Histoire de la construction européenne et mise en données du monde

Frédéric Clavert

C2DH, Université du Luxembourg. frederic.clavert@uni.lu

## Introduction: mise en données du monde, mise en données de l’histoire

Nous vivons dans un monde mis en données. Issu d’un phénomène de long terme – Newton songeait déjà à trouver des lois de la société par la quantification, qui est l’un des aspects de la mise en données –, cette mise en données ou *datafication* ne peut laisser l’historien ou l’historienne de marbre : de nombreux aspects de nos vies sont désormais stockées sous forme de données, au sein de fermes de serveur partout dans le monde, documentant potentiellement des phénomènes jusu’alors difficiles à saisir. Si les éléments nés numériques – tweets, archives du web, documents administratifs ou d’entreprises directement archivés sous leur forme numérique – intéressent plutôt les sociologues dans un premier temps, les programmes de numérisation, massifs (Gallica, Google Books) ou non massifs (citons ici un exemple touchant à l’histoire de l’intégration européenne, le *Centre Virtuel de la Connaissance sur l’Europe*), d’archives « papier » a sans aucun doute des conséquences sur notre pratique de la discipline historique, sur ce « goût de l’archive » qui nous est si important.

Numériser n’est toutefois que la première étape de la mise en données. Suivent alors d’autres étapes – l’amélioration de l’accès aux archives, par exemple – qui, toutes, ont une double caractéristique : une perte (le support physique, le contexte) et un gain (la possibilité de rendre le contenu des sources primaires calculable). La conjonction de l’apparition de l’ordinateur personnel (fin des années 1970), de la mise en réseau (fin des années 1980 avec internet, puis la création du web) et de la production massive de données (notamment mais pas uniquement l’apparition du *smartphone*) a non seulement abouti à transformer notre monde définitivement en monde de données, mais a accéléré la calculabilité de tout.

Les conséquences pour l’historien et l’historienne sont larges. Je vais me concentrer dans ce chapitre sur l’histoire de l’intégration européenne, sur la base de mes recherches, sans toutefois m’interdire ni de faire appel aux recherches des autres, ni de sortir du domaine de l’intégration européenne.

## Sources primaires et mise en données

Il n’y a pas d’histoire sans source. L’historiographie de l’intégration européenne, comme nombre d’autres historiographies, s’est d’ailleurs développée au fil des politiques d’accès aux archives des institutions européennes et, ainsi, à leurs centres d’archives. La compréhension de ces derniers est parfois complexe : chaque institution peut en créer un, tout en devant, théoriquement, verser à terme leurs archives aux Archives historiques de l’Union européenne. Pour « faire » de l’histoire de l’intégration européenne, l’historien ou l’historienne d’aujourd’hui peut égalemnt compter sur des projets de numérisation des sources primaires. D’*Archive of european integration* de l’université de Pittsburgh au *Centre Virtuel de la Connaissance sur l’Europe*, aujourd’hui dépendant du *Centre for Contemporary and Digital History* (C2DH) en passant bien entendu par les centres d’archives des institutions européennes, les sources en ligne sont abondantes. Leur usage, toutefois, doit appeler quelques remarques. En outre, ces archives « nées papier » ne doivent pas faire oublier les archives dites nativement numériques.

### Sources primaires numérisées

Un historien de l’intégration européenne peut aujourd’hui lire, écouter ou regarder des interviews d’acteurs et témoins de la construction européenne sur le site web du CVCE[[1]](#footnote-2) ou des archives historiques de l’Union européenne.[[2]](#footnote-3) Il pourra trouver des éléments iconographiques et audiovisuels sur le site de la Commission - comme un « stockshot » sur le paquet Delors II[[3]](#footnote-4) et auprès des médiathèques des diverses institutions européennes de manière générale. La Banque centrale européenne lui met à disposition les procès-verbaux de son ancêtre, le Comité des gouverneurs des banques centrales des États membres de la Communauté européenne, de sa fondation en 1964 à 1987, dans diverses langues[[4]](#footnote-5). L’ensemble des sources ici mentionnées sont dites numérisées : elles sont nées sur un support « analogique » (papier, celluloïd, support photographique argentique pour donner quelques exemples) et ont été transférées sur des supports numériques (disquettes, cd-rom, disques durs) puis ouverts à un accès en ligne. Toutefois, un programme de numérisation est une affaire de choix et trouver un document en ligne n’est pas équivalent à aller dans un centre d’archives. Si cette affirmation ressort de l’évidence, encore faut-il se demander pourquoi l’est-elle.

#### Sélection

Le premier élément à prendre en compte est la politique de sélection des documents à numériser par l’institutition ayant lancé le programme de numérisation. Il y a ici une rupture souvent non dite mais néanmoins majeure avec un principe fondamental, le respect des fonds. Ce dernier est un principe d’archivistique européenne (à contre balancer par la notion de *series* dans le monde anglo-saxon) : il « implique de respecter l’intégrité matérielle et intellectuelle de chaque fonds sans tenter de séparer les documents les uns des autres et sans les répartir dans des catégories artificielles en fonction de leur sujet » (Wikipédia). Le respect des fonds permet à l’historien d’obtenir des informations qu’il ne pourrait obtenir dans d’autres conditions. Par exemple, pendant ma thèse, un échange de lettre entre le ministre de l’Économie du Reich, Hjalmar Schacht, et le ministre des Finances Chinois ne se trouvait pas dans le fonds du ministère de l’Économie mais dans le fonds de la Chancellerie du Reich. Nous sommes alors en 1937 : la présence de cet échange de lettres dans les archives de la chancellerie montre que le dossier chinois, au moment d’un basculement de la diplomatie allemande de la Chine vers le Japon, n’est plus du ressort d’un ministre comme Schacht, mais directement d’Hitler et de ses collaborateurs directs et, notamment, du secrétaire de la chancellerie, Lammers.

Les institutions menant des programmes de numérisation ne suivent pas nécessairement le principe du respect des fonds et y contreviennent de deux manières : soit en ne numérisant qu’une partie des fonds (tout en les exposant fonds par fonds), soit en numérisant des sources soigneusement sélectionnées et en les exposant dans le cadre de dossiers thématiques. Tout ou partie du contexte des archives est alors perdu. Deux sites web plus particulièrement illustrent cette situation, cvce.eu et aei.pitt.edu. Ces deux sites web reflètent à la fois deux politiques différentes de sélection des sources primaires mises à disposition des chercheurs et chercheuses et deux modes de non respect du principe de respect des fonds.

Le CVCE regroupe les sources primaires numérisées par grands dossiers thématiques (autour du Portugal et de l’intégration européenne, par exemple[[5]](#footnote-6)) ou chronologique (*Événements historiques*[[6]](#footnote-7)). Les sources présentes dans ces dossiers sont de natures extrêmement variées (photographies, coupures de presse, entretiens de témoins menés par les membres du CVCE, textes d’archives, textes produits par le CVCE, traductions, etc.). Le CVCE, anciennement institution autonome, est ainsi autant producteur (interviews) que reproducteurs (par la numérisation) d’archives. Il n’est pas alors question ici de « fonds » mais plutôt de dossiers documentaires, suffisamment riches pour servir des chercheurs et chercheuses, mais où le contexte auquel l’on a accès grâce au respect des fonds disparaît au profit de métadonnées, ces données sur les données qui indiquent les éléments de provenance et de citation du document. Les critères de sélection des documents sont rarement explicités, si ce n’est dans leur pertinence pour le sujet du dossier. Certains documents peuvent n’être reproduits que partiellement.

*Archive or European integration*[[7]](#footnote-8) repose sur un principe différent de cvce.eu. Assemblage hétéroclite (comme cvce.eu), AEI est né de la volonté de préserver des *papiers* d’institutions et organisations, y compris privées – dans les organisations contributrices, on retrouve Bruegel, think tank privé, par exemple – qui étaient publiés sur leurs sites web sans politique de conservation (2003). En 2004, des documents d’archive commencent à y être publiées. AEI est à la fois une « archive » au sens du web (un dépôt pérenne de documents) et un service fourni aux chercheurs qui peuvent ainsi télécharger des documents utiles à leurs travaux.

Dans tous les cas, la politique de sélection est aussi une politique de non-sélection. La numérisation engendre autant une prolifération d’information qu’elle ne cache les informations non sélectionnées, ce que Lara Putnam a appelé les « ombres » de la numérisation[[8]](#footnote-9). Par exemple, certains document du site du CVCE ne sont que numérisés partiellement, sans nécessairement que l’explication soit donnée clairement - il faut ouvrir le document pour le comprendre. Il peut en outre figurer dans une « collection » en entier et dans une autre par extraits seulement. Ces documents, présents à plusieurs endroits du site sous plusieurs formes, mais potentiellement numérisés et disponibles ailleurs deviennent ubiquitaires, rendant la rigueur de citation des historiens et historiennes d’autant plus importante.

#### Numérisation

Ces bibliothèques numériques définissent également une politique de numérisation, qui peut évoluer au cours du temps. Cette politique de numérisation est, là encore, une question de choix et, à nouveau, ces choix et leurs évolutions ne sont pas toujours documentés. Parmi les critères, figurent par exemple celui de remettre en forme le document ou de le garder dans un fac-simile de sa forme d’origine, de procéder à une reconnaissance optique de caractères ou non, de corriger le résultat de cette reconnaissance ou non, ou d’appliquer des traitements plus pointus : reconnaissance d’entités nommées (l’ordinateur devient alors capable de faire la distinction entre un nom de lieu, un nom de personne, *etc*. afin d’améliorer les possibilités de recherche par exemple) ou encore reconnaissance de thèmes (ou *topic modeling*: sur la base, souvent, de la cooccurence des mots, l’on peut dégager de grands thèmes traversant soit un document, soit tout un corpus). Bien évidemment, la qualité du numériseur utilisé est déterminante pour la qualité de ces étapes de numérisation. Plus ces étapes de traitement des sources numérisées sont poussées, plus les sources primaires, notamment les textes, sont mis en données, c’est-à-dire objet possible de calculs et, partant, de nouveaux modes de lecture et d’analyse des sources (voir ci-dessous).

Il existe en outre de plus en plus une « nouvelle » forme de numérisation, celle de chaque historien ou historienne en centre d’archives apportant, lorsque c’est autorisé, leur appareil photo numérique. Des études récentes[[9]](#footnote-10) ont montré la généralisation de cette pratique, mais également leur faible prise en compte dans nos réflexions méthodologiques. Historiens et historiennes aux États-Unis ont ainsi en moyenne 12 000 photographies d’archives sur leurs disques durs, qui deviennent ainsi des bibliothèques numériques personnelles qu’il faut savoir gérer.

### Nativement numériques

La numérisation touche des archives nées «papier» conservées la plupart du temps en centre d’archives. Les historiens de la construction européenne devront toutefois à terme – dès maintenant pour certains – compter avec des sources primaires nativement numériques. Les sources administratives numériques sont soumises aux mêmes règles de mise à disposition que les sources « analogiques » mais d’autres sources primaires, dont certaines de nature nouvelle, sont déjà disponibles. Ainsi, ai-je démarré un projet de recherche autour des « débats non supervisés en ligne » (voir ci-dessous) reposant sur des sources primaires comme les groupes de nouvelles (*newsgroups*), mais aussi, plus récemment, les pages et groupes facebook, malgré des difficultés de plus en plus grandes pour les collecter, ou des tweets. Ces sources en ligne permettront peut-être d’analyser des éléments nouveaux et jusqu’ici peu étudiés et, surtout, peu étudiables en commençant par les échanges directs entre citoyens sur l’Union européenne par exemple.

Toutefois, ces sources primaires ne sont pas des archives, au sens où elles n’ont pas fait l’objet d’un traitement archivistique (de classement et de tri) ou documentaire. L’historien ou l’historienne deviennent alors leurs propres archivistes, devant acquérir des qualités qui ne sont pas nécessairement intégrées à leur formation (développement de bases de données, développement d’un schéma de métadonnées, pérennisation de la base de données, mise en place de logiciels spécifiques pour le traitement des données, logiciels d’analyse), d’autant plus que ces sources peuvent être massives. Ainsi, le terme « eurozone » a-t-il été utilisé dans plus de 1,9 millions de tweets depuis la création de Twitter en mars 2006.

Dans tous les cas, historiens et historiennes sont confrontés à une abondance de données, en tout cas en histoire contemporaine. Cette situation est relativement nouvelle : si tout archiviste pourra rappeler que son métier a profondément changé depuis la fin de la Seconde Guerre mondiale par l’explosion du nombre de documents archivés, l’arrivée des archives nativement numériques fait passer un pallier quantitatif qui engendre des conséquences méthodologiques liées à la gestion de l’abondance. Cet élément n’épargne en aucun cas l’histoire de la construction européenne et l’histoire européenne tout court.

## Modes de lecture et abondance des données historiques

L’abondance des données produites aujourd’hui, y compris par la numérisation des archives historiques, crée un problème fondamental, lorsque la masse des données devient telle que la lire humainement (seul ou au sein d’une équipe de recherche de taille raisonable) n’est plus possible. Il s’agit alors de demander à l’ordinateur de lire pour nous, ce que l’historien de la littérature européenne Franco Moretti a appelé la lecture distante (*distant reading*). Moretti se demandait comment faire l’histoire de la littérature européenne sans se contenter de faire l’histoire des grands textes des grands auteurs, en réintégrant dans cette histoire non seulement les textes plus mineurs des « grands auteurs » mais également les textes des auteurs considérés comme mineurs. La réponse tient en ce que Moretti a qualifié de « petit pacte avec le diable » :

[…] what we really need is a little pact with the devil: we know how to read texts, now let’s learn how *not* to read them. Distant reading: where distance, let me repeat it, *is a condition of knowledge*: it allows you to focus on units that are much smaller or much larger then the text: devices, themes, tropes – or genres or systems[[10]](#footnote-11).

Dès 2007, dans un livre réunissant une série d’articles publiés dans la *New Left Review*, Moretti proposait un raisonnement de lecture distante en trois étapes : *Graphs* – l’approche quantitative héritée des Annales –, *Maps* – adopter les méthodes des géographes pour réaliser une topographie de la littérature –, and *Trees* – user de l’approche de la théorie de la sélection naturelle pour comprendre l’évolution des genres littéraires[[11]](#footnote-12).

Depuis 2007, la notion de lecture distante ou à distance a largement évolué, a été étendue à d’autres méthodes, à d’autres sujets et a fait l’objet de larges critiques également. Parmi ces critiques, la lecture distante a été décrite comme une manière de tordre la réalité à la théorie, comme une manière d’éviter les critiques en faisant montre de la neutralité des statistiques[[12]](#footnote-13). Quoi qu’il arrive, les techniques informatiques que nous pouvons utiliser doivent toujours être interrogées : la plupart repose sur des hypothèses implicites[[13]](#footnote-14). Néanmoins, la notion de lecture distante retient sa pertinence pour de nombreuses sources difficiles à appréhender pour l’humain seul, notamment par leur masse. Il reste surtout intéressant de faire en sorte, contrairement à l’idée d’origine de Moretti, de ne jamais perdre le contact avec sa source et de rester capable de traiter chaque source individuellement ou collectivement : en d’autres termes, d’articuler différents types de lecture, de la lecture proche, classique de l’historien, à la lecture distante, celle du macroscope[[14]](#footnote-15), plus dépaysante.

Concrètement, la lecture distante revient à transformer le texte en matière statistique, d’autant plus exploitable que la mise en données aura été de qualité, et à en déduire différents éléments : les grands thèmes qui traversent un corpus de texte ou encore les réseaux humains (ou autres) que l’on peut déduire du corpus. Pour prendre un exemple qui ne se réfère pas à l’histoire de l’intégration européenne, la lecture distante de millions de tweets sur le Centenaire de la Grande Guerre nous permet ainsi de comprendre les grands thèmes qui sont l’objet des commentaires et discussions des utilisateurs de Twitter et également de les projeter dans le temps du Centenaire. Ainsi, avons-nous pu démontrer que l’hommage aux Poilus morts pour la France, pour les comptes Twitter francophones, était constant mais de nature différente selon les moments de la séquence commémorative du Centenaire : hommage général le jour des grandes commémorations ; hommage particulier, touchant à des Poilus précis et explicitemnet nommés le reste de l’année. Nous avons également pu observer que pour chaque commémoration, un fonds de contestation était toujours présent, avec des manifestations parfois spectaculaires, par exemple à l’occasion de la commémoration franco-allemande de la bataille de Verdun en mai 2016. Enfin, grâce à cette lecture distante de millions de tweets, nous avons pu observer de nouvelles pratiques commémoratives émerger, et qui ont, par exemple, transformé la base de données des Morts pour la France disponible sur « Mémoire des Hommes » (ministère des Armées) en véritable lieux de mémoire en ligne[[15]](#footnote-16).

## Quelques exemples de lecture distante appliquée à l’histoire de l’intégration européenne

Peut-on donner quelques exemples d’utilisation des méthodologies numériques en histoire de la construction européennes ? Nous allons ici nous arrêter sur trois d’entre eux. Le premier montre les possibilités ouvertes par la numérisation de la presse ancienne en histoire de la construction européenne. Le second, première expérimentation de l’auteur de ce chapitre, montre les limites de cette approche quand il y a un défaut de qualité de numérisation. Le troisième, projet de recherche tout juste démarré de l’auteur, montre comment l’exploitation de nouvelles sources permet de poser de nouvelles questions de recherche.

Les projets de numérisation de la presse ancienne, notamment de la part des grandes bibliothèques nationales, ont abouti à la mise à disposition en ligne d’immenses collections, que l’on peut choisir, en tant qu’historien ou historienne, d’exploiter de diverses manières. Si ces grandes collections, notamment de journaux quotidiens, peuvent poser problème[[16]](#footnote-17), il est possible de poser de nouvelles questions de recherche à ces corpus géants. Ainsi peut-on essayer d’évaluer la circulation de l’information au XIXe siècle par exemple[[17]](#footnote-18). Dans le cas de l’histoire de l’intégration européenne, une tentative a été effectuée pour exploiter la presse néerlandaise afin de trouver de nouveaux chemins de recherche en histoire de l’intégration européenne[[18]](#footnote-19).

Partant du point de départ que les médias ont été un acteur peu considéré bien qu’important de l’intégration européenne, les auteurs comptent exploiter la numérisation de masse de la presse et l’émergence de nouvelles méthodologies numériques – qui permettent une lecture distante. Ils se concentrent sur des personnes (*people-centric approach*) afin de procéder à une analyse des réseaux sociaux[[19]](#footnote-20). Le défi technique n’est pas mince, puisqu’il s’agit de repérer, dans un corpus massif et peu structuré, des noms de personne liés à l’intégration européenne et de trouver des liens entre eux. Reposant sur un corpus principalement néerlandais, les résultats de ces analyses sont parfois peu surprenants - on y parle d’Adenauer, Monnet, Schuman. Les auteurs montrent l’importance des acteurs politiques ouest-européens et états-uniens. Ces derniers, dans cette analyse, restent très présents dans la presse néerlandaise après la création de la CECA alors que la littérature, souvent, se contente d’évoquer l’influence de Washington jusqu’en 1950[[20]](#footnote-21). Les auteurs, qui se reposent aussi sur des travaux menés sur la base de données du CVCE[[21]](#footnote-22), notent ainsi que les méthodologies des humanités numériques peuvent faire évoluer l’historiographie de l’intégration européenne. Toutefois, il ne semble pas que des travaux supplémentaires aient été menés par la suite.

Ce type d’approches rencontre des limites manifeste liées à la qualité de la numérisation. La reconnaissance optique de caractères (OCR) est un exemple souvent cité à juste titre, malgré l’apparition d’éléments nouveaux et prometteurs[[22]](#footnote-23). Comme l’a argumenté Tim Hitchcock, même une reconnaissance de texte très performante pourrait mener à des problèmes méthodologiques non-négligeables[[23]](#footnote-24). Voulant écrire une histoire du comité Werner à l’aide de méthodologie de lecture distante, nous avions pu bénéficier au CVCE des archives originales de la famille Werner[[24]](#footnote-25). Malheureusement, cette recherche n’a pu aboutir à quelque chose de convaincant, en raison de la qualité de numérisation des documents contenus dans ce corpus. Nous ne parlons pas ici de la manière, convaincante, dont ils ont été ordonnés, recontextualisés et mis en ligne, mais seulement de leur mise en données ou, dit autrement, de la possibilité de les transformer en matéériau statistique. Si cette recherche a abouti à une publication, ce fut sous la forme d’un chapitre méthodologique[[25]](#footnote-26).

En troisième exemple, nous prendrons la recherche que nous sommes en train de démarrer au *Luxembourg Centre for Contemporary and Digital History*, en binôme avec notre collègue issue de la linguistique computationnelle, Maria Biryukov[[26]](#footnote-27). Il s’agit ici, cette fois, de se reposer non sur des sources numérisées mais sur des sources nativement numériques, celles des groupes de nouvelles (*newsgroup*) qui, depuis 1980, permettent une communication asynchrone via le courrier électronique sur des groupes thématiques. Les groupes qui ici peuvent nous intéresser ont été la plupart du temps fondés dans la deuxième moitié des années 1990, à l’image de fr.soc.politique, qui comme son nom l’indique, couvre la politique française, ou fr.soc.economie qui regroupent des débats autour de l’économie en français. Au sein de ces groupes, la légitimité d’un interlocuteur se construit par la participation au groupe et non par un statut ou un métier. De plus, ces groupes font partie des modes de communication que l’on peut qualifier de *many-to-many*. En d’autres termes, nous pouvons observer au sein de ces groupes des débats qui ne font pas l’objet d’une médiation par des experts (y compris des universitaires) ou par des *gate keepers* (dont les journalistes) de l’accès à l’information : ces débats en ligne peuvent ainsi être qualifiés de non-supervisés. En 2005, année de la fondation du réseau social *Facebook* et du référendum sur le traité constitutionnel européen (TCE) en France et aux Pays-Bas notamment, nous avions pu, ici comme témoin, constater l’importance des débats autour du TCE sur les *newsgroups*, alors à leur apogée avant qu’ils ne déclinent face aux réseaux sociaux numériques que nous connaissons aujourd’hui[[27]](#footnote-28). Cette recherche n’est pas encore au stade des résultats, malheureusement, mais permettra de tenter de répondre à une question importante, celle des relations, dans une campagne mouvementée qui, en France, a abouti au rejet du traité, entre ce qui était encore une marge médiatique (le web et les forums internet comme les *newsgroups*) et le coeur médiatique de l’époque (la presse, la télévision, la radio) : quels ont été les échanges d’idées entre ces deux parties du système médiatique de l’époque ? Des personnalités - outre Étienne Chouard, par exemple - ou des mouvements ont-ils émergé de ces débats ?

## Conclusion

L’ère dans laquelle nous vivons, que nous qualifions de numérique, et qui est marquée par cette mise en données du monde que nous avons tentée de décrire, a sans nul doute des conséquences sur le travail des historiens et historiennes, y compris celles et ceux s’intéressant à l’historiographie de la construction européenne. La mise en données des sources primaires, coeur de ce chapitre, ouvre de grandes possibilités pour l’historien et, en premier lieu, offre de nouveaux modes de lecture permettant de faire ressortir des éléments qu’il serait difficile de faire ressortir sans cette approche nouvelle.

Toutefois, l’exploitation de ces nouvelles possibilités requiert une attention méthodologique nouvelle et de tous les instants, afin de prendre conscience de ce qu’implique l’usage de ces sources primaires numérisées ou de logiciels d’analyse. La lecture distante n’est pas miraculeuse mais, finalement, très insérée dans le coeur de métier de l’historien, la compréhension, l’interprétation et la critique de sources primaires historiques.

L’usage de sources primaires nées numériques n’échappe naturellement pas à cette nécessité méthodologique. Parfois difficile à collecter – l’histoire des *newsgroups* en est un exemple, dont l’archive a été acquise par Google pour devenir *Google Groups* et qui souffre de nombreux problèmes, y compris d’encodage des caractères diacritiques –, ils nous permettront, par leur nature – un medium de communication direct sans (ou du moins avec de nouveaux types de) *gatekeepers* –, d’observer de nouveaux éléments pour mieux comprendre certains moments de l’intégration européenne depuis les années 1990.

## Bibliographie

Alexandre Lamfalussy, Transcription de l’entretien accordé par Alexandre Lamfalussy le 15 février 1999 à Louvain-la-Neuve, 02.1999.

Ardanuy Mariona Coll, Bos Maarten van den et Sporleder Caroline, « Laboratories of Community: How Digital Humanities Can Further New European Integration History », in: Aiello Luca Maria et McFarland Daniel (éds.), *Social Informatics*, Springer International Publishing, 2014 (Lecture Notes in Computer Science 8852), pp. 284‑293. En ligne: <<https://doi.org/10.1007/978-3-319-15168-7_36>>.

Ascari Maurizio, « The Dangers of Distant Reading: Reassessing Moretti’s Approach to Literary Genres », *Genre* 47 (1), 04.2014, pp. 1‑19. En ligne: <<https://doi.org/10.1215/00166928-2392348>>.

Biryukov Maria, Andersen Eva et Wieneke Lars, « Making Sense of Non-Sense. Tracing Topics in a Historical Corpus on Psychiatry Facing Low OCR Quality », 09.2019.

Brennan Timothy, « The Digital-Humanities Bust: After a Decade of Investment and Hype, What Has the Field Accomplished? Not Much », *Chronicle of Higher Education* 64 (8), 10.2017, pp. 1‑1.

Clavert Frédéric, « Face Au Passé : La Grande Guerre Sur Twitter », *Le Temps des médias. Revue d’histoire* (31), 12.2018, pp. 173‑186.

Clavert Frédéric, « L’apport Du Numérique Aux Sciences Historiques: Exemple d’une Analyse Computationnelle Des Archives Werner », in: Danescu Elena et Munoz Susana (éds.), *Pierre Werner et l’Europe : Pensée, Action, Enseignements / Pierre Werner and Europe: His Approach, Action and Legacy*, Bruxelles, PIE - Peter Lang, 2015.

Cordell Ryan, « Viral Textuality in Nineteenth-Century U.S. Newspaper Exchanges », in: Alfano Veronica et Stauffer Andrew (éds.), *Virtual Victorians: Networks, Connections, Technologies*, 2015 edition, New York, Palgrave Macmillan, 2015.

European Central Bank, Comité des gouverneurs : ordres du jour et procès-verbaux, https://www.ecb.europa.eu/ecb/history/archive/agendas/html/index.fr.html.

European Communities, The Delors II Package, 1992.

Fouetillou Guilhem, « Le web et le traité constitutionnel européen : Écologie d’une localité thématique compétitive », *Réseaux* 147 (1), 2008, p. 229. En ligne: <<https://doi.org/10.3917/res.147.0229>>.

Graham Shawn, Milligan Ian et Weingart Scott, *Exploring Big Historical Data:The Historian’s Macroscope*, ICP, 2015. En ligne: <<https://doi.org/10.1142/p981>>.

Hitchcock Tim, « Academic History Writing and Its Disconnects », *Journal of Digital Humanities* 1 (1), 2011.

Larosière Jacuqes de, Interview de Jacques de Larosière (Paris, 25 septembre 2012) (durée: 00:32:00), CVCE.EU by UNI.LU, 09.2012.

Milligan Ian, « Becoming a Desk(Top) Profession: Digital Photography and the Changing Landscape of Archival Research », in: *AHA 2019*, 2019.

Milligan Ian, « Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997-2010 », *Canadian Historical Review* 94 (4), 12.2013, pp. 540‑569. En ligne: <<https://doi.org/10.3138/chr.694>>.

Moretti Franco, *Distant Reading*, Verso Books, 2013.

Moretti Franco, *Graphs, Maps, Trees: Abstract Models for Literary History*, Verso, 2007.

Mullen Abby, Untangling the Mess: Researchers’ Photo Practices, 11.2016.

Novak J., Micheel I., Melenhorst M. et al., « HistoGraph A Visualization Tool for Collaborative Analysis of Networks from Historical Social Multimedia Collections », in: *2014 18th International Conference on Information Visualisation*, 2014, pp. 241‑250. En ligne: <<https://doi.org/10.1109/IV.2014.47>>.

Paloque-Bergès Camille, *Qu’est-ce qu’un forum internet ? Une généalogie historique au prisme des cultures savantes numériques*, Marseille, OpenEdition Press, 2018 (Encyclopédie numérique).

Putnam Lara, « The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast », *The American Historical Review* 121 (2), 01.2016, pp. 377‑402. En ligne: <<https://doi.org/10.1093/ahr/121.2.377>>.

Sculley D. et Pasanek Bradley M., « Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities », *Literary and Linguistic Computing* 23 (4), 12.2008, pp. 409‑424. En ligne: <<https://doi.org/10.1093/llc/fqn019>>.

1. Par exemple: Larosière Jacuqes de, Interview de Jacques de Larosière (Paris, 25 septembre 2012) (durée: 00:32:00), CVCE.EU by UNI.LU, 09.2012. [↑](#footnote-ref-2)
2. Alexandre Lamfalussy, Transcription de l’entretien accordé par Alexandre Lamfalussy le 15 février 1999 à Louvain-la-Neuve, 02.1999 [↑](#footnote-ref-3)
3. European Communities, The Delors II Package, 1992 [↑](#footnote-ref-4)
4. European Central Bank, Comité des gouverneurs : ordres du jour et procès-verbaux, https://www.ecb.europa.eu/ecb/history/archive/agendas/html/index.fr.html [↑](#footnote-ref-5)
5. « Le Portugal dans les COmmunautés européennes », corpus de recherche, [https://www.cvce.eu/recherche/unit-content/-/unit/a7b75b14-91dc-48dd-9df0-a0d51a9edfef/056ec260-fbef-4679-9520-a824090e119c/Resources#4ded0031-fa09-4b29-84c8-eca3e435d437](https://www.cvce.eu/recherche/unit-content/-/unit/a7b75b14-91dc-48dd-9df0-a0d51a9edfef/056ec260-fbef-4679-9520-a824090e119c/Resources%22%20%5Cl%20%224ded0031-fa09-4b29-84c8-eca3e435d437). [↑](#footnote-ref-6)
6. « Événements historiques », <https://www.cvce.eu/epublications/eisc/historical-events>. [↑](#footnote-ref-7)
7. *Archive of European integration*, <https://aei.pitt.edu/>. [↑](#footnote-ref-8)
8. Putnam Lara, « The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast », *The American Historical Review* 121 (2), 01.2016, pp. 377‑402. En ligne: <<https://doi.org/10.1093/ahr/121.2.377>> [↑](#footnote-ref-9)
9. Voir la communication de Ian Milligan à la conférence de l’*American Historical Association* en 2019: Milligan Ian, « Becoming a Desk(Top) Profession: Digital Photography and the Changing Landscape of Archival Research », in: *AHA 2019*, 2019 ainsi que l’étude qui a été faite au moment des premières étapes de l’élaboration du logiciel tropy (logiciel de gestion des photographies prises en centre d’archives): Mullen Abby, Untangling the Mess: Researchers’ Photo Practices, 11.2016. [↑](#footnote-ref-10)
10. Moretti Franco, *Distant Reading*, Verso Books, 2013, édition kindle, emplacement 793. Italique dans le livre original. [↑](#footnote-ref-11)
11. Moretti Franco, *Graphs, Maps, Trees: Abstract Models for Literary History*, Verso, 2007 [↑](#footnote-ref-12)
12. Ascari Maurizio, « The Dangers of Distant Reading: Reassessing Moretti’s Approach to Literary Genres », *Genre* 47 (1), 04.2014, pp. 1‑19. En ligne: <<https://doi.org/10.1215/00166928-2392348>> et Brennan Timothy, « The Digital-Humanities Bust: After a Decade of Investment and Hype, What Has the Field Accomplished? Not Much », *Chronicle of Higher Education* 64 (8), 10.2017, pp. 1‑1. [↑](#footnote-ref-13)
13. Sculley D. et Pasanek Bradley M., « Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities », *Literary and Linguistic Computing* 23 (4), 12.2008, pp. 409‑424. En ligne: <<https://doi.org/10.1093/llc/fqn019>> [↑](#footnote-ref-14)
14. Graham Shawn, Milligan Ian et Weingart Scott, *Exploring Big Historical Data:The Historian’s Macroscope*, ICP, 2015. En ligne: <<https://doi.org/10.1142/p981>> [↑](#footnote-ref-15)
15. Clavert Frédéric, « Face Au Passé : La Grande Guerre Sur Twitter », *Le Temps des médias. Revue d’histoire* (31), 12.2018, pp. 173‑186 [↑](#footnote-ref-16)
16. Milligan Ian, « Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997-2010 », *Canadian Historical Review* 94 (4), 12.2013, pp. 540‑569. En ligne: <<https://doi.org/10.3138/chr.694>> - Depuis la rédaction de cet article où l’auteur montrait que les historiens canadiens utilisaient préférentiellement les journaux quotidiens historiques numérisés, aux détriments de ceux qui ne l’étaient pas, locaux et/ou francophones, la presse canadienne francophone, notamment, a été en grande partie numérisée. [↑](#footnote-ref-17)
17. C’est par exemple l’enjeu du programme de recherche *Oceanic Exchanges* qui étudie la circulation d’information des deux côtés de l’Atlantique de 1840 à 1914. Voir <https://oceanicexchanges.org/>. Le programme repose notamment sur les travaux de Ryan Cordell, son directeur: Cordell Ryan, « Viral Textuality in Nineteenth-Century U.S. Newspaper Exchanges », in: Alfano Veronica et Stauffer Andrew (éds.), *Virtual Victorians: Networks, Connections, Technologies*, 2015 edition, New York, Palgrave Macmillan, 2015 par exemple. D’autres projets en Humanités numériques se sont penchés sur les grands corpus de presse: Impresso (<https://impresso-project.ch/>) ou NewsEye (<https://www.newseye.eu/>). [↑](#footnote-ref-18)
18. Ardanuy Mariona Coll, Bos Maarten van den et Sporleder Caroline, « Laboratories of Community: How Digital Humanities Can Further New European Integration History », in: Aiello Luca Maria et McFarland Daniel (éds.), *Social Informatics*, Springer International Publishing, 2014 (Lecture Notes in Computer Science 8852), pp. 284‑293. En ligne: <<https://doi.org/10.1007/978-3-319-15168-7_36>> [↑](#footnote-ref-19)
19. On parle ici de réseaux sociaux au sens de la sociologie des réseaux sociaux, non au sens des médias dits sociaux comme *Facebook* ou *Twitter*. [↑](#footnote-ref-20)
20. Ce constat est celui des auteurs de l’article. Pour notre part, il nous semble que l’historiographie francophone, notamment les travaux de Gérard Bossuat, a investiguer plus avant l’influence états-unienne sur la construction européenne après 1950. [↑](#footnote-ref-21)
21. Novak J., Micheel I., Melenhorst M. et al., « HistoGraph A Visualization Tool for Collaborative Analysis of Networks from Historical Social Multimedia Collections », in: *2014 18th International Conference on Information Visualisation*, 2014, pp. 241‑250. En ligne: <<https://doi.org/10.1109/IV.2014.47>> [↑](#footnote-ref-22)
22. Le recours à des procédés issus de l’intelligence artificielle dite « faible » (apprentissage machine et apprentissage profon ou *machine learning* et *deep learning*) est en train de permettre de grands progrès, y compris pour l’écriture manuscrite (voir le projet européen Transkribus: <https://transkribus.eu/Transkribus/>). Voir également: Biryukov Maria, Andersen Eva et Wieneke Lars, « Making Sense of Non-Sense. Tracing Topics in a Historical Corpus on Psychiatry Facing Low OCR Quality », 09.2019. [↑](#footnote-ref-23)
23. Hitchcock Tim, « Academic History Writing and Its Disconnects », *Journal of Digital Humanities* 1 (1), 2011 [↑](#footnote-ref-24)
24. « Une relecture du rapport Werner du 8 octobre 1970 à la lumière des archives familiales Pierre Werner », <https://www.cvce.eu/recherche/unit-content/-/unit/ba6ac883-7a80-470c-9baa-8f95b8372811>. [↑](#footnote-ref-25)
25. Clavert Frédéric, « L’apport Du Numérique Aux Sciences Historiques: Exemple d’une Analyse Computationnelle Des Archives Werner », in: Danescu Elena et Munoz Susana (éds.), *Pierre Werner et l’Europe : Pensée, Action, Enseignements / Pierre Werner and Europe: His Approach, Action and Legacy*, Bruxelles, PIE - Peter Lang, 2015 [↑](#footnote-ref-26)
26. Pour une histoire de Usenet (le réseau sur lequel se reposaient les *newsgroups*) et sur les forums internet de manière générale : Paloque-Bergès Camille, *Qu’est-ce qu’un forum internet ? Une généalogie historique au prisme des cultures savantes numériques*, Marseille, OpenEdition Press, 2018 (Encyclopédie numérique). [↑](#footnote-ref-27)
27. Sur l’importance du web dans la campagne référendaire française de 2005 : Fouetillou Guilhem, « Le web et le traité constitutionnel européen : Écologie d’une localité thématique compétitive », *Réseaux* 147 (1), 2008, p. 229. En ligne: <<https://doi.org/10.3917/res.147.0229>>. [↑](#footnote-ref-28)