

APPEL À CONTRIBUTIONS

pour le premier numéro de la revue en ligne sur l'Internet¹.

Études Culturmétriques

Publication fin 2018

Big data et sciences humaines et sociales :

critique des données massives, critique par les données massives ?

Coordination :

— Éric BAILBLÉ, directeur de publication de la revue, docteur en histoire et en linguistique, enseignant en histoire et géographie à Budapest.

— Frédéric CLAVERT, coordinateur du département histoire de la revue, *Senior Research Scientist* à l'université du Luxembourg (Center for contemporary and digital History).

— Jean-Baptiste LE CORF, rédacteur en chef de la revue, maître de conférences en Sciences de l'Information et de la Communication (Université Rennes 2 – PREFics)

Si le terme « big data » - en Français « mégadonnées » ou « données massives » - existe depuis plusieurs décennies, il a fait l'objet d'un intérêt grandissant depuis l'émergence du web dans les années 1990, qui s'est accéléré avec l'importance croissante des services de Google et la création des réseaux sociaux numériques comme Twitter ou Facebook, plateformes web dites « sociales » qui engendrent un volume de données inédit.

Les définitions des données massives sont nombreuses. Nous nous tiendrons ici à la suivante : des données sont massives lorsqu'elles échappent non seulement à l'intuition humaine mais également aux traitements classiques de l'informatique personnelle. Le grand intérêt des années 2010 pour les données massives est d'abord lié au potentiel des applications commerciales de son traitement. Toutefois, très vite, les sciences en général mais aussi les sciences humaines et sociales s'en sont saisies.

Le traitement des données massives a été permis grâce aux mutations de la nouvelle sémantique informatique et des corpus, ainsi qu'avec le développement des modèles probabilistes et soutenu par l'accroissement continu depuis plusieurs décennies des capacités de calcul des ordinateurs.

L'ère du *big data* se caractérise par une forme de pragmatisme, les *data scientists* pouvant réduire l'utilisation d'*a priori* dans la construction des modèles et mesurer leur qualité par leur pouvoir prédictif. Appuyés sur ces nouvelles technologies, le calcul intensif et la modélisation permettent d'envisager un nombre infini de visualisations des données ; chiffres, lettres, graphiques, présentation dynamique, etc. Dans le même temps, nombreuses sont les plateformes qui proposent une lecture massive de données textuelles. Aux États-Unis,

¹ <https://etudes-culturmetriques.com/>

de nombreux projets ont ainsi vu le jour. Le laboratoire Cesta à Stanford, créé en 2012, se consacre par exemple à des analyses massives de textes dans le temps et l'espace. L'université du Massachusetts, de son côté, fournit un ensemble d'outils statistiques de traitement automatique du langage appelé MALLET et permettant aux chercheurs en sciences humaines et sociales un traitement très poussé de leurs données textuelles.

Dans ce contexte des données massives, le premier numéro de la revue *Études Culturmétriques* visera donc à explorer cette partie des humanités numériques – que nous définissons comme se situant « dans la convergence des démarches scientifiques et de l'ingénierie informationnelle » (Cormerais, 2014, p. 133) – qui utilise les données massives dans un sens large. Les différentes contributions pourront donc s'intéresser aux processus de traitement de données scientifiques (constitution de corpus, analyse, traitement, exploitation et visualisation des données), qu'elles soient ou non nativement numériques. Plus précisément sont attendus des articles apportant des éclairages théoriques et/ou des études de cas de projets de recherche mettant en œuvre certains de ces processus et exposant les instruments et techniques mobilisés.

Trois axes de recherches seront plus particulièrement travaillés :

- la conception de la science au prisme du *big data*
- la vigilance méthodologique face aux données massives
- la transdisciplinarité des humanités numériques induite par l'usage de ces données massives

Ces trois défis auxquels doivent faire face les chercheurs en SHS, mais aussi les professionnels du savoir et de la documentation scientifique, pourront faire l'objet d'analyses empiriques ou conceptuelles.

Axe 1. La conception de la science au prisme du big data

Ce premier axe vise à prendre une distance critique vis-à-vis de certains discours véhiculés par les promoteurs du *big data*, qui laissent à penser que les données massives et leur traitement ont radicalement changé la conception même des sciences. Emblématique de cette thèse, Chris Anderson, rédacteur en chef de *Wired Magazine*, publiait en juin 2008 un article au titre provocateur : « The end of theory : the data deluge makes the scientific method obsolete ». Selon lui, lorsque nous disposerons de suffisamment de données, les nombres parleront d'eux-mêmes et les corrélations ainsi dévoilées remplaceront les relations de causalité avancées par les théories. La science pourrait désormais se passer d'hypothèses explicites, attestant d'une véritable rupture épistémologique.

La conclusion est sans doute hâtive, et cette approche a été notamment déconstruite par Danah Boyd et Kate Crawford (« *6 provocations about big data* »). Les promoteurs du *big data* n'occultent-ils pas en effet le fait que les données ne sont jamais brutes - « *raw data is an oxymoron* » - mais issues d'un processus social soulignant que la question d'un corpus n'est pas neutre ? Des études ont par exemple mis en avant la non-représentativité de corpus tels que celui de Google Books. On peut également citer le cas du fameux *like* de Facebook ; les nombreuses données générées par cette fonctionnalité, indéniablement liées à la finalité commerciale de Facebook, ne sont-elles pas difficilement interprétables ?

Plus largement, *Études culturmétriques* souhaite ici soumettre à l'épreuve l'hypothèse selon laquelle l'exploitation des traces numériques, facilitée par les industries de la communication que sont les plateformes GAFa, entraînerait inévitablement l'avènement d'une nouvelle génération de sciences sociales (Boullier, 2017).

Enfin, il s'agira de surcroît d'apprécier les rapports entre humanités numériques et épistémologie. Les humanités numériques ont-elles un rôle épistémologique à jouer dans la rencontre entre SHS et données massives ? Et constituent-elles une épistémologie singulière ou leur vocation serait-elle plutôt d'être au service d'autres épistémologies plus établies, comme la critique (Granjon, Magis, 2016), la neutralité axiologique ou la recherche-action ?

Axe 2. La vigilance méthodologique face aux données massives

Le deuxième axe de réflexion proposé concerne la position des chercheurs en SHS face à leur objet d'étude. Ils sont en effet amenés à développer une attitude réflexive sur les conditions de leurs propres pratiques de recherche et leur rapport technique à la science. Il s'agit ici d'examiner l'usage que le chercheur fait de l'informatique, son utilisation du numérique comme outil et méthode pour construire de nouveaux corpus ou explorer certains terrains de recherche liés aux traces numériques, comme les usages du web, l'identité numérique des organisations ou l'image des territoires.

Par ailleurs, les contributeurs pourront discuter des opportunités ou au contraire des menaces liées à l'utilisation de certains outils, notamment des outils en ligne qui peuvent mener l'internaute jusqu'à des occurrences inattendues. Comment ces dispositifs proposent-ils de rénover les procédures méthodologiques à partir de mégacorpus ? Pour l'historien par exemple, l'application en libre accès *Ngram Viewer* (visualiseur de fréquence de mots) permet d'étudier la production éditoriale mondiale des XIX^e et XX^e siècles. Cet instrument numérique autorise une approche assez précise des occurrences lexicales des deux derniers siècles (Bailblé et Bailblé, 2015) et permet de repérer des signaux faibles et rares. Les textes et images ainsi automatisés permettent un traitement massif des données et leurs corrélations aident à déconstruire, par exemple, un certain nombre de mythes.

Il est donc crucial de comprendre comment de grands volumes de données, traités par des algorithmes pour en extraire des corrélations et des connaissances (*data mining*) et quelques fois pour représenter ces résultats sous forme visuelle (*data visualisation*), posent pourtant des difficultés épistémologiques. En effet, l'existence de corrélations (*data analytics*) ne prouve pas l'existence d'une causalité entre les données référentes, et une corrélation n'équivaut pas à une signification ou une à connaissance. Dès lors, la tension fondamentale entre une science fondée sur la causalité et une analyse qui s'appuie sur la corrélation restera au cœur de nos débats.

Axe 3. Vers une transdisciplinarité induite par le traitement des données massives ?

Le troisième et dernier axe permettra de se demander si le traitement des données massives oriente le chercheur vers une posture de transdisciplinarité ou d'interdisciplinarité. Cette question rejoint plus largement la problématique de savoir si les humanités numériques peuvent se revendiquer comme transdisciplinaires. Autrement dit, face aux enjeux posés par le traitement des données massives, quels espaces interdisciplinaires ou transdisciplinaires (entre SHS, ingénieurs et informaticiens notamment) peut-on créer autour de concepts et de méthodes communes relevant des *data sciences* ? Quelles applications concrètes pourrait-on en retirer ?

Par ailleurs, dans quelles conditions pourrait-on par exemple proposer des formations en informatique et en ingénierie de l'information scientifique afin d'utiliser pleinement leur potentiel technique dans un projet de recherche ?

Cette dernière question est d'une importance cruciale à l'heure où se multiplient de nouvelles offres transdisciplinaires de formation universitaire mentionnant les humanités numériques².

Modalités de soumission

Les propositions sont à adresser pour le **30 juin 2018** à : redaction@etudesculturmetriques.fr

Tous les textes d'*Études Culturmétriques* font l'objet d'une double évaluation pour décider de leur publication.

- Le comité de rédaction peut être amené à demander des compléments, des précisions ou à proposer des reformulations.
- La version définitive du texte (prenant en compte les éventuelles remarques et retours des évaluateurs), incluant les corrections mineures ou majeures demandées, sera soumise au comité éditorial de la revue qui est souverain pour l'accord définitif de publication en **décembre 2018**.

Quant à la forme attendue des contributions, vous pouvez consulter [ici](#) notre charte éditoriale.

Les textes seront accompagnés d'une photographie et d'une notice biographique fournies par l'auteur, indiquant ses fonctions, son institution/organisation de rattachement et éventuellement ses publications récentes. Elle peut intégrer un lien hypertexte renvoyant à sa page web.

Bibliographie indicative

BAILBLÉ, Olivier ; BAILBLÉ, Éric. « Les nouvelles technologies face à l'historiographie positiviste de l'histoire de France : étude sur le XIX^e siècle en français : enjeux et perspectives », *Revista Electrónica Matices en Lenguas Extranjeras*, numero 9, pp. 68-90.

BOULLIER Dominique. (2017). « Les sciences sociales face aux traces du big data ? Société, opinion, répliques », FMSH-WP, n° 88, avril 2015.

BOURDELOIE, Hélène. (2013). « Ce que le numérique fait aux sciences humaines et sociales », *tic & société*, Vol. 7, N° 2, <http://ticetsociete.revues.org/1500>

CHARTRON, Ghislaine. (2014). « Open access et SHS : Controverses », *Revue européenne des sciences sociales*, 52-1. URL : <http://ress.revues.org/2658> ; DOI : 10.4000/ress.2658

CARDON, Dominique (2015). *A quoi rêvent les algorithmes ?*, Seuil.

CARMES, Maryse. & NOYER, Jean-Max. (2014). « L'irrésistible montée de l'algorithmique: méthodes et concepts en SHS ». *Les Cahiers du numérique*, vol. 10,(4), p. 63-102.

CASILLI Antonio, « Pratiquer la transdisciplinarité dans la discipline. Temporalité, territorialité et réalisme des professions scientifiques », in DARBELLAY Frédéric, PAULSER Theres (dir.), *Au miroir des disciplines. Réflexions sur les pratiques d'enseignement et de recherche inter- et transdisciplinaires*. Oxford, Peter Lang, 2011, p. 65-81.

² À l'Université Rennes 2 par exemple, citons l'existence du master Sciences Humaines et Sociales, mention « humanités numériques ».

CASILLI Antonio, « Comment les usages numériques transforment-ils les sciences sociales ? », Read/Write Book 2. Une introduction aux humanités numériques, 2012, p. 239-247.

CLAVERT, Frédéric. (2014). « Vers de nouveaux modes de lecture des sources », dans LE DEUFF, Olivier (dir). Le temps des humanités digitales. La mutation des sciences humaines et sociales, FYP, p. 33-47.

COLLECTIF, Manifeste des digital humanities, 26 mars 2011, URL : <http://tcp.hypotheses.org/318>.

CORMERAIS, Franck. (2014). « Humanités digitales et (ré) organisation du savoir », dans LE DEUFF, Olivier (dir). Le temps des humanités digitales. La mutation des sciences humaines et sociales, FYP, p. 129-142.

GEFEN, Alexandre. (2015). « Les enjeux épistémologiques des humanités numériques », Socio, 4 |, p.61-74.

GITELMAN, Lisa dir. (2013). « *Raw data* » is an oxymoron, MIT Press.

MAGIS, Christophe. GRANJON, Fabien. Critique et humanités numériques. Pour une approche matérialiste de l'« immatériel », Variations, 19 | 2016.

MOUNIER, Pierre. (2012). « Qu'apportent les digital humanities ? Quelques exemples », dans P. MOUNIER (dir.), Read/Write Book 2, OpenEdition Press, p. 75-83, <http://books.openedition.org/oep/247>