



HAL
open science

Recherches digitales et production des données, bouleversement des épreuves pour le chercheur en SIC

Franck Cormerais, Amar Lakel

► To cite this version:

Franck Cormerais, Amar Lakel. Recherches digitales et production des données, bouleversement des épreuves pour le chercheur en SIC. Etudes digitales, 2018, <10.15122/isbn.978-2-406-09563-7.p.0155>. <hal-02485339>

HAL Id: hal-02485339

<https://hal.science/hal-02485339v1>

Submitted on 24 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Recherches digitales et production des données, bouleversement des agencements pour le chercheur en SIC

Franck Cormerais,
Université Bordeaux-Montaigne
MICA (EA 4426) - Équipe E3D
franck.cormerais@u-bordeaux-montaigne.fr

Amar Lakel
Université Bordeaux-Montaigne
MICA (EA 4426) - Équipe E3D
amar.lakel@u-bordeaux-montaigne.fr

Citation :

Franck Cormerais et Amar Lakel. "Recherches digitales et production des données, bouleversement des agencements pour le chercheur en SIC", *Etudes Digitales*, N°6, avril 2019, p.94-111.

« Entre les données désirables et les données accessibles, entre les recommandations idéales, théoriques et méthodologiques faites et les conceptions de chacun, il existe une notable différence. »
Estivals R., *Les sciences de l'écrit*, 1993, p. 6

Introduction

Le «tournant digital» de la société est désormais une évidence même si sa définition fait encore débat dans les milieux académiques. À tout le moins, il est possible de dire que les technologies qui y sont impliquées permettent une explosion, dans les sociétés humaines, d'interfaces interconnectées à des systèmes d'information globaux. Un ensemble, chaque jour plus important, de phénomènes est médiatisé par des interfaces qui fonctionnent comme des dispositifs signifiants enregistrant des données (data). Des millions de traces sont, chaque seconde, stockées dans des bases de données et dans des serveurs de fichiers. Ordinateurs, tablettes, smartphones, caméras, objets mobiles sont un essaim de capteurs au service d'une production d'archives qui échappent à toute forme de régulation. Les sciences humaines et sociales ont été dans ce contexte, elles aussi, touchées par la digitalisation des traces et toute la valorisation de la donnée qui s'en est suivi. Le passage du mémorisable au prédictif ouvre sur des enjeux économiques, politiques et scientifiques qui reconfigurent profondément la structure des actants dans la production de savoir.

Cet article se veut une contribution critique du «tournant digital» des SIC. Comment un régime neuf de la trace engendre-t-il un remaniement profond du document

entendu comme source du savoir ? Comment cette numérisation du signe fait-elle entrer l'algorithme dans un outillage du chercheur ? Comment cette masse de traces, produites parfois en temps réel, redessine les collaborations scientifiques, voire remet au goût du jour une inter ou transdisciplinarité ? Notre terrain critique est l'aboutissement de deux projets de recherche innovants en sciences de l'information et de la communication et en sociologie qui semblent dessiner une prise de position pratique «commune». Ces travaux viennent éclairer les changements majeurs à l'œuvre dans la recherche en communication numérique¹. À travers l'étude critique de ces réalisations scientifiques récentes, engageant des méthodes digitales issues tant des SHS que de l'informatique et du monde professionnel, nous interrogeons cette nouvelle économie scientifique en trois moments dans la volonté de dégager les nouvelles pratiques scientifiques opérationnelles en humanités digitales et numériques.

Dans un premier temps, les datafications des corpus et les données massives ouvrent un champ de pratiques hybrides en cours. Nous voulons souligner les dynamiques induites par une économie numérique du document, où le mémorisable devient une source de revenus et donc de pouvoir. Dans un second temps, nous abordons le passage entre data et corpus engageant les recherches algorithmiques dans un processus de signifiante qui tente de dépasser les oppositions traditionnelles entre « quanti » et « quali ». Enfin, dans une troisième partie, il apparaît que les enjeux des études digitales rendent nécessaire une interrogation sur le statut et l'autonomie des agencements de chercheurs dans la perspective d'un questionnement et d'un dialogue entre les disciplines.

1. Le document digital et les données : l'épreuve de l'accès aux données

Les problèmes documentaires renvoient à une histoire du document qui redistribue les relations entre la mémoire et l'oubli, le visible et l'invisible². Avec le digital, la documentologie comme science du document³ connaît une nouvelle étape de problématisation au sein des SIC qu'il nous faut expliciter afin de mieux éclaircir les enjeux de la recherche dans le domaine. Trois points peuvent être successivement évoqués pour en décrire la dynamique : le témoignage, l'accessibilité et l'interconnexion. Ces trois «épreuves» du document numérique permettent de mieux saisir les voies de la documentologie de notre temps en ce qu'ils redéfinissent l'épistémè des savoirs numériques.

1.1. Le défi de l'hétérogénéité des données massive face à l'information et à la connaissance

¹ Bastard, I. Et al. (2017). «Facebook, pour quoi faire : Configurations d'activités et structures relationnelles». *Sociologie*, vol. 8,(1), 57-82. ; Lakel, Amar. (2014a). «Enjeux de la recherche digitalisée dans les humanités numériques», *My Web Intelligence*. Consulté à l'adresse <http://mywebintelligence.net/enjeux-de-la-recherche-digitalisee-dans-les-humanites-numeriques/>; Lakel, A., & Le Deuff, O. (2016). «À quoi peut bien servir l'analyse du web ?», *Les Cahiers du numérique*, 13(3), 39-62.

² Escarpit, R. (1995). *L'information et la communication. Théorie générale*, Paris, Hachette.

³ Meyriat J. (1981). « Document, documentation, documentologie ». *Schéma et schématisation*, 2e trimestre, no 14, p. 51-63

Séries binaires enregistrées en bases de données, fichiers numériques, codes informatiques : cette pluralité de formats liée à la convergence des régimes documentaires apparaît comme la première épreuve du chercheur en SHS. Les dispositifs digitaux sont la source d'une production de documents très hétérogènes, constitués de couches multiples et imbriquées de traitements, qui viennent bousculer les techniques traditionnelles de l'enquête⁴. Notons d'abord l'existence de l'immense corpus possible en temps réel des conversations⁵ auquel il faut ajouter les formes variées de l'éditorialité numérique⁶ qui sont autant de traces des échanges. Ajoutons à cela que ces traces sont de plus en plus hétérogènes dans leurs formats. Image, vidéo et son, par leur massification dans les échanges intersubjectifs, interrogent la place des méthodes visuelles en SHS face à des corpus de plus en plus multimédias. Il faut ajouter à cela l'ensemble des documents produits par des captations des comportements (caméras, téléphone mobile, capteurs autonomes, objets connectés, etc.) qui engendrent une surveillance généralisée dans un quantify self plus ou moins volontaire⁷. Ces captations numériques ont pu paraître comme une promesse d'un témoignage généralisé dont les sciences seraient l'un des bénéficiaires (loin derrière l'industrie de l'attention). Mais pour chacun de ces systèmes sociotechniques, la première des prudences scientifiques n'est-elle pas d'interroger l'écart entre l'usage et ses traces ?

La production des données est en réalité le symptôme d'une extension indéfinie de la captation du monde dont le «ghost» numérique n'est qu'un sous-ensemble⁸. Derrière les données se pose le problème de la capacité de la signifiante comprise comme processus interprétatif qui mélange une culture matérielle et une culture logicielle⁹. Si l'information est le résultat d'une trace au service d'une intention, il faut en conclure que les données impliquent toujours un traitement vu comme processus de signifiante problématique¹⁰. C'est donc à une volonté de savoir que la data doit être assignée pour assurer la production d'une connaissance à partir de l'information. La promesse de la data science réside dans un découplage entre collecte a priori généralisée du monde et problématisation de l'enquêteur finie et a posteriori. Mais la multiplicité des rationalités de la captation¹¹ implique, comme préalable à la connaissance, la construction d'un cadre de « traduction » où les traces accessibles émergent comme première étape de la signifiante.

Nous faisons désormais face à un territoire de la donnée si vaste qu'il transforme chaque chercheur en archéologue (c'est-à-dire aussi en géologue). Le chercheur dans les

⁴ Paloque-Berges, C. & Schafer, V. (2015). «Quand la communication devient patrimoine». *Hermès, La Revue*, (71), 255--261.

⁵ Proulx, S., & Rueff, J. (2018). *Actualité des méthodes de recherche en sciences sociales sur les pratiques informationnelles*. Consulté à l'adresse http://epe.lac-bac.gc.ca/100/200/300/univ_laval/centre_etudes_medias/CEM-ActualiteMethode.pdf

⁶ Vitali-Rosati, M., & Sinatra, M. E. (Éd.). (2014). *Pratiques de l'édition numérique*, Montréal: Presses de l'Université de Montréal. Consulté à l'adresse <http://books.openedition.org/pum/306>

⁷ Pucheu, D. (2016). «Effacer l'interface : Une trajectoire du design de l'interaction homme-machine», *Interfaces numériques*, 5(2), 257--276.

⁸ Rieder, B. (2010,). «Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention», *Études de communication. langages, information, médiations*, (35), 91--104

⁹ Serres, M. (2015). *Le gaucher boiteux, - Figures de la pensée*. Paris: LE POMMIER

¹⁰ Boullier, D. (2015). «Les sciences sociales face aux traces du big data, Big Data traces and social sciences». *Revue française de science politique*, 65(5), 805-828.

¹¹ Ferraris, M. (2012,). *Documentality: Why It Is Necessary to Leave Traces*, New York: Fordham University Press.

humanités digitales va devoir puiser dans la tradition archivistique des sciences historiques pour élaborer son nouvel éthos. Cependant, il n'est plus dans un terrain fait de raretés enfouies dans une terre familière (qu'elle soit de glaise ou de papiers), mais, au contraire, face à des signes numériques abondants, toujours traces secondes d'un monde familier disparut. Un terrain médiatisé par de fortes compétences numériques qui nécessite, pour le scientifique, un outillage de traitement de la data non encore stabilisé. Ce bouillonnement d'innovations continues (et dont l'intelligence artificielle pointe un voyage vers l'inconnu) ouvre un nouvel horizon d'épreuves pour nos disciplines¹² si elles veulent s'engager dans la compréhension des mondes massivement captés.

1.2. L'accès à des données structurées : de l'open access à l'open science

Si la complexité de la donnée est une épreuve a priori nécessaire au processus de savoir, la question de l'accessibilité des données et des documents apparaît, donc, comme la seconde épreuve des humanités digitales¹³. L'accès à ces données pose des limites matérielles qui sont autant de révélateurs de la reconfiguration des jeux d'acteurs dans l'économie des savoirs. Le caractère globalement privé de la propriété des données est le premier frein à la production du savoir. La confidentialité, le respect de la vie privée et la nécessaire anonymisation apparaissent comme le second frein à leur utilisation. Enfin, la question encore délicate de l'usage de ces données a posteriori à des fins de recherche et du consentement du sujet pose un troisième frein non pas tant juridique qu'éthique¹⁴. Alors même que la signifiante de la data dans un processus de recherche s'impose comme une question délicate, il faut aussi construire les conditions d'accessibilité des données pour fabriquer des corpus constitués à des fins de recherche¹⁵.

Plusieurs possibilités s'offrent au chercheur qui voudrait s'engager dans les études digitales. La première voie que peut prendre le chercheur est d'actualiser les pratiques d'enquête en se dotant des nouveaux outils numériques de captation. Si l'enquête en ligne est d'ores et déjà devenue un classique, l'expérience récente de l'ANR Algopol¹⁶, visant à construire un capteur des données des profils Facebook des utilisateurs, a montré la capacité d'innovation des nouvelles méthodes digitales en matière d'enquête sur les réseaux sociaux. Le Projet My Web Intelligence¹⁷ qui vise à développer un outil d'absorption et de gestion d'un corpus web ouvert et hétérogène élargit la logique de la démarche à l'ensemble de l'éditorialité publique en ligne en captant les expressions hétérogènes du web en un corpus centralisé directement utile à la recherche.

La seconde voie d'accès des données vise à contractualiser (de façon pécuniaire ou pas) avec les détenteurs de ces ressources leur usage au service de problématiques de

¹² Lucas, J.-F. (2014). «Le projet Magic Ring : expérimentation d'une méthode de recueil de données qualitatives dans Second Life», *Tic&société*, (Vol. 7, N° 2).

¹³ Dacos, M., & Mounier, P. (2015). *Humanités numériques*, Institut français. Rapport consulté à l'adresse <https://hal.archives-ouvertes.fr/hal-01228945/document>

¹⁴ La controverse sur l'étude de EU Desinfolab qui avait usé de tweets pourtant publics à des fins de recherche politique dans le contexte «d'affaire Benalla» en août 2018 est, à elle seule, paradigmatique de notre propos.

¹⁵ Cardon, D. (2013b). «Dans l'esprit du PageRank», *Réseaux*, (177), 63-95; Lakel, A., & Le Deuff, O. (2016). «À quoi peut bien servir l'analyse du web ?», *op.cit.*

¹⁶ Bastard, I. et al. (2017). «Facebook, pour quoi faire : Configurations d'activités et structures relationnelles». *op.cit.*

¹⁷ Lakel, Amar. (2014b). «My Web Intelligence : Un projet d'outils open source au service des humanités numériques.» *My Web Intelligence*. Consulté à l'adresse <http://mywebintelligence.net/outils-open-source-humanites-numeriques/>

recherche. L'enjeu ici étant de connaître la nature et la structure de ces données comme condition de réalisation des projets de recherche et de pouvoir formaliser les contrats de collaboration public-privé dans un champ peu habitué à voir les entreprises devenir des acteurs de la connaissance¹⁸. Facebook, à l'instar de tous les géants du web, a développé un programme académique de collaboration avec la recherche au sein de sa division Facebook Research qui a tout d'une politique de recherche¹⁹. Mais doit-on laisser aux seuls géants du web l'initiative d'une telle ouverture aux données et aux traitements ? N'y-at-il pas en la matière une véritable politique publique dans le soutien à l'ouverture des données au service de la recherche qui serait appuyé par le cadre législatif ? Où en est-on des projets d'archivage public du web²⁰ ?

La troisième voie enfin, initiée par le mouvement de l'Open Data search²¹, est celle d'un partage ouvert des données de la recherche en SHS afin de constituer des corpus communs et réutilisables²². Les enjeux de normalisation et de développement en cours ouvrent peut-être une nouvelle ère pour l'accessibilité des données à des fins scientifiques. Cette voie déboucherait alors sur une rencontre des disciplines²³. L'Institut pour les Sciences Sociales Quantitatives (IQSS) de l'Université de Harvard, à l'origine du projet Dataverse auquel participent déjà de nombreux laboratoires européens, est un des pionniers en la matière²⁴.

1.3. L'enjeu d'une interconnexion des données vers des archives ouvertes

Une des questions majeures de l'accessibilité des données pour la recherche repose sur l'interconnexion²⁵. En effet, au-delà des banques de données de la recherche, une accumulation de la connaissance est possible si des corpus répartis ont la possibilité de s'interconnecter par des architectures et des nomenclatures communes. La TEI est aujourd'hui un des projets les plus aboutis dans la réflexion d'un (ré)encodage commun à des corpus d'humanités numériques²⁶. Dans ce cas, ce qui relie les enquêtes singulières serait une même définition de l'individu. Une sorte de métacodage d'un comportement

¹⁸ Monnoyer-Smith, L. & Plantin, J.-C.(2014). «Ouvrir la boîte à outils de la recherche numérique. Trois cas de redistribution de méthodes», *Tic&société*, (Vol. 7, N° 2).

¹⁹ <https://research.fb.com/programs/>

²⁰ Paloque-Berges, C. & Schafer, V. (2015). «Quand la communication devient patrimoine». *Hermès, La Revue*, (71), 255-261.; Musiani, F. et al. (2019). *Qu'est-ce qu'une archive du web ?* Marseille: OpenEdition Press. Consulté à l'adresse <http://books.openedition.org/oep/8713>

²¹ Lire le rapport de Elsevier(2016), *Open Data : the researcher perspective*, https://www.elsevier.com/_data/assets/pdf_file/0004/281920/Open-data-report.pdf

²² A titre d'exemple, nous renvoyons le lecteur au projet PROGEDO (<http://www.progedo.fr>) cité in (Boulier, 2015)

²³ Cormerais Franck (2015b), « Humanités digitales, transdisciplinarité et entretexualité. Éléments pour une organisation des corpus partagés », in Carayol Valérie et Morandi Franck (sous la dir. de), *Le tournant numérique des sciences humaines et sociales*, Bordeaux, Maison des sciences d'Aquitaine, p. 55-65.

²⁴ <https://dataverse.org/>

²⁵ Lakel Amar (2007) « Interconnexion et identités électroniques : le pouvoir d'enquête à l'heure des systèmes de fichiers répartis » in *Terminal*, dossier spécial « Administration électronique », 2007.

²⁶ Ben Henda, M. (2015). "Adaptation normative des Big Data et du Learning Analytics." dans : Évelyne Broudoux éd., *Big Data - Open Data : Quelles valeurs ? Quels enjeux ?*, Actes du colloque « Document numérique et société », Rabat, 2015 (pp. 197-212). Louvain-la-Neuve, Belgique: De Boeck Supérieur. doi:10.3917/dbu.chron.2015.01.0197; Ben Henda, M. & Hudrisier, H. (2013). "Penser, classer, apprendre et communiquer. Normalisation et nouveaux modes de classification du savoir.", *Hermès, La Revue*, 66(2), 160-166. <https://www-cairn-info.ezproxy.u-bordeaux-montaigne.fr/revue-hermes-la-revue-2013-2-page-160.htm>.

(l'éditorialisation numérique par exemple) permettrait de normaliser collaborativement les corpus de recherche et leur constitution dans le but d'une mise en commun. Une autre piste s'appuie sur les corpus partagés permettant des collaborations qui permettent à plusieurs projets de recherche de travailler sur une population d'individus (au sens statistique) commune, mais en variant les protocoles de recherche. Cette voie «additive» consisterait à cumuler, deux ou plusieurs études ayant en commun les individus de l'enquête (l'actant ou les phénomènes étudiés). Sur le modèle des études en cohorte, cette approche multiplierait les regards sur un actant ou sur des phénomènes. Enfin, une troisième piste, encore plus intégrative, vise à étudier plusieurs itérations d'un même protocole de recherche faite sur des populations aux définitions plus ou moins similaires (avec au minimum une constante différenciante) dans une approche comparative. Cette conduite considère que chaque enquête forme une variation à l'intérieur d'une enquête globale reposant sur des objets communs. En tout état de cause, toutes ces formes d'interconnexions nécessitent une volonté de normalisation et des standards favorisant l'interopérabilité des données.

Dans les mondes numériques, l'interopérabilité est la question fondamentale depuis l'origine. Sur ce modèle, l'interopérabilité des recherches en SHS pourrait vite devenir un enjeu crucial. Dans le cadre du projet My Web intelligence, la possibilité de partager un même corpus de pages web caractérisées par une nomenclature centrale, mais ouverte à la construction de formulaires de qualifications spécifiques à chaque recherche est un enjeu majeur des développements à venir. Une même population de pages peut être sujette à des qualifications, des analyses et des interprétations diverses, voire concurrentes. Le problème des dispositifs à venir est bien celui d'une invention d'un dialogue réglé entre les disciplines autour d'un corpus partagé²⁷. Autour de l'acte de communication d'une part et de son enregistrement, d'autre part, des objets communs pourraient permettre d'établir les fondements d'un travail collectif critique. Le document et l'archive normalisée deviendraient alors la plateforme d'une science partagée.

2. La signifiante algorithmique au service de la recherche

Si l'on définit la signifiante comme le procès d'un travail sur la donnée qui la fait passer de l'information à la connaissance, force est de constater la nécessité d'un cadre théorique qui repose sur le lien entre supports, signes, et information²⁸. Toute l'algorithmisation à la source du document numérique²⁹ subit des recadrages multiples dans le processus de recherche en humanités numériques. Les activités d'enrichissement, de modélisation et de contextualisation par les algorithmes sont une nouvelle série d'épreuves nécessaires au développement du «tournant digital» qui doit fonder les nouvelles compétences du chercheur en SHS.

2.1. L'enrichissement des données ou le recodage

²⁷ Cormerais Franck (2015a), « Humanités digitales et Ré/organisation du savoir », in Le Deuff Olivier (sous la dir. de), *Le temps des Humanités digitales*, Limoges, fyp éditions.

²⁸ Venturini, T. et al. (2016). «Hors champs». *Réseaux*, (199), 11-42.

²⁹ Carmes, M., & Noyer, J.-M. (2014). «L'irrésistible montée de l'algorithmique», *Les Cahiers du numérique*, 10(4), 63-102.

L'activité de recherche traditionnelle traite très rarement des données premières ; elles lui sont souvent fournies. Aussi, il faut un long travail de nettoyage et de recodage de ces données en formes secondes pour pouvoir répondre à une problématique. Et c'est précisément dans ce champ que le traitement algorithmique a connu ses innovations les plus remarquées. Extraire des données complexes des items utiles, par un processus de requalification des données grâce à des outils numériques, n'est pas un phénomène récent. Mais nous constatons un flux d'innovation (plus par diffusion que par invention) rendu nécessaire par l'épreuve de gérer de grandes masses de données. Ces innovations par la multiplication d'outils de recherche ne cessent d'interpeller le chercheur en humanités numériques au point de remettre en cause son portefeuille de compétences. L'opération d'annotation repose sur une grande diversité de méthodes pouvant fonder une science de l'écrit autour d'une herméneutique révisée. L'analyse stylistique des textes, la classification thématique automatique, la lemmatisation des paragraphes, la détection des noms propres, des lieux et places, la classification affective... ne sont que les prémices d'un traitement automatique du texte (ou de la langue) en vue d'une problématique de classification assistée par ordinateur³⁰.

Le traitement des corpus d'images ou de vidéos numérisées connaît également des évolutions significatives. Les recherches avancées en analyse d'image, notamment sous l'impulsion de Google (Google Vision Analysis) et Amazon (Amazon rekognition), permettent la reconnaissance faciale, la localisation géographique, l'analyse des pratiques (comme dans la recherche d'images pornographiques), la présence d'objets grâce à des algorithmes d'identification issus de la recherche en intelligence artificielle³¹. Enfin, la détection des empreintes audio permet aujourd'hui une transcription en texte, une identification de la voix, des contenus similaires, etc. Dans les sciences historiques, la transcription des manuscrits scripturaires en textes numérisés a été un des moteurs de l'innovation dans le mouvement de la reconnaissance d'image en SHS³².

Cette activité de recodage forme l'activité essentielle d'une science de l'écriture. Nous avons ici les bases d'une ingénierie des connaissances dont les orientations peuvent être mises au service d'un nouveau du programme de formation en humanités numériques³³ en permettant effectivement de travailler sur de très grands corpus à faible coût. Il ne s'agit pas ici de plonger dans une «passion algorithmique», mais bien de s'engager dans un processus d'innovation mesuré et prudent remettant en cause les frontières entre qualitatifs et quantitatifs³⁴. Les nouvelles formes d'opérationnalité des hypothèses, difficiles à mettre en œuvre jusque-là, ouvriront par là même de nouvelles formes de problématisation dans un cercle dynamique et créatif.

³⁰ Poibeau, T. (2014). «Le traitement automatique des langues pour les sciences sociales, Automatic language processing for the social sciences», *Réseaux*, (188), 25--51.

³¹ Nous renvoyons le lecteur aux travaux de Lev Manovich (Manovich, 2012) et de son Cultural Analytics Lab. <http://lab.culturalanalytics.info/> (cité in Monnoyer-Smith et Plantin, 2014)

³² Camps, J.-B. (2017). «Homemade manuscript OCR (1): OCRopy [Billet]». Consulté 10 septembre 2018, à l'adresse <https://graal.hypotheses.org/786>; Massot, M.-L., Sforzini, A., & Ventresque, V. (2018). «Transcribing Michel Foucault's reading notes with the Transkribus software: test report», Consulté à l'adresse <https://hal.archives-ouvertes.fr/hal-01794139>

³³ Cormerais, F., Le Deuff, O., Lakel, A. & Pucheu, D. (2017). «L'école et l'avenir de la culture digitale». *Hermès, La Revue*, 78,(2), 87-95.

³⁴ Rebillard, F. (2011). «L'étude des médias est-elle soluble dans l'informatique et la physique ? À propos du recours aux digital methods dans l'analyse de l'information en ligne», *Questions de communication*, (20), 353--376.

2.2. La modélisation des algorithmes : régression, classification et clusterisation

La fonction des algorithmes est d'aider à interpréter les données par un principe d'automatisation en étudiant les corrélations statistiques entre les variables. Cette activité forme un des éléments avantageux de la recherche. L'implémentation récente des logiciels prédictifs (que nous préférons appeler «interprétatifs») en briques logiciels permet l'analyse des relations entre plusieurs variables afin de les classer et de les lier en des catégories. Dans ce foisonnement d'innovation, il existe trois groupes d'algorithmes qui sont présentés dans le tableau suivant :

	objectif
Algorithme de régression	modéliser en équation la relation d'une variable avec plusieurs variables sur la base de leur variabilité concomitante.
Algorithme de classification	établir une relation statistique sur la base des coprésences de modalités de variables entre les réponses à certaines questions et la réponse à une question précise en analysant le poids de chacune des relations explicatives
Algorithmes de clustering	Ranger les réponses à une série de variables en structures récurrentes afin de faire apparaître des régularités entre les réponses d'une enquête sous forme de typologie formelle ou «profil»

Tableau n°1 : « Typologie des algorithmes interprétatifs »

Le progrès de la recherche en intelligence artificielle dans le traitement des variables permet aujourd'hui de travailler avec une immense quantité de réponses sur de nombreuses dimensions (nombre de variables très grand). Les langages open source comme R ou Python en facilitent l'usage à condition que la communauté des chercheurs puissent accéder à des formations adéquates. Il ne s'agit pas ici de transformer chaque chercheur en SHS en informaticien, mais de le doter d'une culture suffisamment solide pour en faire un utilisateur critique plus ou moins engagé dans l'usage de ces algorithmes. Les critiques se sont basées plus sur les limites d'une idéologie messianique que sur une véritable évaluation de leurs usages³⁵. Mais l'émergence des chercheurs en humanités numériques permet de pouvoir appréhender les nouveaux usages d'une façon plus constructive. Les recherches récentes en SIC et en sociologie du web ont montré la pertinence de l'usage de ces algorithmes. Dans l'étude de l'ANR Algotopol, citée

³⁵ Cardon, D. (2015). *A quoi rêvent les algorithmes*, Paris: Le Seuil.

précédemment, l'utilisation du K-Mean, un algorithme de clusterisation, a permis de dégager une typologie comportementale de l'éditorialité sur Facebook très riche en informations³⁶. Plus généralement, la massification des données numériques rend nécessaire l'usage de ces outils pour en tirer une classification préalable à l'exploration assistée.

2.3. La contextualisation structurale des documents : des relations réelles aux relations logiques

Avec l'apparition des études digitales, l'hypertexte fut mis à l'honneur dans la littérature grise de SHS. La nature interreliée des documents renvoyait à la tradition bibliologique d'étude de la citation qui puise ses sources dans la scientométrie. Le document est contextualisé, tant en matière sémiotique qu'au niveau sémantique, par des relations qui le «situent» à l'intérieur d'un corpus. Au tournant des années 90, le réseau est devenu une figure à la fois paradigmatique et heuristique de la société³⁷. L'explosion du web a nourri l'impérieuse nécessité de le mettre au goût du jour. Depuis 2009, des logiciels comme Gephi des frères Jacomy ont fait sortir de la confidentialité, en France, les méthodes en analyse structurale des réseaux. Ce logiciel lui-même est le résultat d'un engouement majeur, outre-Atlantique, pour des méthodes opérationnelles d'analyse des réseaux.

Si les cartographies de liens ont régulièrement impacté les démarches scientifiques, c'est souvent au détriment d'une véritable analyse de réseaux. À la base, il s'agit d'une structure en double enquête sur, d'une part, une population d'individus et d'autre part sur un relevé de leurs relations (des pages et leurs hyperliens, des entreprises et leurs transactions, des individus et leurs interactions). L'analyse d'un réseau est une méthode statistique qui développe des algorithmes pour traiter les liens de chaque nœud, afin de décrire sa structure. Ces algorithmes sont de deux natures : ceux qui informent sur le réseau dans une perspective holiste et comparatiste, ceux qui informent sur la place du nœud dans un réseau donné. Les sciences de l'information et de la communication utilisent ces méthodes notamment dans la réflexion qu'elles portent sur l'hypertextualité³⁸. En étudiant les pages web traitant des humanités numériques et digitales, l'analyse holiste du réseau a permis d'appréhender la structure d'une communauté savante et les facteurs pesant sur la socialisation des auteurs³⁹. L'étude de l'ANR Algopol est encore plus emblématique, car elle couple l'usage d'un algorithme interprétatif pour comprendre les comportements et une analyse des réseaux pour situer les acteurs⁴⁰.

3. Un horizon des réseaux de chercheurs est-il possible ?

³⁶ Bastard, I., Cardon, D., Charbey, R., Cointet, J. & Prieur, C. (2017). «Facebook, pour quoi faire : Configurations d'activités et structures relationnelles». *op.cit.*

³⁷ Musso, P. (2003b). *Réseaux et société*. Paris: Presses Universitaires de France - PUF.

³⁸ Koszowska-Nowakowska, P., & Renucci, F. (2012), «L'hypertexte : approches expérimentale et herméneutique», *Les Cahiers du numérique*, Vol. 7(3), 71--91; Saemmer, A. (2015). «Hypertexte et narrativité». *Critique*, n° 819-820(8), 637-652.

³⁹ Lakel, A., & Le Deuff, O. (2016). «À quoi peut bien servir l'analyse du web ?», *op.cit.*

⁴⁰ Bastard, I., Cardon, D., Charbey, R., Cointet, J. & Prieur, C. (2017). «Facebook, pour quoi faire : Configurations d'activités et structures relationnelles». *op.cit.*

La tradition des SHS depuis la fin du 19^{ème} siècle fut souvent le résultat d'une recherche personnelle originale pouvant, seulement ensuite, faire «école» en empruntant des canaux particuliers (revues, sociétés savantes, congrès, etc.). Pour autant, nous avons vu que face à l'architecture digitale, la somme des compétences et des techniques en jeu ne permettait plus le travail d'un seul⁴¹. Cette situation est-elle en passe de changer ? Sommes-nous, face aux enjeux de la donnée et de la méthode digitale, en présence d'une injonction du collectif de recherche en SHS, seul garant d'un développement «contrôlé»⁴².

3.1. Mathématiques, Informatique et SHS autour d'un enjeu partagé : comprendre l'humain

La dernière épreuve des humanités digitales est d'ordre social. Si les informaticiens et les statisticiens ont leurs propres préoccupations, ces trois dernières décennies, les sciences (et les écoles) de l'ingénieur ont su faire émerger une approche hybride orientée vers la résolution de problèmes. Comprendre l'humain a toujours été une préoccupation d'une partie des chercheurs en sciences informatiques et mathématiques. De la cybernétique au cognitivisme, ces chercheurs nous ont montré la capacité des mathématiques et de l'informatique à penser l'humain et la société. Franck Rebillard⁴³ nous rappelle qu'encore récemment, avant même l'émergence des digital methods, ce sont bien les informaticiens et les mathématiciens qui se sont engagés à transformer les champs d'études des SHS. Avons-nous encore les moyens de retarder cette alliance des compétences aux risques de l'échec de nos projets d'humanités numériques⁴⁴ ? Sans une incitation majeure d'une politique de la science engageant les ressources et les formations, a-t-on une chance d'assurer le passage des débats a priori à une série de projets construisant des sciences en action ?

Pourtant comme le souligne Jean-Guy Meunier⁴⁵, la modélisation scientifique révèle une très grande proximité entre les sciences qu'elles soient humaines ou des matériaux et du vivant. Le processus de modélisation formelle, matérielle et conceptuelle qu'il propose pour les humanités numériques souligne la proximité entre les différentes disciplines. L'engagement du chercheur en SHS dans les processus de traduction numérique pose la question de la «distance informationnelle» médiatisée⁴⁶ que le chercheur ne manquera pas d'avoir avec la gestion informatique de son corpus numérique. Franck Rebillard, dans la critique qu'il porte sur la recherche des mêmes dans un grand

⁴¹ Imad Saleh et Hakim Hachour, « Le numérique comme catalyseur épistémologique », *Revue française des sciences de l'information et de la communication* [En ligne], 1 | 2012, mis en ligne le 25 juillet 2012, consulté le 24 mars 2019. URL : <http://journals.openedition.org.ezproxy.u-bordeaux-montaigne.fr/rfsic/168> ;

⁴² Rebillard, F. (2011,). «L'étude des médias est-elle soluble dans l'informatique et la physique ? À propos du recours aux digital methods dans l'analyse de l'information en ligne», *Questions de communication*, (20), 353--376.

⁴³ Ibid.

⁴⁴ Bourdaloie, H. (2014). «Ce que le numérique fait aux sciences humaines et sociales. épistémologie, méthodes et outils en questions». *Tic&société*, (Vol. 7, N° 2).

⁴⁵ Meunier, J.-G. (2017). «Humanités numériques et modélisation scientifique», *Questions de communication*, (31), 19-48.

⁴⁶ Rieder, B. (2010,). «Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention», *Études de communication. langages, information, médiations*, (35), 91--104.

corpus, montre bien la limite d'une vision purement informatique de la traduction en SHS. Le simple usage de Google dans la constitution d'un corpus numérique place le chercheur dans la nécessité d'inventer, voire de bricoler des voies d'accès aux données qui sont autant de négociations avec une modélisation informatique d'une pratique non destinée à l'approche scientifique. Une fois passés les évidents appels à la prudence scientifique, il faut bien nous engager dans la rénovation des sciences sociales⁴⁷ par une pratique innovante de nouveaux agencements scientifiques.

3.2. Des formations au service du nouveau lettré du digital

La formation au code informatique connaît une longue histoire dans les institutions scolaires. Au cœur de cette politique publique, une tension très forte entre un discours d'autonomisation des élèves et une politique d'équipement et de soutien aux industriels du secteur numérique. Entre une vision de formation des citoyens à la puissance du code produisant, au cœur d'outils modulables et open sources, des œuvres collaboratives et celle d'un étudiant compétent dans l'usage des boîtes noires des solutions ergonomiques du marché au service de son insertion professionnelle, la question de la littératie numérique⁴⁸ est au cœur d'une problématisation conflictuelle du rôle de la formation dans les sociétés numériques. Du Plan Informatique Pour Tous (IPT) en 1985 au C2I des universités en 2002, on a vu émerger le mythe du digital native qui n'a besoin que d'un accompagnement professionnalisant dans une supposée maîtrise des usages numériques qu'il aurait acquis spontanément dans une maison fortement dotée. Pourtant cette logique, entièrement soumise à la compétence professionnelle et à l'employabilité, a fait le deuil, si ce n'est a enterré, la figure du lettré numérique capable de passer de la lecture à la production du sens, de l'usage à la critique, de la compétence à la programmation dans un nouvel environnement de production de sens où il serait apte à interroger sa place. Compréhension, déconstruction, programmation, production des nouveaux schèmes de connaissances à l'heure des dispositifs numériques n'ont guère avancé alors même que l'on fête joyeusement la révolution internet qui a, en 2019, 30 ans.

Que ce soit par une approche de vulgarisation mutuelle ou dans une approche projet, nous faisons le pari que les conditions d'une véritable hybridation disciplinaire se fonderont sur l'innovation des programmes de formations interdisciplinaires au cœur des écoles d'ingénieurs et des écoles doctorales en SHS⁴⁹. En relation avec la formation d'un «lettré du digital»⁵⁰, il est possible de décliner l'importance de la formation des chercheurs en SHS autour de trois recommandations qui pourraient bien former un cadre pour les sciences de l'écrit.

⁴⁷ Boullier, D. (2015). «Les sciences sociales face aux traces du big data, Big Data traces and social sciences». *Revue française de science politique*, 65(5), 805-828.

⁴⁸ Deuff, O. L. (2012). «Humanisme numérique et littératies.» *Semen - Revue de sémio-linguistique des textes et discours*, (34), 117-134.

⁴⁹ Wolff, É. (2014). «Cheminement sur les terres inconnues des humanités numériques. À la recherche de la e-diaspora réunionnaise» *Tic&société*, (Vol. 7, N° 2).

⁵⁰ Cormerais, F. et al. (2017). «L'école et l'avenir de la culture digitale». *Hermès, La Revue*, 78,(2), 87-95.

- Recommandation n° 1 : Appuyer une approche conceptuelle (interprétative et dialogique) en rapport avec une approche fonctionnelle évite le danger d'un primat du «code informatique » sur le langage naturel. L'insertion du numérique au cœur des projets SHS doit être l'occasion d'interroger le code informatique dans ce qu'il fait au processus de véridicité.
- Recommandation n° 2 : Favoriser les rencontres matérielles autour des algorithmes interprétatifs. Des avancées sont possibles avec des réseaux qui associent, dans l'expérimentation, les chercheurs de différentes disciplines. L'architecture distribuée open source incarne une démarche de coopération et d'alternative critique qui marche. Le développement, et le financement de prototype ouvert peut servir de base à cette approche critique pragmatique.
- Recommandation n° 3 : Passer des emplois « occultes » des technologies dans la recherche, sur un modèle de sous-traitance, à un emploi réfléchi par une maîtrise de pratiques intégrées dans un projet pensé avant tout comme un collectif basé sur des transferts de compétences et du co-apprentissage.

Au sein de l'université, les URFIST⁵¹ peuvent jouer un rôle fondamental dans la montée en compétence des chercheurs en SHS par un travail d'adaptation de leur programme de formations aux enjeux des humanités numériques. Les humanités digitales ne peuvent s'imaginer sans la formation de chercheurs capables de naviguer d'une formalisation mathématique à une formalisation conceptuelle en passant par une formalisation informatique⁵².

3. 3. *L'écrit comme base d'un discours commun à-venir*

Nous proposons alors de partir de la recherche d'un langage commun à partir de l'écriture qui est un objet transversal à toutes les démarches scientifiques. Les sciences de l'écrit possèdent un passé qui consistait à repositionner l'écriture dans la perspective de la communication : « L'écrit, mémoire objective de l'écrivain est perçu, décodé, compris par l'acte de lecture. La transformation de l'objet écrit en document réel, subjectivise à nouveau le message. L'acte de communication, de mise en commun se réalise. Mais cette transformation de l'écrit en document introduit une nouvelle notion : l'information. L'information, c'est le contenu, le sens, la connaissance mise à l'intérieur d'une forme. On comprend mieux pourquoi la mémoire fixée dans le médium, devenant document, fournit une information. »⁵³ Aujourd'hui, une nouvelle synthèse est nécessaire, dès lors que l'acte de lecture se fait en fonction d'un stock énorme de formes et de connaissances mémorisées sur les supports que sont les serveurs et les bases de données. Le processus initié par le «tournant digital» dans la communication écrite associe un circuit informationnel auquel viennent se superposer un circuit interprétatif et un circuit dialogique.

Les sciences de l'écrit partent du document pour comprendre les conditions de possibilités de son existence. Production, édition, distribution et usages, les sciences de

⁵¹ Unités Régionales de Formation à l'Information Scientifique et Techniques

⁵² Meunier, J.-G. (2017). «Humanités numériques et modélisation scientifique», *Questions de communication*, (31), 19-48.

⁵³ Estivals, R. (1987). *La Bibliologie*, Presses Universitaires de France.

l'écrit ont toujours développé un art exégétique pour reconstruire le processus complet de la communication. Avec l'émergence du document numérique, de nombreux chercheurs en SHS ont su dépasser la stupéfaction sociotechnique pour réinterroger l'objet des humanités numériques à l'aune de ce nous voulons rappeler comme science de l'écrit. De la documentarisation⁵⁴ à l'extension progressive du concept d'éditorialisation⁵⁵, la recherche sur le document numérique montre la nécessité d'appréhender le document dans toute l'épaisseur de son existence⁵⁶. Il s'agit alors de comprendre que pour chaque document numérique il faut, sur un premier axe, interroger tout le processus de production de document tout en déclinant cette description sur un second axe, au regard d'une contextualisation technologique, sémiotique, économique, sociologique et politique. Or notre thèse consiste à rappeler que cette approche est le fondement même des sciences de l'écrit à l'origine de la création des Sciences de l'Information et de la Communication hantées par le fantôme de «la méthode Foucault». Les humanités digitales seraient alors comprises comme un nouveau moment des sciences de l'écrit.

Dès l'origine, cette «position» a favorisé de nouvelles alliances entre les démarches quantitatives et qualitatives. Elle permet de s'orienter vers une « archi-écriture » qui rassemblerait les dimensions suivantes : paléographique, grammatologique, philologique, informatologique. Cette approche suppose un dialogue entre des méthodologies endogènes et exogènes. C'est dans ce croisement des méthodes que se trouve remis en question le blocage épistémologique suivant : « L'introduction de la statistique dans les sciences humaines a été l'une des conditions de leur essor aux 19e et 20e siècles. Néanmoins la statistique n'élimine pas totalement tant s'en faut, la subjectivité. Celle-ci intervient au départ dans les présuppositions et les hypothèses, et à l'arrivée, après l'observation, au niveau des interprétations.»⁵⁷ La différence pointée entre sciences humaines et l'approche statistique, sans demeurer invalide, se rétrécit lorsque l'alliance du logiciel et du matériel permet une plus grande facilité de la calculabilité manipulatoire. C'est de la synthèse entre une imagination créatrice et un solide usage de la statistique que peut naître une transdisciplinarité qui, si elle demeure à l'ordre du jour, est assez difficile à réaliser, car peu perméable aux habitudes institutionnelles universitaires.

Conclusion

⁵⁴ Zacklad M. (2007), «Espace documentaire participatif et gouvernance» *Congress of the European Regional Science Association* (47th Congress) and ASRDLF (Association de Science Régionale de Langue Française, 44th Congress) PARIS - August 29th - September 2nd, 2007., Aug 2007, France. 2007; Zacklad, M. (2012). «Organisation et architecture des connaissances dans un contexte de transmédia documentaire : les enjeux de la pervasivité», *Études de communication. langages, information, médiations*, (39), 41-63.

⁵⁵ Pour une plongée dans ce concept, lire : Souchier E, Emmanuel. (1998). «L'image du texte pour une théorie de l'énonciation éditoriale». *Les cahiers de médiologie*, 6(2), 137; Souchier E, Emmanuel. (2007),). «Formes et pouvoirs de renonciation éditoriale», *Communication et langages*, 154(1), 23--38. Jeanneret, Y., & Souchier, E. (2005). «L'énonciation éditoriale dans les écrits d'écran», *Communication et langages*, 145(1), 3--15; Bachimont, B. et al. (2011),). «Enjeux et technologies : des données au sens.» *Documentaliste-Sciences de l'Information*, 48(4), 24--41; Merzeau Louise, (2014). «Entre événement et document : vers l'environnement-support», *Les Cahiers de la SFSIC*, Société française des sciences de l'information et de la communication, 2014, pp.230-233; Vitali Rosati, M. (2016). «Qu'est-ce que l'éditorialisation ?», *Sens Public*. Consulté à l'adresse <http://sens-public.org/article1184.html>

⁵⁶ Pour une bibliographie assez exhaustive de la problématique en français voir le travail de Jérôme Valluy : Valluy, Jérôme. (2015, mai 18). Editorialisation - Terra-HN. Consulté 10 septembre 2018, à l'adresse <http://www.reseau-terra.eu/article1333.html>

⁵⁷ Estivals, R. (1987). *La Bibliologie, op.cit.*, p.116.

Les sciences de l'information et de la communication connaissent, depuis 20 ans, un «digital turn» qui s'exprime avec intensité dans la problématisation des humanités digitales. Face à la numérisation généralisée du quotidien, le chercheur en SHS se retrouve devant un océan de «données» qui forment un potentiel immense dans la compréhension du fait social par la communication. Pour autant, cet océan de données est tout sauf donné. Il existe des limites techniques à l'interopérabilité, des limites économiques d'accès à ces données, souvent propriétés privées jalousement gardées, limite symbolique du mythe de la data sans l'information. La question d'une véritable politique publique de la donnée au coeur de la recherche en SHS et en SIC s'inscrit est à un tournant avec les initiatives d'open datas et de dataverse. Notre réflexion se veut dans un premier temps une recommandation pour embrasser la problématique de l'accès aux données au service de la recherche. Doter les chercheurs d'outils de récoltes et de production de la data, faciliter la contractualisation public-privé d'accès aux corpus, ouvrir un vaste chantier d'accès aux données publiques de la recherche interopérable. La promesse des humanités numériques de demain passe avant tout par la question politique du statut de la donnée. La figure du «lettré digital» relève le défi, face à l'épreuve quantitative, de la question de l'accès à la méthodologie algorithmique en environnement big data et à l'outillage qui le rend accessible aux chercheurs en SHS. Ce lettré d'un nouveau type veut relever le défi d'une interrogation du sens assisté par ordinateur et se donne les moyens d'accéder à des processus interprétatifs actuels, de les critiquer voire de les améliorer en participant à la construction de nouveaux dispositifs de véridicité dans l'épistémé des humanités numériques. Enfin, nous affirmons que la dernière épreuve, qui permettra aux humanités numériques de dépasser les discours métascientifiques, consiste à faire tomber les frontières entre disciplines mathématiques, informatiques et sciences du vivant, d'une part, et sciences humaines et sociales, d'autre part. Autour de problématiques et de projets communs, des chercheurs de différents horizons sont en capacité d'échanger leurs objets et leurs méthodes par un transfert de compétences. En fait, les plus grands tournants épistémologiques des SHS ne sont-ils pas souvent le fruit d'une hybridation d'avec les sciences «dures»? Au début du siècle, les mathématiques rencontraient les humanités pour fonder les sciences sociales (sociologie, économie et psychologie). Dans l'après-guerre, l'informatique rencontrait les humanités pour fonder la systémique et le connexionnisme. Aux fondements des sciences de l'information et de la communication, on retrouve cette interdisciplinarité entre science de la quantification et science de l'interprétation⁵⁸. Robert Escarpit et Robert Estival, dans une «École de Bordeaux», trop vite oubliée, nous avaient enjoins d'opérer la grande jonction autour des sciences de l'écrit. Passé l'effet de surprises, nous gageons que les humanités numériques vont [se sont déjà] s'engager dans le «digital turn» en multipliant les projets pragmatiques. Cette profusion d'expérimentation est une période nécessaire à l'innovation scientifique. Certains y voient l'émergence d'une «zone d'échange» favorable à la créativité scientifique⁵⁹. À

⁵⁸ Meyriat J., Miège B., 2002, « Le projet des sic : de l'émergent à l'irréversible (milieu des années 60-milieu des années 80) », pp. 45-70, in : Boure R., *Les origines des sciences de l'information et de la communication. Regards croisés*, Lille, Presses universitaires du Septentrion.

⁵⁹ Svensson P. (2010). «The Landscape of Digital Humanities», *Digital Humanities Quarterly*, 4(1). Consulté à l'adresse <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>

l'ère du numérique, les défis sont immenses, mais ils sont à l'échelle des épreuves qui s'annoncent pour les humanités et pour les institutions universitaires qui les défendent.

Bibliographie

Akrich, M., Callon, M., & Latour, B. (2006). *Sociologie de la traduction : Textes fondateurs* (1re éd.). Paris: Transvalor - Presses des mines.

Bruno Bachimont B., (2007), «Nouvelles tendances applicatives. De l'indexation à l'éditorialisation», dans Patrick Gros (dir.), *L'indexation multimédia : description et recherche automatiques*, Paris, Hermès Sciences., 2007

Bachimont, B., Poupeau, G., Vatan, B., Troncy, R., Gandon, F., Pouyllau, S., ... Zacklad, M. (2011), «Enjeux et technologies : des données au sens.» Documentaliste-Sciences de l'Information, 48(4), 24--41.

Bastard, I., Cardon, D., Charbey, R., Cointet, J. & Prieur, C. (2017). «Facebook, pour quoi faire : Configurations d'activités et structures relationnelles». *Sociologie*, vol. 8,(1), 57-82.

Ben Henda, M. (2015). Adaptation normative des Big Data et du Learning Analytics. Dans : Évelyne Broudoux éd., *Big Data - Open Data : Quelles valeurs ? Quels enjeux: Actes du colloque « Document numérique et société »*, Rabat, 2015 (pp. 197-212). Louvain-la-Neuve, Belgique: De Boeck Supérieur. doi:[10.3917/dbu.chron.2015.01.0197](https://doi.org/10.3917/dbu.chron.2015.01.0197).

Ben Henda, M. & Hudrisier, H. (2013). Penser, classer, apprendre et communiquer. Normalisation et nouveaux modes de classification du savoir. *Hermès, La Revue*, 66(2), 160-166. <https://www-cairn-info.ezproxy.u-bordeaux-montaigne.fr/revue-hermes-la-revue-2013-2-page-160.htm>.

Bouchardon, S., Cailleau, I., Crozat, S., Bachimont, B., & Thibaud, H. (2011). «Explorer les possibles de l'écriture multimédia», *Les Enjeux de l'information et de la communication*, (12/2), 11-24.

Boullier, D. (2015). «Les sciences sociales face aux traces du big data, Big Data traces and social sciences». *Revue française de science politique*, 65(5), 805-828.

Bourdeloie, H. (2014). «Ce que le numérique fait aux sciences humaines et sociales. épistémologie, méthodes et outils en questions». *tic&société*, (Vol. 7, N° 2).

Boure, R. (2002). *Les origines des sciences de l'information et de la communication. Regards croisés*. Villeneuve-d'Ascq: Presses Universitaires du Septentrion.

Camps, J.-B. (2017). «Homemade manuscript OCR (1): OCRopy [Billet]». Consulté 10 septembre 2018, à l'adresse <https://graal.hypotheses.org/786>

Cardon, D. (2013b). «Dans l'esprit du PageRank» *Réseaux*, (177), 63-95.

- Cardon, D. (2015). *A quoi rêvent les algorithmes*, Paris: Le Seuil.
- Carmes, M., & Noyer, J.-M. (2014). «L'irrésistible montée de l'algorithmique», *Les Cahiers du numérique*, 10(4), 63-102.
- Cormerais Franck (2015a), « Humanités digitales et Ré/organisation du savoir », in Le Deuff Olivier (sous la dir. de), *Le temps des Humanités digitales*, Limoges, fyp éditions.
- Cormerais Franck (2015b), « Humanités digitales, transdisciplinarité et entretrextualité. Éléments pour une organisation des corpus partagés », in Carayol Valérie et Morandi Franck (sous la dir. de), *Le tournant numérique des sciences humaines et sociales*, Bordeaux, Maison des sciences d'Aquitaine, p. 55-65.
- Cormerais, F., Le Deuff, O., Lakel, A., & Pucheu, D. (2016). «Les SIC à l'épreuve du digital et des Humanités: des origines, des concepts, des méthodes et des outils». *Revue française des sciences de l'information et de la communication*, (8).
- Cormerais, F., Le Deuff, O., Lakel, A. & Pucheu, D. (2017). «L'école et l'avenir de la culture digitale». *Hermès, La Revue*, 78,(2), 87-95.
- Cormerais, Franck, & Le Deuff, Olivier. (2018). «Robert Estivals, entre bibliométrie et bibliologie, quel apport pour les SIC dans la recomposition des savoirs», dans *Fondements épistémologiques et théoriques de la science de l'information-documentation* (ISTE Editions, p. 311).
- Dacos, M., & Mounier, P. (2015). *Humanités numériques*. Institut français. Rapport consulté à l'adresse <https://hal.archives-ouvertes.fr/hal-01228945/document>
- Escarpit, R. (1995). *L'information et la communication. Théorie générale*, Paris, Hachette.
- Estivals, R. (1987). *La Bibliologie*, Presses Universitaires de France.
- Estivals, R., & Meyriat, J. (Éd.). (1993). *Les sciences de l'écrit: encyclopédie internationale de bibliologie*. Paris: Retz.
- Ferraris, M. (2012),). *Documentality: Why It Is Necessary to Leave Traces*, New York: Fordham University Press.
- Genette, G. (2004). *Fiction et diction, précédé de Introduction à l'architexte*. Paris, Seuil.
- Hartog, F. (2003). *Régimes d'historicité : Présentisme et expériences du temps*, Paris: Seuil.
- Jeanneret, Y., & Souchier, E. (2005). «L'énonciation éditoriale dans les écrits d'écran», *Communication et langages*, 145(1), 3--15.
- Koszowska-Nowakowska, P., & Renucci, F. (2012),). «L'hypertexte : approches expérimentale et herméneutique», *Les Cahiers du numérique*, Vol. 7(3), 71--91.

Lakel Amar (2007) « Interconnexion et identités électroniques : le pouvoir d'enquête à l'heure des systèmes de fichiers répartis » in Terminal, dossier spécial « Administration électronique », 2007.

Lakel, Amar. (2014a). «Enjeux de la recherche digitalisée dans les humanités numériques», My Web Intelligence. Consulté à l'adresse <http://mywebintelligence.net/enjeux-de-la-recherche-digitalisee-dans-les-humanites-numeriques/>

Lakel, Amar. (2014b). «My Web Intelligence : Un projet d'outils open source au service des humanités numériques.» My Web Intelligence. Consulté à l'adresse <http://mywebintelligence.net/outils-open-source-humanites-numeriques/>

Lakel, A., & Le Deuff, O. (2016). «À quoi peut bien servir l'analyse du web ?», Les Cahiers du numérique, 13(3), 39-62.

Le Deuff, O. (2012). "Humanisme numérique et littératies." Semen - Revue de sémiolinguistique des textes et discours, (34), 117-134.

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). «Meme-tracking and the dynamics of the news cycle». In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09 (p. 497). Paris, France: ACM Press.

Lucas, J.-F. (2014). «Le projet Magic Ring : expérimentation d'une méthode de recueil de données quali-quantitatives dans Second Life», Tictic&société, (Vol. 7, N° 2).

Massot, M.-L., Sforzini, A., & Ventresque, V. (2018). «Transcribing Michel Foucault's reading notes with the Transkribus software: test report», Consulté à l'adresse <https://hal.archives-ouvertes.fr/hal-01794139>

Merzeau Louise, (2014). «Entre événement et document : vers l'environnement-support», Les Cahiers de la SFSIC, Société française des sciences de l'information et de la communication, 2014, pp.230-233.

Meunier, J.-G. (2017). «Humanités numériques et modélisation scientifique», Questions de communication, (31), 19-48.

Meyriat J. (1981). « Document, documentation, documentologie ». Schéma et schématisation, 2e trimestre, no 14, p. 51-63.

Meyriat J., Miège B., 2002, « Le projet des sic : de l'émergent à l'irréversible (milieu des années 60-milieu des années 80) », pp. 45-70, in : Boure R., Les origines des sciences de l'information et de la communication. Regards croisés, Lille, Presses universitaires du Septentrion.

Monnoyer-Smith, L. & Plantin, J.-C.(2014). «Ouvrir la boîte à outils de la recherche numérique. Trois cas de redistribution de méthodes», tic&société, (Vol. 7, N° 2).

Musiani, F. et al. (2019). Qu'est-ce qu'une archive du web ? Marseille: OpenEdition Press. Consulté à l'adresse <http://books.openedition.org/oep/8713>

Musso, P. (2003a). Critique des réseaux. Paris: Presses Universitaires de France - PUF.

Musso, P. (2003b). Réseaux et société. Paris: Presses Universitaires de France - PUF.

Paloque-Berges, C. & Schafer, V. (2015). «Quand la communication devient patrimoine». *Hermès, La Revue*, (71), 255--261.

Poibeau, T. (2014). «Le traitement automatique des langues pour les sciences sociales, Automatic language processing for the social sciences», *Réseaux*, (188), 25--51.

Proulx, S., & Rueff, J. (2018). Actualité des méthodes de recherche en sciences sociales sur les pratiques informationnelles. Consulté à l'adresse http://epe.lac-bac.gc.ca/100/200/300/univ_laval/centre_etudes_medias/CEM-ActualiteMethode.pdf

Pucheu, D. (2016). «Effacer l'interface: Une trajectoire du design de l'interaction homme-machine» *Interfaces numériques*, 5(2), 257--276.

Rebillard, F. (2011),). «L'étude des médias est-elle soluble dans l'informatique et la physique ? À propos du recours aux digital methods dans l'analyse de l'information en ligne» *Questions de communication*, (20), 353--376.

Rebillard, F., Fackler, D., & Marty, E. (2012),). «L'offre d'informations est-elle plus diversifiée sur le web qu'à la télévision ?», *Réseaux*, (176), 141--172.

Rieder, B. (2010),). «Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention», *Études de communication. langages, information, médiations*, (35), 91--104.

Saemmer, A. (2015). «Hypertexte et narrativité». *Critique*, n° 819-820(8), 637--652.

Saleh, I., & Hachour, H. (2012). «Le numérique comme catalyseur épistémologique». *Revue française des sciences de l'information et de la communication*, (1).

Serres, M. (2015). *Le gaucher boiteux, - Figures de la pensée*. Paris: LE POMMIER.

Severo, M., & Venturini, T. (2016),). «Enjeux topologiques et topographiques de la cartographie du Web», *Réseaux*, (195), 85--105.

Souchier E, Emmanuel. (1998). «L'image du texte pour une théorie de l'énonciation éditoriale». *Les cahiers de médiologie*, 6(2), 137.

Souchier E, Emmanuël. (2007),). «Formes et pouvoirs de renonciation éditoriale», *Communication et langages*, 154(1), 23--38.

Spitzer, L. (1980). *Études de style*. Paris: Gallimard.

Svensson P. (2010). «The Landscape of Digital Humanities», *Digital Humanities Quarterly*, 4(1). Consulté à l'adresse <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>

Venturini, T., Jacomy, M., Baneyx, A., & Girard, P. (2016). «Hors champs». *Réseaux*, (199), 11-42.

Vitali Rosati, M. (2016). «Qu'est-ce que l'éditorialisation ?» *Sens Public*. Consulté à l'adresse <http://sens-public.org/article1184.html>

Vitali-Rosati, M., & Sinatra, M. E. (Éd.). (2014). *Pratiques de l'édition numérique*. Montréal: Presses de l'Université de Montréal. Consulté à l'adresse <http://books.openedition.org/pum/306>

Wolff, É. (2014). «Cheminement sur les terres inconnues des humanités numériques. À la recherche de la e-diaspora réunionnaise» *tic&société*, (Vol. 7, N° 2).

Zacklad M. (2007),. «Espace documentaire participatif et gouvernance» Congress of the European Regional Science Association (47th Congress) and ASRDLF (Association de Science Régionale de Langue Française, 44th Congress) PARIS - August 29th - September 2nd, 2007,., Aug 2007, France. 2007

Zacklad, M. (2012). «Organisation et architecture des connaissances dans un contexte de transmédia documentaire : les enjeux de la pervasivité», *Études de communication. langages, information, médiations*, (39), 41-63.