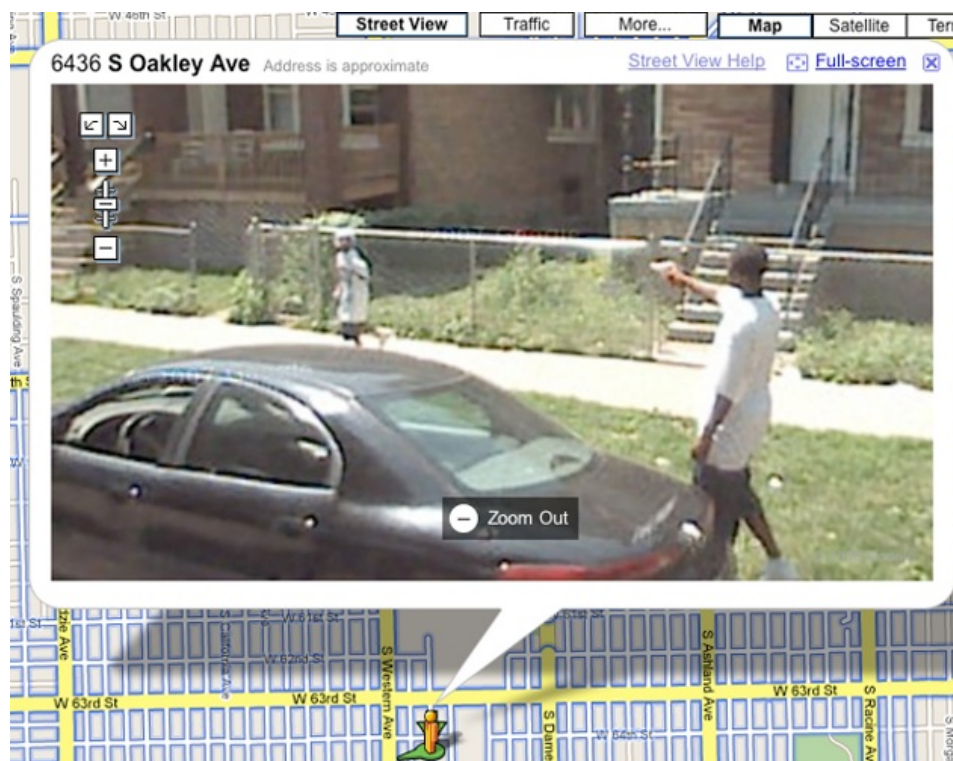


CHICAGO: LE PARRAIN DU DATAJOURNALISME?

LE 20 JUIN 2011 MAEL INIZAN

Des premiers hackers-journalistes à la News Application Team du Chicago Tribune, le sociologue Sylvain Parasio revient sur l'essor du datajournalisme dans la ville d'Al Capone.

[Liens en anglais sauf mention contraire] Sociologues au **Latts** [fr], le Laboratoire Techniques, Territoires et Sociétés de l'Université Paris-Est Marne-la-Vallée, Sylvain Parasio et Éric Dagiral ont séjourné trois semaines à Chicago en septembre dernier, pour rencontrer les acteurs locaux du datajournalisme. Plus connue à cause d'Al Capone et des *Incorruptibles*, "la ville du crime" est l'une des villes pionnière en Amérique du Nord en matière d'utilisation des bases de données dans la production de contenu journalistique. Des premiers hackers-journalistes à la News Application Team du *Chicago Tribune*, en passant par ChicagoCrime.org : Sylvain Parasio revient pour Silicon Maniacs sur 15 ans de datajournalisme à Chicago.



Avant de nous intéresser à Chicago, peux-tu rapidement revenir sur la naissance du datajournalisme aux États-Unis ?

L'idée d'utiliser des bases de données comme support à la production d'information apparaît en Amérique du nord à la fin des années 1960, notamment sous la houlette de Philip Meyer. C'est un pionnier qui écrit dès 1971 un manuel, **The New Precision Journalism**, dans lequel il explique que les journalistes doivent utiliser des ordinateurs, mais également construire des bases de données dans un tableur pour produire l'information. Son idée est de s'inspirer des outils statistiques des sciences sociales pour produire une information en rupture avec les préjugés. Les bases de données viennent appuyer le journalisme d'investigation. En 1967, il réalise ainsi une première enquête sur des émeutes raciales dans la ville de Détroit. En utilisant un ordinateur IBM disponible à Harvard, il montre, chiffres à l'appui, que contrairement à ce qu'on pensait à l'époque les Afro-américains étudiant à l'université avaient autant de chances que les autres de participer aux émeutes.

Comment ces méthodes d'investigation vont-elle se généraliser ?

Même si le manuel de Philip Meyer remporte un certain succès, ce modèle de journalisme ne se diffuse pas véritablement avant la deuxième moitié des années 1980. Plusieurs journalistes remportent alors des prix Pulitzer avec des enquêtes qui s'appuient sur l'utilisation de bases de données. Je pense notamment à Bill Dedman, qui obtient le prix en 1989 pour une série d'articles, **The color of money**, sur les questions de discriminations raciales pour l'obtention de prêts bancaires.

À l'époque, l'utilisation de bases de données reste encore anecdotique. Elles sont surtout utilisées par des journalistes isolés, pour faire des coups et se faire connaître. C'est seulement à partir des années 1990 que le journalisme de données change progressivement de statut. Cette évolution est portée par l'apparition d'organisations professionnelles qui essaient de commercialiser des bases de données auprès des rédactions. La plus connue d'entre-elles est la **Nicar**, qui se spécialise dans ce qu'on appelle à l'époque le CAR, le Computer Assisted Reporting. Ils se proposent de collecter des bases de données, de les compiler, et de les nettoyer pour les vendre à des journaux. Ils offrent également des formations aux journalistes. À cette époque, ce spécialiste des ordinateurs et des bases de données qu'est le Computer Assisted Reporter n'est pas vraiment reconnu dans les rédactions.

C'est à partir des années 2000 que le journalisme de données prend de l'importance...

La véritable rupture ne se fait pas avant le milieu des années 2000. Les grands titres comme le *New York Times*, le *Los Angeles Times*, le *Chicago Tribune* et le *Washington Post* ont déjà commencé à recruter des spécialistes, mais c'est la récupération de l'ensemble des données concernant la guerre, qui popularise ces méthodes. En 2006 par exemple, le *Washington Post* met en ligne l'application **The Faces of the Fallen**, qui tient la comptabilité de tous les militaires tués en Irak et en Afghanistan. C'est une grande base de données dans laquelle on peut voir le visage des militaires tombés au combat, savoir où ils sont nés, où ils ont grandi, connaître les circonstances et la date de leur mort. À la même époque le *Washington Post* publie également des données qui permettent de suivre l'activité des parlementaires américains.



Vous avez concentré votre étude sur la ville de Chicago. Retrouve-t-on ces différentes étapes dans l'essor du datajournalisme ?

Il nous est apparu qu'il y a eu véritablement trois temps dans le développement du journalisme de données à Chicago, et plus largement aux États-Unis. Un premier temps qui concerne essentiellement les organisations de presse traditionnelle, entre la fin des années 90 et le début des années 2000, avec des journalistes isolés qui utilisent des bases de données. Un second temps pendant lequel on voit apparaître de nouveaux acteurs, issus du développement web et du logiciel libre. Et enfin, un troisième temps marqué par la constitution d'équipes dédiées au sein des rédactions des grands journaux.

Qui sont ces journalistes isolés qui utilisent les premiers des bases de données ?

Nous avons rencontré **Darnell Little**, un pionnier du journalisme de données, à Chicago. C'est une personne fascinante, qui a une histoire et un parcours atypique. Ingénieur de formation, il travaille d'abord pour l'opérateur américain ATT, avant de se reconvertir dans la presse, dans le milieu des années 1990. Il obtient un Master de journalisme à la North Western University et est embauché au *Chicago Tribune* en 1996, la semaine même où le site internet du journal est lancé. Au sein de la rédaction, il sera à l'origine de grandes séries d'investigations sur des sujets qui concernent souvent la justice sociale.

Il travaille en collaboration avec les journalistes. Son travail à lui consistait à essayer de trouver les grandes lignes, les angles et des idées d'articles pertinentes, en travaillant à partir des bases de données de l'État de l'Illinois ou de la ville de Chicago. Ce sont ensuite des reporters qui vont sur le terrain pour poursuivre l'enquête. C'est une tradition que l'on connaît très peu en France : un journalisme écrit, littéraire, mais qui repose énormément sur l'exploitation statistique et de bases de données.

Les journalistes arrivent souvent avec des idées préconçues sur les gens à interviewer, sur ce qui est pertinent sur un sujet donné. Darnell Little intervenait notamment pour montrer aux journalistes qu'ils parlaient avec des idées reçues, en appuyant ses arguments sur des chiffres. Ça donnait parfois lieu à des relations un peu difficiles, d'autant plus qu'à l'époque le Computer Assisted Reporter n'est pas considéré comme une personne cool. Ce n'est pas un métier sur lequel on fantasme. C'est quelqu'un qui a des compétences extraordinaires, mais tellement particulières qu'il peut être rapidement ostracisé, isolé, en étant relégué au rang de simple technicien.

■ Comment ces data-journalistes vont-ils gagner en reconnaissance ?

À partir de 2004-2005, plusieurs initiatives émergent à Chicago. Cette fois-ci, elles ne concernent pas du tout les médias traditionnels, mais plutôt des startups, en marge du journalisme. On voit apparaître tout un ensemble de nouveaux acteurs issus du développement web : des militants pour l'ouverture des données, des informaticiens et des hackers. Ils essaient de construire quelque chose de nouveaux à l'intersection de la presse et de l'open data. Tout ces gens-là ont des pieds dans différents mondes : celui de la presse, du développement, parfois même des institutions, de la mairie. C'est notamment le cas d'Adrian Holovaty, le fondateur de Chicago Crimes en 2005, puis d'**EveryBlock**.

■ Qu'est ce que Chicago Crimes ?

En 2005, Chicago Crimes est l'un des premiers mashups qui permet de géolocaliser sur une carte les données criminelles (homicides, agression, vol...), qui sont diffusées par le site du Chicago police département depuis 1996. Les données existaient, mais elles étaient classées d'une manière qui les rendait très difficiles à comprendre et à analyser. De plus, elles n'étaient disponibles que pendant deux semaines. Adrian Holovaty va « hacker » ces données. Il les interface sur une carte Google – ce qui est déjà en soit une performance puisque l'API de Google Map n'était pas encore ouverte à cette époque -, les archive, mais crée également de nouvelles catégories, pour les rendre plus exploitables. Il était dès lors possible d'afficher les crimes par zones d'habitation. À cette époque là, c'est un gros coup.

E CHICAGOCRIME.ORG
A freely browsable database of crimes reported in Chicago.

Browse by: [Crime type](#) · [Street](#) · [Date](#) · [Police district](#) · [ZIP code](#) · [Ward](#) · [Location](#) · [Route](#) · [City map](#)
Search: [Go](#)

Crime map

68 crimes found.

Filter crimes

All locations

Burglary

1 (Central)

Jan. 1, 2007 to April 11, 2009

Midnight to Next midnigl

Update map

Crime classifications key

- Person
- Person (domestic)
- Property
- Property (domestic)
- Society
- Society (domestic)

Important disclaimer: This site is not affiliated with the Chicago Police Department. This site uses crime data obtained from the CPD's [Citizen ICAM Web site](#), which is a publicly available database of reported crime. Please read the [Citizen ICAM disclaimer](#) to understand the data fully.

[About this site](#) · [RSS](#) · [Developer services](#) · [Search](#) · [Press](#) · [Donate](#) · Powered by [Django](#)

L'image des hackers s'améliore. Ils acquièrent un rôle social beaucoup plus ouvert, avec un message politique, citoyen. Ils veulent permettre à tout le monde d'accéder facilement à l'information et aux données. Adrian Holovaty est déjà un développeur reconnu dans la communauté. Avec Chicago Crimes, il se fait une réputation non seulement à Chicago, mais également en Amérique du Nord.

Il bénéficie d'une certaine légitimité vis-à-vis de la presse puisqu'il est également journaliste...

Au départ, c'est effectivement un journaliste. Il a notamment travaillé au *Washington Post*. La légitimité de ces acteurs vient également de fondations qui les soutiennent. La **Knigh Foundation**, par exemple, a financé à hauteur de plusieurs millions de dollars Adrian Holovaty pour le développement de EveryBlock. Chicago Crimes était l'œuvre d'une personne, EveryBlock est celle d'une équipe.

Est-ce que tu peux expliquer ce qu'est EveryBlock ?

Là où Chicago Crimes reprenait simplement les données policières, EveryBlock étend le principe. On a accès à un ensemble de données et d'informations relatives à sa situation géographique. Ça concerne évidemment les crimes, les homicides, mais également, par exemple, les ouvertures de commerces, les transactions immobilières, les licences de débit d'alcool... EveryBlock agrège un ensemble d'informations disponibles en ligne et les centralise à l'échelle d'un quartier (« block »), d'une rue. C'est la même chose pour tous les avis concernant les restaurants du coin du quartier, ou des photos postées sur Flickr.

Pour ce type de site, toute la difficulté réside dans la récupération des données. Ils nous ont raconté à quel point c'était difficile, même à Chicago. Pour eux, l'enjeu est d'être en mesure d'avoir des données qui répondent à deux qualités : d'abord qu'elles soient soutenables, c'est à dire mises à jour régulièrement et toujours de même qualité, ensuite, qu'elles soient granulaires, c'est à dire qu'elles correspondent à un emplacement extrêmement précis dans un quartier, dans la ville. Contrairement à la France, ce modèle journalistique est totalement reconnu en Amérique du Nord. Le site s'est monté à Chicago et s'est rapidement étendu à d'autres villes avant d'être racheté par MSN en août 2009.

Comment ont réagi les médias traditionnels face à l'arrivée de ces nouveaux acteurs qui possèdent un savoir faire spécifique et qui les concurrencent sur un nouveaux terrain ?

En 2009 à Chicago, la situation était assez catastrophique pour la presse. Le *Chicago Tribune* est lui-même déclaré en faillite, fin 2008. Il se relève malgré tout et décide contre toute attente, mi-2009, de recruter des hackers-journalistes pour lancer une **News Application Team** de quatre personnes. Ces nouveaux journalistes viennent du monde de l'open data, du développement web et ont souvent une formation en journalisme. Ils sont recrutés pour faire des applications, de la visualisation et traiter des données. Ils doivent servir d'appui au journalisme d'investigation. Derrière cette News application Team, il y a donc un projet de journalisme ambitieux.

Les médias historiques ont une approche très différente du journalisme de données. Les journalistes traditionnels reprochaient souvent à des sites comme EveryBlock, le fait de simplement diffuser des données, sans les mettre en perspective. Dans la presse traditionnelle, on observe donc une forme de compromis : on mêle la mise à disposition des données, l'investigation et le cadrage journalistique. Le *Chicago Tribune* a par exemple réalisé une enquête sur **les maisons de retraite**, avec une série de papiers très fouillés sur les personnes victimes d'abus, mais également un moteur de recherche qui permet aux lecteurs de savoir quelles sont les caractéristiques des maisons de retraite dans lesquelles ils ont placé leurs parents ou leurs beaux parents. C'est également le même procédé qui a été employé dans une autre enquête du *Chicago Tribune* sur l'**Agent Orange** utilisé lors de la guerre du Vietnam.

The screenshot shows the 'Chicago Tribune Watchdog' website. At the top, there's a navigation bar with categories like Home, News, Business, Sports, Entertainment, Living, Opinion, Real Estate, Cars, Jobs, Deals, and Weekly Ad. Below that, there's a search bar and a section titled 'Nursing home safety' with a search tool. The search tool asks for a town, address, cross street, or nursing home name. Below the search tool, there's a section 'Recent reports' with three thumbnails: 'Compromised care', 'Drugged without cause', and 'Dangerous neighbors'. To the right, there's a section 'In case you missed it ...' with several news snippets, including 'Unearthing of Elizabeth Taylor's 3rd husband's grave still a Chicago mystery' and 'Breaking down the NCAA's Sweet 16'. At the bottom right, there's a 'CHICAGO SHOPPING' section with offers like '\$15 for \$30 of healthy Thai cuisine from Dharma Garden' and '\$15 for a tour of all your favorite Chicago movies by Chicago Film Tour (\$30 value)'. There's also a 'More than 50% off Tribune Wine Club' offer.

Dans ce contexte de crise de la presse, la datajournalisme participerait donc à une certaine forme de re-légitimation du journalisme, en appuyant l'information sur des données chiffrées réputées plus objectives et ouvertes au lecteur...

Il y a de ça, en partie. Le datajournalisme permet de développer une certaine proximité avec le lecteur. On traite l'information à l'échelle des personnes individuelles, de leur quartier, voisinage. Le datajournalisme participe à un mouvement qui voudrait qu'une des solutions à la crise de la presse soit de renouveler l'intérêt des lecteurs en parlant de chose ultra-proches de lui. Là où il faut tempérer, c'est que ces hackers-journalistes ont été recrutés pour assister les journalistes d'investigation. Ces formats sont loin d'être ceux qui font le plus d'audience. Mais, ils apportent beaucoup en terme de crédibilité et de notoriété, d'image pour le journal... C'est à la fois une composante hyperlocale, mais également une composante de prestige.

On ne peut pas s'empêcher de faire le parallèle avec la France. On assiste depuis quelques années à l'arrivée de nouveaux acteurs. Je pense par exemple au site NosDéputés.fr, dont l'objectif est de donner plus de transparence à l'activité parlementaire. Il y a également le site OWNI qui se revendique du hacking-journalisme...

Nos Députés.fr est effectivement porté par des gens qui viennent du développement web, du logiciel libre et qui ont une vraie implication citoyenne. Mais, il n'y a quasiment aucune connexion avec la presse, à part une enquête avec Médiapart et quelques partenariats très

ponctuels. On est très loin des États-Unis où c'est le *Washington Post* lui-même diffuse les données sur le travail des sénateurs américains.



Le cas d'OWNI est très différent. D'abord, il est beaucoup plus récent. Cependant, c'est effectivement une tentative très originale de construire un pont entre le journalisme traditionnel et la mise à disposition des données. Évidemment, on reste à la marge, mais cette fois-ci on n'est plus en dehors du journalisme. Le prix remis en octobre dernier à OWNI par la Online News Association (ONA) participe à la reconnaissance de ce site, mais cette légitimité joue plutôt au niveau international pour l'instant.

Comment expliques-tu le manque d'investissement français dans le datajournalisme ?

Déjà, il n'y a pas en France de fondations pour l'innovation, qui soutiennent financièrement de tels projets innovants. Nous avons beaucoup d'aides à la presse en France, mais elles ne fonctionnent pas de la même manière et ne touchent pas les acteurs à la marge. Ensuite, la presse traditionnelle a été assez peu sensible au développement du datajournalisme. Mais, on peut être surpris. Ce serait totalement logique qu'un site comme *Le Monde.fr* se rapproche de *NosDéputés.fr*, par exemple.

Enfin une difficulté supplémentaire dans le cas français est le manque de disponibilité des données. Pour que le datajournalisme se développe, il faut d'abord faciliter l'accès aux données. Je prend l'exemple du *Figaro* : il y a quelques semaines, le titre a lancé une enquête concernant les maisons de retraites. Elle s'accompagne d'une **visualisation** [fr] dans laquelle on peut parcourir, département par département, un classement des maisons de retraite. La différence entre le traitement de ce sujet par *Le Figaro* et le *Chicago Tribune*, c'est la méthode utilisée pour récupérer des données. *Le Figaro* a envoyé un questionnaire aux maisons de retraite. À Chicago, ce sont des statistiques de l'État de l'Illinois qui ont été croisées avec des données policières. Tout est dit. On ne peut pas avoir la même qualité des données quand on demande aux maisons de retraites de se positionner elles-mêmes.

En France, les données disponibles sont essentiellement nationales. De ce qu'on a pu observer à Chicago, les récits qui fonctionnent le mieux sont ceux qui sont ancrés localement. C'est typiquement de genre de données qui est très difficile à collecter en France. C'est possible, mais ça demande énormément de temps et de travail.

Est-on encore loin du développement d'un modèle de datajournalisme à la française ?

Un des autres enseignements de notre enquête est la spécialisation des acteurs. Il y a des experts du traitement de données, mais également des personnes qui se spécialisent dans la récupération de données. On les appelle les « People Person ». Leur travail quotidien consiste à « harceler » les autorités pour récupérer des données exploitables. Étant donné la différence des cultures entre la France et les États-Unis, beaucoup de journalistes sont moins sensibilisés à la valeur que peuvent représenter des données brutes. Je pense malgré tout qu'un mouvement est amorcé. Les journaux commencent à mesurer l'intérêt qu'ils peuvent tirer de se positionner sur ces formats journalistiques.

>> Article initialement publié sur **Silicon Maniacs**.

>> Photo Flickr  par **alamodestuff**.

MARTIAL MAURETTE REPORTER-PHOTOGRAPHE

le 20 juin 2011 - 15:54 • SIGNALER UN ABUS - PERMALINK



Je suis fort content de tous ces "data-papiers" qui nous expliquent le naufrage de la presse, et de notre noyade...

Et à force de les lire, je vais bien me faire à l'idée : que l'expérience, la qualité, la déontologie,... (Culture acquise: "le beau, le bon, le bien") sont vraiment d'une époque révolue, à la limite primitive,

(et vénale...salaires corrects, copyright, droits d'auteurs...Beurk?)

C'est donc Culture Numérique, l'Histoire Contemporaine !

Ordo Ab Chaos ...et le Renouveau viendra de la sous-culture.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

1 ping

DataJournalism by natiou schka - Pearltrees le 7 décembre 2011 - 11:24

[...] Chicago: le parrain du datajournalisme? » OWNI, News, Augmented La véritable rupture ne se fait pas avant le milieu des années 2000. Les grands titres comme le New York Times , le Los Angeles Times , le Chicago Tribune et le Washington Post ont déjà commencé à recruter des spécialistes, mais c'est la récupération de l'ensemble des données concernant la guerre, qui popularise ces méthodes. Datajournalisme articles web data journalisme Articles sur le datajournalisme Data journalisme Data journalisme L3 NPJ [-+] Data journalism election données datajournalism Data-journalisme, infographie & programmation pioneering data journalism medialab Ressources Références 5 commission chronomath cultures numériques technologiques Comment le journalisme a été augmenté sécuritaires journalisme au futur cablegate Journalismisme journalistes nostalgiques Veille quatre prospection web de données voies journalism Pure Players pulps formation tools institute guide Journalismisme de DONNEES flare Datajournalisme section utilisation exemples exemples visualization OpenData datalift Culture Geek ritchie Webjournalisme model lille Owni forme google droite Réseaux génération lectures diverses médias journalisme data mining atlas voies data journalism journaliste The art of data visualization ideas Documentaries birke cinema borealis data journalism journalism gérées database The New York Times infographic approach innovation Réseaux Sociaux movim sociale yahoo eBooks & ressources bibliothèque numériques Outils infochimps solution umaps Data directory crunchbase media update Traitements texas bureau schuster Réseaux Sociaux twitter Google effet milliard match Android premier smartwatch Nourris le geek qui est en toi ! métro montréal luxury datajournalism exemples ...I Just Cant Get Enough dissent culture Recherche 9 Sciences Humaines jalons online domitia journalisme de données database information Journalismisme home • contact • blog • fb • twitter to experience pearltrees activate javascript. [...]