

Data Journalism

La transformation digitale du journalisme



I Définition du journalisme de données

Tout peut être de la donnée

Le journalisme de données est-il simplement du journalisme fait avec des données ? Cette définition est trop large car quasiment toutes les informations peuvent être traduites en données, à savoir en bits, des zéro et des uns.

Nouvelles opportunités grâce à la digitalisation

Comme tout travail de journaliste, le data journalism consiste à mener un travail d'investigation pour le communiquer au public, mais il combine cette compétence traditionnelle avec l'opportunité offerte par le volume grandissant d'information digitale.

Le data journalisme utilise donc de la donnée et traite de sujets qui ne pourraient l'être sans un travail d'analyse des données. C'est plus ou moins la définition qu'en fait Van Es : "Data journalism is journalism based on data that has to be processed first with tools before a relevant story is possible."

Preuves irréfutables

La digitalisation seule n'est pas la raison pour laquelle les journalistes ont pris le tournant digital. En effet le pouvoir réel des données est d'apporter des preuves dures voire irréfutables, à des arguments qui jusque là pouvaient être trop subjectifs ou anecdotiques.

C'est cet élément de preuve empirique que les journalistes cherchent à incorporer dans leur travail pour donner plus de crédibilité et donc de confiance de la part du lecteur.

II Processus de transformation

Le journalisme de données n'est pas entièrement nouveau, les journalistes utilisaient déjà des données les chiffres et les statistiques pour écrire leurs articles, ainsi que des visualisations.

Computer Assisted Reporting

Le journalisme des données était ce qu'on appelait le journalisme assisté par ordinateur (CAR) dans les années 60 et il s'agissait principalement de l'usage d'ordinateurs et de bases de données. Le National Institute for Computer Assisted Reporting (NICAR) a organisé la première conférence dédiée au CAR en 1990. Elles sont tenues tous les ans depuis et sont le plus grand rassemblement de data journalistes.

Journalisme de précision

Dans les années 90 un nouveau terme apparaît, le journalisme de précision, pour désigner l'usage de statistiques et de méthodes des sciences sociales.

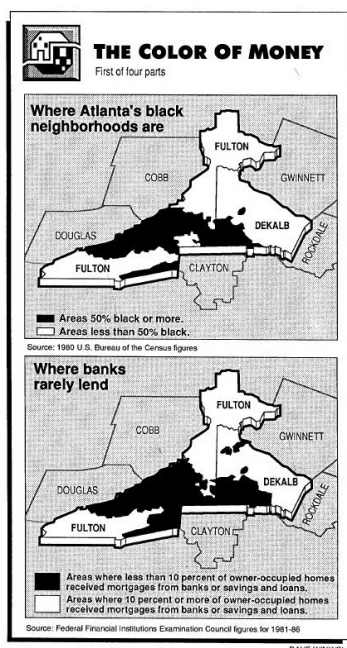
Aujourd'hui de multiples possibilités

Aujourd'hui étant donné la quantité de données générées et accessibles (des chiffres mais aussi du texte, des images, vidéos, audio) ainsi que le développement de techniques permettant de manipuler ces données, la pratique du data journalisme offre de multiples possibilités.

D'abord au niveau des sources disponibles : les données open-source ou publiques ainsi que les fuites de données, offre de nouvelles opportunités pour des sujets méritant d'être investigués.

Il est également possible que des journalistes veuillent automatiser le processus d'acquisition et d'assemblage de ces données, comme le montre Adrian Holovaty dans ChicagoCrime et EveryBlock avec des données du gouvernement, de la police et d'autres institutions locales. Ou encore, dans le projet du Telegraph MPs' expenses, des algorithmes ont été employés pour trouver des connexions dans un grand volume de documents.

Enfin les techniques de visualisation permettent d'interagir avec le lecteur d'une façon plus ludique et innovante. Il y a en effet une tendance à faire des infographies beaucoup plus engageantes pour le lecteur. Information is Beautiful est un exemple de l'état de l'art à ce sujet. Une autre visualisation très connue qui a été pionnière dans le domaine est celle de Gapminder de Hans Rosling's sur l'évolution de la pauvreté mondiale.



III Processus journalistique

On peut voir l'impact de cette transformation digitale dans le processus journalistique.

Similaire a l'analyse de données

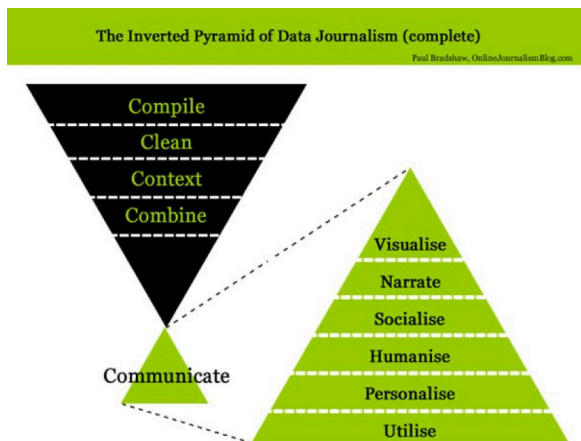
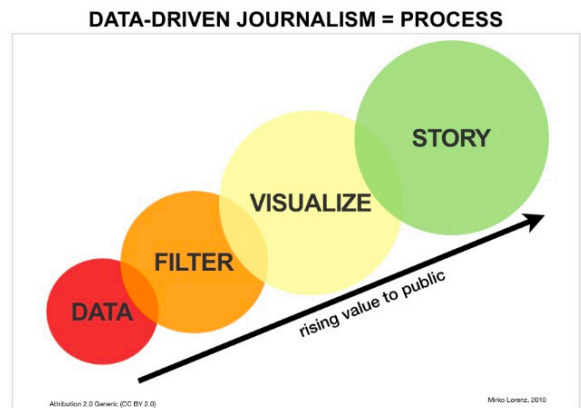
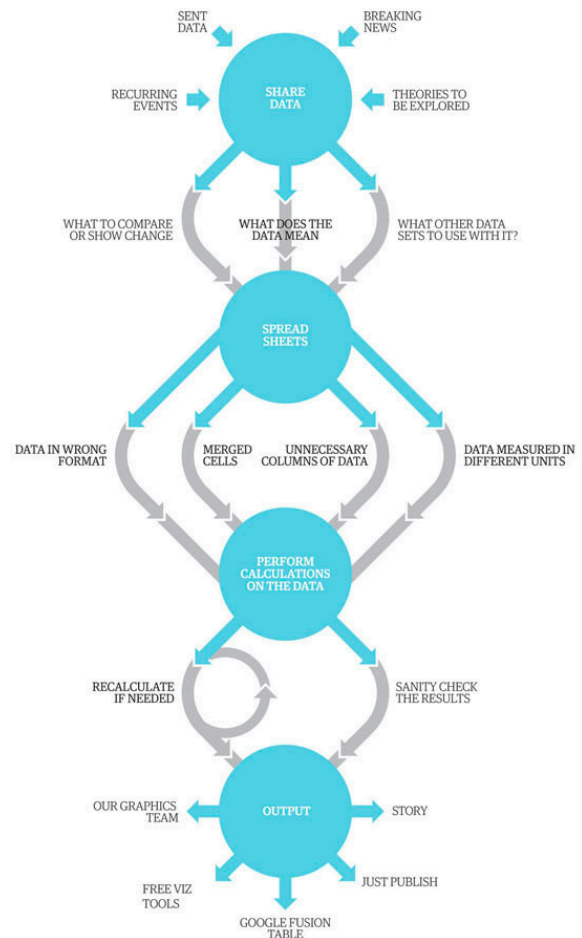
Les grandes étapes de ce processus sont similaires à un processus d'analyse de données. Premièrement un travail de récolte, de nettoyage et de structuration, suivi d'un processus de filtrage sur des données plus spécifiques et enfin une étape de visualisation.

Emphase sur la communication

L'étape de visualisation s'imbrique avec la phase du communication qui en est la finalité. Elle est particulièrement importante puisqu'il s'agit d'interagir avec le lecteur. En effet, dans cette autre modélisation du processus journalistique, on peut voir comment l'analyse de données est un processus qui permet de parvenir à la communication de l'histoire, qui est en lui-même un processus demandant une réflexion propre au journalisme traditionnel.

Limites

Aucune définition du workflow n'est tout à fait figée ni complète. Notamment, la **qualité** des données reste un enjeu essentiel. De plus, le travail d'analyse n'est pas limité à un enchaînement d'étapes à suivre, mais la dimension critique face aux données et des biais potentiels doit également être soulevée.



IV De nouveaux challenges

Nouvelles compétences nécessaires

Le principal challenge est que les journalistes doivent **acquérir les compétences** pour conduire l'analyse des données qui sont souvent hors du champ principal de leur éducation. Pour y remédier, il existe des **formations** spécifiques au data journalisme surtout en ligne. Autrement, les agences **emploient** des personnes qui sont plus qualifiées dans ce domaine.

Les teams data

Dans les grandes agences de presse, les teams data ont vu le jour spécifiquement pour permettre de produire là où des compétences plus scientifiques sont requises. Elles sont souvent composées de **statisticiens, de programmeurs, data analysts et data scientists**. Ces métiers **travaillaient indépendamment l'un de l'autre auparavant**, c'est donc un phénomène de rassemblement des compétences.

Cependant un des challenges est la **communication avec les journalistes**, c'est pourquoi certaines agences emploient une personne pour faire le pont entre les deux, et celles qui ont adopté cette stratégie ont pu produire des data driven stories de bien meilleure qualité.

V Newsrooms

Aujourd'hui la plupart des sites de presse ont des équipes spécialisées dans les investigations basées sur la data. Les **grandes newrooms ou salles de presse sont les premières à adopter des teams data** dans leur organisation.

Parmi elles, la **presse américaine** est très présente tel The New York Times, Washington Post, Wall Street Journal, FiveThirtyEight... qui sont une référence pour leurs data stories, ainsi que The Guardian, Reuters, Pro Publica, La Nación et Berliner Morgenpost.

En plus de la presse généraliste, d'autres acteurs font du journalisme des données tel que des **ONG, des institutions, des petites agences de presse ou des start-ups**. On peut voir la diversité des nouveaux acteurs dans le podcast Conversations with data, par exemple avec Amnesty International.

The Guardian - Wikileaks

Les anglais The Guardian sont considérés comme les pionniers du data journalisme et ont un data blog dédié depuis 2011. Wikileaks est sûrement l'un des sujets les plus fameux qu'ils ont relaté.

Dans leur article Wikileaks data journalism : how we handle the data, le détail du travail autour des données peut y être apprécié, en particulier lorsqu'elles ont été issues d'une **fuite**. En effet, les journalistes ont du prendre en compte la dimension de la **sécurité** et de l'intégrité des personnes qui pouvaient être mise en jeu dans la révélation de certaines histoires.

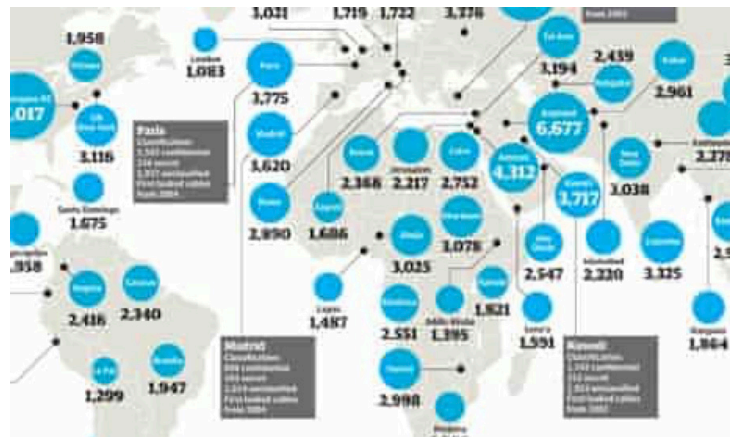
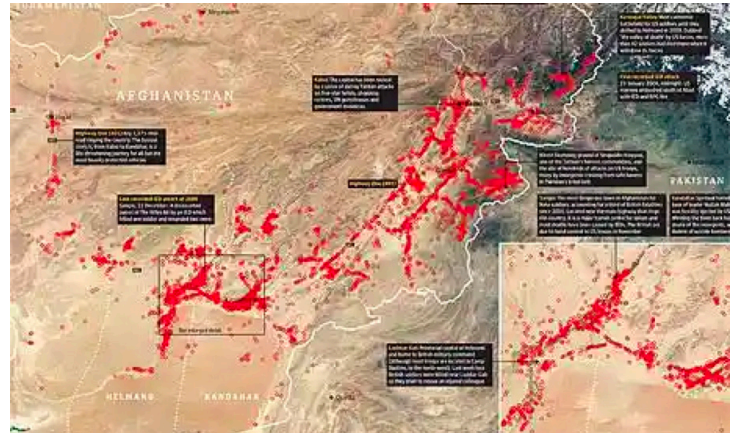
Le **SIGACTS** (US Military Significant Actions Database) est une base de données recensant les actions de guerre. Elle recouvre plusieurs événements en Afghanistan, en Irak et l'affaire des câbles.

Les données concernant l'Afghanistan sont composées de 92201 lignes. Leur analyse a permis de faire émerger le thème particulier des **Explosive Device attacks**. Ce sont des attaques d'explosifs consistant pour la plupart de bombes installées au bord de la route activées soit par télécommande, commande radio, téléphone portable ou par la victime lors de son passage. L'analyse de ces actes meurtriers a mis en exergue un fait déjà connu, les attaques avaient lieu dans la région sud des Helmand, là où étaient le plus souvent les troupes canadiennes et britanniques. Les journalistes du Guardian ont pu observer leur évolution dans le temps. Il y a eu 7500 de ces attaques entre 2004 et 2009 et ont été la première cause de décès parmi les soldats britanniques.

Pour l'Iraq on compte 391000 lignes de données. La plupart des morts étaient des civils mais il y en a sûrement eu beaucoup plus du fait que les personnes sur place n'avaient pas le temps de reporter chaque tragédie sur le terrain.

Wikileaks a aussi pu faire émerger l'affaire des câbles qui est une analyse de la diplomatie américaine à travers toutes les dépêches diplomatiques qui passaient par un réseau secret, le SIPRNet (Secret Internet Protocol Router Network). 250 ambassades et consulats à travers le monde font partie de ce réseau tenu séparé du réseau ordinaire et dirigé par le département de la Défense à Washington. Les journalistes ont dénombré 251287 documents dont 15652 classées secret, 101748 classées confidentiels et 133887 non classés. Leur travail est gigantesque et n'est pas terminé étant le volume très important. Néanmoins il en est ressorti que le plus de dépêches diplomatiques émanaient d'Ankara en Turquie, suivi de Bagdad et Tokyo et sont principalement classés confidentiels ou PTER à savoir la prévention du terrorisme.

Cette vaste investigation est une preuve de la presque hégémonie diplomatique des Etats-Unis, ces dépêches remontant de plus depuis 1966.



Conclusion

Nous avons vu qu'en tirant parti de l'accessibilité grandissante aux données, les journalistes ont acquis une nouvelle dimension à leur mission qui est de transformer ces données pour communiquer des histoires ayant un intérêt public. D'autres acteurs s'emploient à cette mission, des agences de presse plus petites ou des institutions tel que les ONG. Wikileaks est une des affaires les plus importantes démontrant ce phénomène mais d'autres projets voient continuellement le jour. Notamment, les SIGMA Awards récompensent les meilleurs projets de Data Journalism (cf. couverture).

Il reste encore beaucoup de questions à préciser pour que le data journalisme soit un domaine bien défini et comprendre comment la data joue un rôle dans le travail quotidien des journalistes. Quel est l'état de l'art du data journalisme ? Quelles sont les compétences requises ? Quel rôle joue les teams data dans une agence de presse ? Quelles sont les bonnes pratiques et quels sont les challenges ?

Un autre phénomène vient impacter la transformation du journalisme. Avec la diffusion ultra rapide devenue possible sur des plateformes tel que Twitter, la stratégie de l'attention n'est plus suffisante. Ainsi les médias sérieux cherchent désormais à adopter une stratégie qui ne se base plus sur la rapidité avec laquelle ils tentent d'informer le lecteur mais sur la confiance en la qualité des articles qui sont produits. Cette transition est largement aidée par l'analyse de la "Big Data" et les preuves parfois irréfutables qu'il est possible d'en tirer.

Bibliographie

Definition du data journalisme

« Data Journalism ». In Wikipedia, 18 janvier 2021. https://en.wikipedia.org/w/index.php?title=Data_journalism&oldid=1001237834.

« Data-Driven Journalism ». In Wikipedia, 9 mars 2021. https://en.wikipedia.org/w/index.php?title=Data-driven_journalism&oldid=1011113086.

InternetActu.net. « Les données pour comprendre le monde ». InternetActu.net. Consulté le 22 mars 2021. <https://www.internetactu.net/2011/07/19/les-donnees-pour-comprendre-le-monde/>.

DataJournalism.com. « Podcast Conversation with Data ». Consulté le 22 janvier 2021. <https://datajournalism.com/listen>.

Processus de transformation

Bradshaw, Paul. « The Inverted Pyramid of Data Journalism ». Online Journalism Blog (blog), 7 juillet 2011. <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>.

DataJournalism.com. « Datajournalism.Com ». Consulté le 22 janvier 2021. <https://datajournalism.com/>.

Gray, Jonathan. Repository journalism courses, 2021. <https://github.com/jwyg/data-journalism-courses>.

Knight Center for Journalism in the Americas. « Knight Center for Journalism ». Consulté le 22 mars 2021. <https://knightcenter.utexas.edu/>.

Dropbox. « Machine learning for the rest of us.pdf ». Consulté le 23 janvier 2021. <https://www.dropbox.com/s/nr8cshr27p6vobr/Machine%20learning%20for%20the%20rest%20of%20us.pdf?dl=0>

Newsrooms dédiées au data journalism

« Atlanta Shootings: Why US Hate Crime Data Is so Lacking | Mona Chalabi ». The Guardian, 20 mars 2021, sect. US news. <http://www.theguardian.com/us-news/datablog/2021/mar/20/asian-american-hate-crime-data-mona-chalabi>.

Silver, Nate. « How Popular Is Joe Biden? » FiveThirtyEight, 28 janvier 2021. <https://projects.fivethirtyeight.com/biden-approval-rating/>.

the Guardian. « Wikileaks Data Journalism: How We Handled the Data », 31 janvier 2011. <http://www.theguardian.com/news/datablog/2011/jan/31/wikileaks-data-journalism>.

« Operation Iraqi Freedom - Iraq Significant Activities (SIGACTS) ». Consulté le 22 mars 2021. https://www.globalsecurity.org/military/ops/iraq_sigacts.htm

Projets de data journalism

Made in France

« Made in France ». Consulté le 22 mars 2021. <https://made-in-france.disclose.ngo/>.

« “Made in France” - Itinéraire d’une livraison secrète d’armes françaises - YouTube ». Consulté le 22 mars 2021. https://www.youtube.com/watch?v=BKU1iHmaJL0&ab_channel=Disclose

Pointer

KRO-NCRV. « Het verhaal achter een identiteitsroof ». Consulté le 02 mars 2021. <https://pointer.kro-ncrv.nl/het-verhaal-achter-een-identiteitsroof>

DataJournalism.com. « Podcast Conversation with Data ». Consulté le 22 février 2021. <https://datajournalism.com/listen>.

DataJournalism.com. « Sigma Data Journalism Awards ». Consulté le 22 mars 2021. <https://datajournalism.com/awards>