateur équipé de compétences pour le recueil de données ou encore, pour certains, des compétences techniques spécifiques avec MySQL ou Python qui rassemblent les données pour vous.

2. Interroger les données

Interroger les données signifie que vous devez avoir une bonne compréhension du jargon et du contexte plus large dans lequel les données sont englobées, de même que pour les statistiques. Être familier avec les tableurs peut aider à gagner beaucoup de temps.

3. Visualisation de données

Visualiser et nettoyer les données était historiquement la responsabilité des concepteurs et des codeurs, mais un nombre croissant de rédacteurs s'y essaient. En partie grâce à une conscience élargie de ce qui est possible, mais aussi grâce à une baisse des barrières en ce qui concerne l'expérimentation avec les données.

4. Agréger les données

Grâce à leurs possibilités, des outils comme ManyEyes pour la visualisation et Yahoo!Pipes pour l'agrégation m'ont permis d'accrocher rapidement des étudiants en journalisme. Beaucoup attrapent le virus de journalisme de données peu après.

Comment commencer ?

Par où un journaliste de données débutant doit-il commencer ? Une réponse évidente serait "*avec les données*" - mais une deuxième réponse est également possible : "*avec une question*". Les journalistes doivent jouer leur rôle. Répondre à l'actualité et aux événements en cherchant activement des sujets et des angles - et c'est pareil pour les données. Aron Pilhofer du *New York Times* recommande que vous "*commencez petit et commencez par quelque chose que vous connaissez déjà et faites déjà. Et toujours, toujours, se rappeler que le but est ici le journalisme.*" Charles Arthur du *Guardian*, suggère "*de trouver une histoire qui sera meilleure racontée à travers des nombres*", tandis que Jonathan Richards du *Times* et Conrad Quilty-Harper du *Télégraphe* recommandent tous les deux de trouver ses marques et des idées en suivant les *blogs in the field* et en assistant à des rencontres telles que Hacks/Hackers.

Il n'y a aucune sorte de données publiées sur lesquelles vous ne pouvez vous faire vos dents journalistiques. Le mouvement de données libres au Royaume-Uni et maintenant à l'international engendre une sortie continue de données valant la peine d'être publiées. Aujourd'hui, il est relativement facile de trouver des ensembles de données étant publiés par des régulateurs, des groupes de consommateurs, de charité, des institutions scientifiques et de business. Vous pouvez

aussi contrôler les réponses des organismes comme *Freedom of Information* et *What Do They Know* ainsi que les publications internes d'organisations. Et bien sûr, il y a le datablog du *Guardian*.

Une deuxième approche, cependant, est de commencer par une question. Un exemple : "*Des radars de vitesse coûtent ou économisent de l'argent ?*" Voilà une question d'actualité que l'on a récemment demandée sur *Help Me Investigate*, le site de journalisme d'investigation crowdsourcing que j'ai mis en place. Nous avons ensuite cherché les données qui pourraient y répondre (pour le moment cela est venu d'une revue gouvernementale et un rapport DfT). Soumettre une requête à Freedom of Information est aussi une voie utile (assurez-vous que vous demandez les données en CSV ou dans un format semblable).

Peu importe l'approche que vous prenez, il est probable que le travail réel se situera dans la découverte des nouveaux morceaux d'informations et de données pour remplir l'image que vous essayez de clarifier. Les données gouvernementales, par exemple, viendront souvent encombrer avec un jargon et des codes que vous devrez comprendre. Un appel à l'organisation appropriée peut apporter un peu de lumière. Si cela prend trop longtemps, une recherche avancée pour un des codes les plus obscurs peut aider aussi - limiter votre recherche, par exemple, en incluant à la fin « `site:gov.uk filetype:pdf` » (ou des limitations équivalentes pour votre recherche particulière).

Vous aurez aussi besoin de contextualiser des données initiales avec de nouvelles données. Dans le cas d'informations sur le changement de la masse salariale d'un département gouvernemental, par exemple : la main-d'œuvre de département s'est-elle étendue ? Quelle est la comparaison avec d'autres départements gouvernementaux ? Qu'en est-il de salaires plus larges dans l'industrie ? Qu'en est-il de l'inflation et changements du coût de la vie ? Ce contexte peut faire une différence entre manquer et découvrir une histoire.

Assez souvent, vos données auront besoin d'être nettoyées : cherchez des noms différents qui désignent la même chose, les erreurs de ponctuation et d'orthographe, des champs mal renseignés (par exemple les dates qui sont formatées comme le texte), des données inexactly entrées et des informations qui manquent entièrement. Des outils comme Freebase Gridworks peuvent aider ici.

À d'autres moments, l'ensemble de données dont vous avez besoin sera dans un format inopportun, comme un PDF, Powerpoint ou même une page web mal fichue. Si vous avez de la chance, vous pourrez copier-coller les données dans un tableur. Mais vous n'aurez pas toujours de la chance. Dans ces moments, un peu de connaissance en programmation arrive à point nommé. Il y a deux écoles : d'un côté ceux qui peuvent écrire des scénarios à partir de rien, comme un extrait d'une page Web, et qui stockent les informations dans un tableur. Les autres peuvent utiliser un site web comme *Scraperwiki* qui a déjà des scénarios d'exemple que vous pouvez personnaliser à vos fins - et une communauté pour aider. Il y a aussi des outils en ligne comme *Yahoo!Pipes* et *the Firefox plugin OutWit Hub*. Si les données sont dans un tableau HTML, vous pouvez même écrire une formulaire en ligne dans un tableur Google pour les mettre dedans. Vous pouvez aussi juste le consigner à la main - mais quoi que vous faites, assurez-vous que vous publiez en ligne votre tableur et bloguer à son propos, de manière à ce que d'autres n'aient pas à répéter votre travail acharné.

Une fois que vous avez les données vous devez raconter l'histoire, vous devez l'obtenir prête à être visualiser. Taillez tout ce qui est périphérique à vos besoins dans le but de visualiser votre histoire. Il y a des douzaines d'outils en ligne libres que vous pouvez utiliser pour ceci. ManyEyes et Tableau Public sont de bons sites pour commencer les diagrammes. Explorez. Si vous vous débrouillez avec l'aspect graphique, essayez de faire la visualisation la plus claire possible grâce aux couleurs et l'étiquetage. Et incluez toujours un texte donnant un lien aux données et leur source - car les

infographies ont tendance à être séparées de leur contexte original au fur et à mesure qu'elles font leur chemin sur le web.

Pour les cartes, le merveilleux OpenHeatMap est très facile à utiliser - tant que vos données sont classées par pays, collectivité locale, circonscription, région... Ou vous pouvez utiliser Yahoo!Pipes pour cartographier des points d'intérêt. Ces deux sites sont en réalité des exemples d'agrégation, ce qui est utile si vous aimez le concept d'agrégation et que vous voulez l'utiliser par moments. D'autres outils existent, mais si vous voulez devenir efficace en agrégation, vous devrez explorer le monde de la programmation et APIs. À ce moment-là, vous pouvez vous asseoir et penser : "le data-journalisme de données est énorme."

Et vous savez quoi ? Je vous l'avais dit.

Paul Bradshaw

Fondateur de *Help Me Investigate* et *Reader in Online Journalism*, Birmingham City University. Il enseigne à Université de Londres et publie le *Online journalism blog*.