



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

Qu'est-ce que l'histoire numérique ?

Frédéric Clavert\*

  
Fondation Pierre du Bois  
pour l'histoire du temps présent

### Introduction: l'histoire numérique, une méthodologie pour l'ère des archives sans fin ?

L'histoire numérique, souvent associée aux humanités numériques<sup>1</sup>, est une notion qui a émergé au début des années 1990 si l'on s'en tient à ce que peut nous en dire Google Books ngrams viewer<sup>2</sup>. Cette première période d'utilisation de l'expression est marquée par des projets comme The Valley of the Shadows<sup>3</sup> - qui présentait, tout d'abord sur cd-rom puis, toujours aujourd'hui, en ligne, la vie de deux communautés, l'une confédérée, l'autre du Nord, avant, pendant et après la guerre de Sécession aux États-Unis - mais aussi la fondation de centres comme le Center for History and New Media<sup>4</sup> (CHNM). Cette première génération d'histoire numérique était très marquée par l'histoire publique (public history) qui étudie les relations du grand public à l'histoire et produit des narrations historiques à destination du et avec le grand public<sup>5</sup>.

L'apparition, somme toute assez tardive, de l'expression digital history ne doit toutefois pas nous faire oublier que les historiens, dans de nombreux pays, ont utilisé l'informatique dès les années 1950<sup>6</sup>. Les générations qui ont suivi, celles de l'histoire quantitative ou, ensuite, de la micro-histoire, de l'histoire culturelle et d'autres courants, ont toutes utilisé cet outil - parfois à la marge de leur discipline - et se sont toutes penchées sur les conséquences méthodologiques et épistémologiques de cet usage.

Ce qui nous semble faire la spécificité de l'histoire numérique telle qu'elle commence à apparaître dans les années 1990 est sa concomitance avec la création d'Internet et, surtout, celle du Web. Ce dernier, à la fois archive née numérique et dispositif permettant de mettre en lien des archives numérisées issues, notamment, des grands programmes de numérisations massives publics<sup>7</sup> ou privés<sup>8</sup>, modifie les relations entre histoire et informatique, tant les sources primaires en ligne abondent. Le lien entre Web et Internet d'une part, histoire numérique, d'autre part, est manifeste dans deux publications de 2007 et 2008. George W. Thomas III, porteur du projet The Valley of the Shadows avec Edwards L. Ayers, définit ainsi, en 2008, ce qu'est l'histoire numérique<sup>9</sup>:

« On one level, digital history is an open arena of scholarly production and communication, encompassing the development of new course materials and scholarly data collections. On another, it is a methodological approach framed by the hypertextual power of these technologies to make, define, query, and annotate associations in the human record of the past. To do digital history, then, is to create a framework, an ontology, through the technology for people to experience, read, and follow an argument about a historical problem<sup>10</sup>. »



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

Cette définition fait écho à un livre publié un an auparavant par Roy Rosenzweig, alors directeur du CHNM, et Dan Cohen, qui lui a succédé cette même année, définition, qui, elle aussi, insiste sur l'hypertextualité<sup>11</sup>: ces liens qui sont devenus si courants dans nos pratiques numériques aujourd'hui sont une caractéristique forte de l'histoire numérique, intrinsèquement liée au web. Cette réflexion sur les liens entre le web et l'histoire numérique nous amène à formuler ce qui sera l'axe fort de ce texte. William J. Turkel, historien canadien, parle, au sujet du web et, de manière plus large de l'ère numérique, d'une « archive infinie ». De notre point de vue, la pratique de l'histoire à l'ère numérique est justement liée à l'évolution de ce qu'est une source primaire aujourd'hui, évolution que l'on ne peut séparer du déluge de données caractéristiques du début du XXIe siècle. Qu'est-ce qu'une source primaire à l'ère numérique? Comment le volume incommensurable de données disponibles modifie-t-il aujourd'hui la relation des historien.ne.s à leurs sources?

Répondre à ces questions n'est pas si aisé. Nous nous pencherons d'abord sur la notion de mise en données des sources primaires, sur les caractéristiques distinguant sources numérisées et sources nées numériques, puis sur la manière dont nous pouvons, aujourd'hui, lire ces sources.

### Mise en données du monde et mise en données de l'histoire

Dans un ouvrage relativement récent, deux auteurs ont tenté de définir ce qu'était la mise en données (datafication) du monde :

To datafy a phenomenon is to put it in a quantified format it can be tabulated and analysed<sup>12</sup>.

La notion de mise en données vise des phénomènes qui peuvent être antérieurs à l'ère numérique et à celle des ordinateurs. Ainsi, les auteurs citent-ils la manière dont la marine militaire états-unienne a rationalisé les voies employées par ses bâtiments en les quantifiant systématiquement puis en les analysant. La mise en données prend toutefois une ampleur toute spéciale avec l'apparition des ordinateurs puis avec leur mise en réseau avec internet et avec l'accessibilité facilitée aux données avec le Web. La mise en données du monde liée à l'informatique et au Web a débouché sur ce que l'on appelle aujourd'hui le big data, c'est-à-dire l'émergence de données massives sur de nombreux aspects de notre vie. Ainsi, les réseaux sociaux numériques comme Facebook ou Twitter peuvent-ils être vus comme une mise en données, par ailleurs très imparfaite et très critiquable, des relations sociales.

La mise en données du monde concerne les sciences historiques, car elle touche aux sources primaires, par la numérisation massive de grands corpus de sources, mais aussi en engendrant des sources nées numériques amenées, dans un futur proche et, pour certains d'entre nous, dès aujourd'hui, à devenir des sources primaires.

Nous devons ainsi faire face à des enjeux méthodologiques et épistémologiques dont dépend la manière dont nous serons capables d'utiliser, dans nos travaux, les sources numérisées et nées numériques. Nous allons distinguer sources numérisées et sources nées numériques dont la nature, selon notre expérience, n'est pas similaire. Les sources numérisées sont les documents produits avant l'ère numérique (ou pendant, mais sur un support papier ou analogique) et qui sont portées sur support numérique. Nous verrons néanmoins que cette transformation implique des choix qui peuvent limiter les modes de lecture et l'information qu'ils portent. Les sources nées numériques sont immédiatement produites dans un format numérique, souvent immatériel, et captées par des dispositifs techniques variés. La notion de captation est une notion-clé, car, comme la numérisation, elle implique des choix, parfois - c'est le cas du fameux Like de Facebook, y



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

compris dans sa version plus évoluée aujourd'hui - issus d'une logique commerciale, sociale ou politique précise et que les historien.ne.s doivent prendre en compte dans leur traitement de ces sources.

### **Mettre en données l'histoire: sources numérisées et sources nées numériques**

Les sources primaires numérisées et disponibles pour les historien.ne.s sont aujourd'hui très nombreuses. En se restreignant à deux exemples helvétiques, on peut citer les documents diplomatiques suisses en ligne (dodis.ch) ou encore les archives du Temps, impressionnante base de données de deux siècles d'archives du Journal de Genève, de la Gazette de Lausanne et du plus récent Nouveau Quotidien (<http://letempsarchives.ch/>), qui fournit des outils, développés à l'École Polytechnique Fédérale de Lausanne (EPFL), extrêmement intéressants pour l'historien.ne.<sup>13</sup>.

Toutefois, ce second exemple montre aussi les limites de la numérisation de sources primaires, limites notamment abordées dans un article<sup>14</sup> de l'historien britannique Tim Hitchcock, qui a été l'un des responsables de la numérisation et de la mise en ligne des archives de l'Old Bailey<sup>15</sup>. La première limite réside dans le processus de numérisation elle-même: numériser un document textuel revient d'abord, du point de vue de l'ordinateur, à en faire non du texte, mais une image. Il faut alors transformer cette image en un document vu comme du texte par l'ordinateur. Il y a plusieurs manières d'opérer cette transformation: soit recourir à des techniques dites de reconnaissance optique de caractères (OCR), soit faire un appel aux «foules» (crowdsourcing) pour, finalement, avoir recours à l'expertise humaine. La seconde grande limite est liée à la notion de « méta-données ». Ces dernières permettent de décrire - un peu comme le faisaient les fiches cartonnées dans les bibliothèques avant l'informatisation des catalogues - le contenu des documents numérisés: ce sont des données sur les données. Elles accélèrent le fonctionnement des outils nous permettant de trouver ces documents, c'est-à-dire les moteurs de recherche. Elles sont toutefois la plupart du temps, sinon très imparfaites, du moins très incomplètes. N'y figurent souvent pas les personnalités, les lieux, les traces de temps présents dans les documents. Si une recherche dite « plein texte » permet de rechercher directement dans ces documents, la rapidité de traitement n'est toutefois pas la même. Là aussi, il existe des stratégies pour contourner cette limite: soit faire appel à un procédé informatique et linguistique appelé la reconnaissance d'entités nommées - c'est-à-dire le fait d'indiquer, dans le texte, que certains mots correspondent à des personnes, des lieux, des concepts, des traces de temps, etc. -, soit, à nouveau, faire appel à la participation humaine<sup>16</sup>.

Dans le cas des archives du Temps, les procédés informatiques et linguistiques ont été favorisés. Mais d'autres projets de grande ampleur ont préféré faire appel à l'intelligence humaine. La Public Library de New York archive les menus des restaurants de la ville. Cet archivage « papier » ne permet pas de répondre à des questions que l'on pourrait juger assez simples: combien coûtait en moyenne une bière à Brooklyn dans les années 1930? La mise en données de ces menus, c'est-à-dire leur numérisation (le passage du papier à une image numérique) puis la transformation de cette image en texte reconnu comme tel par l'ordinateur, permettrait de répondre à cette question. Si numériser les menus n'est pas un problème pour la bibliothèque, la reconnaissance de texte de cette masse documentaire où chaque archive peut avoir des caractères d'impression et une mise en page différents et faire appel à plusieurs langues était jugée impossible. Alors, la bibliothèque a lancé le projet What's on the Menu<sup>17</sup>? qui a pour but une indexation collective des menus new-yorkais archivés et devrait permettre, à terme, de transformer ces archives en données à proprement parler.

Le ministère de la Défense français, qui a publié en ligne le site Mémoire des Hommes<sup>18</sup>, a fait de même pour la base de données des Morts pour la France<sup>19</sup> qui recense 1,3 millions de soldats ou civils déclarés morts pour la France ou du moins pour qui la demande a été faite pendant ou après la Première Guerre mondiale: un module d'indexation permet aux



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

passionnés de la Grande Guerre de transformer ces images d'actes administratifs en texte reconnu comme tel par l'ordinateur, un grand service rendu aux historien.ne.s.

Les sources numérisées ne sont toutefois pas le seul type de sources touchées par la mise en données. Si elles représentent nos sources d'hier adaptées (et transformées) en fonction de notre ère numérique, il ne nous faut pas oublier que nos sources de demain, et, dans certains cas, déjà d'aujourd'hui, pour beaucoup, sont et seront nées numériques.

Depuis 1979 et la création du réseau UseNet qui permettait l'accès aux groupes de nouvelles (newsgroups), système de discussion dérivé du mail<sup>20</sup>, nous disposons d'archives nées numériques dont le volume va en s'accroissant au point de pouvoir parler de Big Data, ces volumes gigantesques de données engendrés aujourd'hui par les grandes plateformes du web.

Si le processus de mise en données d'archives «analogiques» implique des choix qui peuvent leur faire perdre certains de leurs aspects originaux, il en est de même pour les archives nées numériques: les données doivent être collectées ou captées. Que capte-t-on? Que préserve-t-on? Comment structurer les données et, notamment, quelles sont les métadonnées qui définiront les données et simplifieront leur recherche?

Nous nous attarderons sur deux exemples emblématiques: les archives du Web et les réseaux sociaux numériques.

La question des archives du web<sup>21</sup>, c'est-à-dire la préservation et l'archivage des sites web, s'est posée relativement tôt dans l'histoire du web, proposée en 1989 par Tim Berners-Lee: dès 1996, Internet Archive, fondation à but non lucratif, commence à sauvegarder de nombreux sites web. Suivi, notamment depuis les années 2000, par les grandes bibliothèques et archives nationales, l'archivage du web permet la mise à disposition, aujourd'hui, de milliards de pages web pour les chercheurs. Mais ici aussi, les choix opérés par les archivistes et conservateurs des bibliothèques sont déterminants: si, par exemple, la Bibliothèque nationale de France (BNF) a mission, avec l'Institut national de l'audiovisuel pour les sites à vocation audiovisuelle, d'indexer et d'archiver l'ensemble du .fr au titre du dépôt légal et, parfois, de procéder à des archivages plus ciblés à l'occasion de certains événements comme l'attentat de janvier 2015, la Bibliothèque nationale suisse (BNS) préfère se concentrer sur des sites d'abord institutionnels liés à la Confédération ou aux cantons. La BNF archive de manière plus large, mais la profondeur et la fréquence d'archivage seront inférieures à celles des sites, nettement moins nombreux, pris en compte par la BNS, mais archivés de manière plus complète.

Les choix opérés ont ainsi une grande conséquence sur ce qui sera (et est déjà) disponible pour les chercheurs. S'il est incontestable que les centres d'archives ont toujours dû procéder à des choix, si les grands fonds d'archives ont toujours été incomplets, il est néanmoins incertain que nous ayons compris toutes les conséquences de ces choix<sup>22</sup>.

Plus problématique encore est le cas des réseaux sociaux numériques. Apparus au milieu des années 2000, Facebook ou Twitter, par exemple, engendrent tous les jours des masses de données sans commune mesure. Twitter - très connu des chercheurs car ses données sont faciles à collecter - permet l'émission d'un demi-milliard par jour de messages pouvant compter jusqu'à 140 caractères et être accompagnés de liens vers des pages web, d'images, de vidéos, etc.

Surtout, les réseaux sociaux numériques incarnent l'un des aspects les plus marquants de la nature de nombreuses sources nées numériques: le flux. Si le flux - par exemple, le flux constant d'émission de tweets, ce demi-milliard de messages émis quotidiennement - renvoie au sens originel du terme «source», il pose des enjeux techniques non négligeables. Ainsi, la



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

bibliothèque du Congrès états-unien archive-t-elle Twitter en vertu d'un accord avec la firme californienne mais n'est pas en mesure, notamment pour des raisons techniques, d'autoriser des chercheurs à l'exploiter: la bibliothèque du Congrès ne dispose tout simplement pas des technologies de recherche permettant d'exploiter une telle masse de données.

Toutefois, les éléments que nous évoquons ici ne doivent pas empêcher les chercheurs et chercheuses, en histoire et de manière plus générale en sciences humaines et sociales, de mener des travaux sur la base de ces sources. Nous nous permettrons de donner l'exemple de nos propres recherches autour des échos des commémorations organisées pour le Centenaire de la Grande Guerre sur Twitter. Utilisant des dispositifs techniques fournis par Twitter, cette recherche permet d'étudier de très près les relations entre le temps long de la mémoire collective de la Grande Guerre et le temps court, en flux, des réseaux sociaux numériques, sur la base de données dont le volume est important (3,5 millions de tweets).

Le volume des données nées numériques et la disponibilité croissante d'archives numérisées, néanmoins, pose d'autres enjeux que celui d'archivage: si nous continuons l'exemple de nos recherches, comment pouvons-nous lire puis interpréter des masses considérables de données? Comment lire les sources lorsque les historien.ne.s font face à un océan de données?

### Lire les sources primaires à l'ère numérique: vers un pacte avec le diable?

Lire 3,5 millions de tweets, dans le cas de mes recherches, n'est humainement pas possible. Les lire reste pourtant nécessaire pour mener une recherche. Franco Moretti, historien de la littérature européenne, chercheur à Stanford et à l'EPFL, a, depuis 2003<sup>23</sup>, développé l'idée de distant reading ou lecture distante ou distanciée, qu'il définit de la manière suivante:

[...] 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<sup>24</sup>.

Ce pacte avec le diable que nous propose Moretti revient en fait à demander à l'ordinateur de lire à notre place. Cette médiation informatique entre les historiens et historiennes et leurs sources permet de répondre à un besoin de distance envers nos sources pour pouvoir les comprendre dans leur ensemble, besoin exacerbé par la masse des données numérisées ou nées numériques que nous devons affronter.

Dans le cas de notre projet de recherche sur le Centenaire, nous utilisons par exemple un logiciel reposant sur des recherches en statistiques relativement anciennes<sup>25</sup> qui nous permet de dégager les grands thèmes discutés dans notre corpus de tweets, séparé en plusieurs sous-corpus en fonction des langues des tweets et «projetés» dans le temps afin de savoir à quels moments ces grands thèmes ont été discutés sur le réseau social numérique<sup>26</sup>. Avec ce type de méthodes que l'on regroupe sous le terme de « fouille de texte » (textmining) et qui n'est pas le seul à pouvoir entrer dans le cadre de la lecture distante, nous pouvons ainsi tenter de lire nos sources par le biais de l'ordinateur.

Toutefois, le risque d'une lecture distante de nos sources est de perdre contact avec elles. Si nous devons faire un reproche, ou, tout du moins, poser une limite au pacte avec le diable que nous propose Franco Moretti, c'est bien celle-là: ne séparons pas les historiens et historiennes de leurs sources. Les outils et méthodes informatiques que nous utilisons doivent nous permettre de constants allers-retours entre lecture distante et lecture proche - ce mode, qui, finalement, est le mode traditionnel de lecture de nos sources.



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

Pratiquer une lecture distante implique en outre que les sources aient été mises en données de manière rigoureuse, notamment lors de l'étape de la reconnaissance optique de caractères, qui est une étape clé dans le cas des sources numérisées, et lors de la captation des données dans le cas des sources nées numériques. Ces étapes sont nécessaires pour assurer la rigueur d'une recherche.

### Conclusion: promesses et risques de l'histoire numérique

Qu'est-ce que l'histoire numérique? Si le fait d'utiliser des outils et méthodes dites numériques joue un grand rôle pour définir ce qu'est l'histoire dite «numérique», il nous semble également qu'elle peut se définir aussi comme une très vaste discussion méthodologique, en tout cas si nous considérons l'angle « données » suivi dans cet article.

Le point méthodologique le plus saillant touche à la mise en données, c'est-à-dire un processus fait de choix. Ces choix impliquent une perte : l'aspect de l'artefact numérique du document original, par exemple, ne sera plus le même ; un tweet stocké dans une base de données est décontextualisé, soustrait à l'une des interfaces qu'un utilisateur de Twitter peut choisir pour le lire, pour ne prendre que deux exemples. Ce que nous collectons ne peut, la plupart du temps, prendre en compte la dimension hautement personnalisée de ce que les utilisateurs consultent sur le web.

Mais cette perte autorise un gain non négligeable: le traitement de corpus immenses de sources primaires reliées entre elles, dont la masse interdit une lecture attentive, humaine, traditionnelle. Embrasser la masse permet de faire ressortir des grandes tendances que, parfois, nous ne pouvons voir quand nous sommes trop proches de nos sources.

Ce gain ne doit toutefois pas se faire au risque de la quantophrénie ou frénésie du quantitatif. Dans les années 1960 et 1970, la quantophrénie a précipité l'ère de l'histoire quantitative fondée sur les grandes séries statistiques dans une impasse, critiquée avec raison par la microhistoire. Ce risque existe fondamentalement toujours, à l'ère du renouveau du quantitatif que nous vivons par le numérique. Le retour récent, illustré par la publication de l'History Manifesto<sup>27</sup>, d'une version très réinterprétée de la longue durée chère à Fernand Braudel, et les débats parfois tendus qui l'ont suivie notamment dans les Annales<sup>28</sup>, illustrent plutôt bien ce risque.

Le second point caractérisant les grands débats récents de l'histoire numérique touche à une notion évoquée plus haut: si l'ère numérique est l'ère de l'archive sans fin, elle n'est pas celle de l'archive sans ombres ni sans absences. L'historienne Lara Putnam<sup>29</sup> l'a relevé: la disponibilité de grandes masses de sources en ligne ne doit pas faire oublier celles qui ne le sont pas. La mise en données engendre ou amplifie des inégalités en termes d'accès et même d'existence des sources primaires. Le choix des sources à numériser et de celles que nous ne numérisons pas orientent la recherche en sciences historiques. Ian Milligan, prenant le cas des citations de la presse canadienne dans la recherche en histoire du Canada, a ainsi montré que les journaux numérisés sont privilégiés par les historiens et historiennes ayant publié dans la Revue d'histoire du Canada, entraînant une sorte d'éviction des sources non-numérisées par les sources numérisées<sup>30</sup>. Les grands journaux canadiens, de Toronto, bien numérisés, sont plus souvent cités, non seulement en défaveur des journaux francophones mais aussi des nombreux journaux locaux existant au XIXe siècle quelle que soit leur langue.

L'histoire numérique, tant comme méthode que comme méthodologie, tant comme communauté de pratiques informatiques que comme espace de débat, doit devenir plus générique, plus diffusée au sein de la communauté des historiens et historiennes, non au titre du progrès technique, mais parce que la plupart des chercheurs sont devenus des



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

historiens numériques en tant qu'utilisateur de nombreuses sources numérisées, restant toutefois largement en retrait des grands débats méthodologiques que cette utilisation entraîne.

Frédéric Clavert

\* Maître Assistant, Université de Lausanne



## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

<sup>1</sup> Stephen Robertson, directeur actuel du Roy Rosenzweig Center for History and New Media, a toutefois apporté de fortes nuances à cette association: Robertson, Stephen, « The Differences between Digital Humanities and Digital History » in Gold Matthew K et Klein Lauren Frederica (éds), *Debates in the Digital Humanities 2016*, Minnesota, Minnesota University Press, 2016. En ligne:

[www.upress.umn.edu](http://www.upress.umn.edu), <<https://www.upress.umn.edu/book-division/books/debates-in-the-digital-humanities-2016>>.

<sup>2</sup> Google Books Ngrams Viewer (<https://books.google.com/ngrams>) permet de visualiser simplement l'évolution de la fréquence de mots ou groupes de mots dans l'immense mais néanmoins fragmentaire corpus de livres numérisés par Google au travers du monde. En utilisant cet outil sur le corpus anglais, Google ngrams note un usage de l'expression « digital history » dès les années 1960, mais dans un contexte, jusqu'aux années 1990, qui n'est pas lié aux sciences historiques. Le corpus numérisé par Google pour créer Google Books d'une part, les incertitudes sur le nombre de livres qui est la base de Google ngrams d'autre part, ne peuvent que nous amener à prévenir le lecteur que cet outil est à utiliser avec de très grandes précautions.

<sup>3</sup> « Interchange: The Promise of Digital History », *The Journal of American History* 95 (2), 09.2008. En ligne: History Cooperative, <<http://www.historycooperative.org/journals/jah/95.2/interchange.html>>, consulté le 14.09.2008.

<sup>4</sup> Renommé aujourd'hui Roy Rosenzweig Center for History and New Media (RRCHNM) en hommage à son fondateur, décédé en 2007. <http://chnm.gmu.edu/>.

<sup>5</sup> Pour plus de renseignements sur l'histoire publique, se rendre sur le site de la fédération internationale d'histoire publique: <http://ifph.hypotheses.org/>

<sup>6</sup> Un exemple parmi d'autres: François Furet et Adeline Daumard, « Méthodes de l'Histoire sociale: les Archives notariales et la Mécanographie », *Annales ESC* 14 (4), 1959, pp. 676-693.

<sup>7</sup> Voir par exemple Gallica, de la Bibliothèque nationale de France: <http://gallica.bnf.fr/>

<sup>8</sup> Google Books, avec parfois des partenariats avec des institutions publiques: <http://books.google.com/>

<sup>9</sup> Pour aller plus loin sur les modifications de la notion d'histoire numérique à cette période, voir: Noiret Serge, « Y a t-il une Histoire Numérique 2.0 ? », in: Genet Jean-Philippe et Zorzo Andrea, *Les historiens et l'informatique : Un métier à réinventer.*, Rome, École française de Rome, 2011 (Collection de l'École Française de Rome 444), pp. 235-288. En ligne: <<http://cadmus.eui.eu/handle/1814/18074>>, consulté le 21.07.2011.

<sup>10</sup> William G. Thomas III. « Interchange: The Promise of Digital History », *The Journal of American History* 95 (2), 09.2008, p. 454.

<sup>11</sup> Daniel Cohen et Roy Rosenzweig, *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*, Pennsylvanie, University of Pennsylvania Press, 2005. En ligne: <<http://chnm.gmu.edu/digitalhistory/>>, consulté le 05.06.2008.

<sup>12</sup> Mayer-Schönberger Viktor et Cukier Kenneth, *Big data: a revolution that will transform how we live, work, and think*, Boston, Houghton Mifflin Harcourt, 2013, p. 72.

<sup>13</sup> Voir Rochat Yannick, Ehrmann Maud, Buntinx Vincent et al., « Navigating through 200 years of historical newspapers », in *ACM 2016 (proceedings)*. <https://www.semanticscholar.org/paper/Navigating-through-200-years-of-historical-Rochat-Ehrmann/6c70d5ba329fd0f4412a9fe94450a4756718e90a>. On pourra voir un usage simple possible de cette base de données en se reportant à un billet en ligne: Clavert Frédéric, « Guerre et neutralité au travers des archives de deux journaux suisses romands », *Billet, L'histoire contemporaine à l'ère numérique*, <<http://histnum.hypotheses.org/2655>>.

<sup>14</sup> Hitchcock Tim, « Academic History Writing and its Disconnects », *Journal of Digital Humanities* 1 (1), Winter.2011. En ligne: <<http://journalofdigitalhumanities.org/1-1/academic-history-writing-and-its-disconnects-by-tim-hitchcock/>>, consulté le 27.06.2012.

<sup>15</sup> L'Old Bailey était la cours de justice criminelle principale de Londres. <https://www.oldbaileyonline.org/>

<sup>16</sup> Les deux approches ne sont pas exclusives et peuvent d'ailleurs être mélangées.

<sup>17</sup> <http://menus.nypl.org/>

<sup>18</sup> <http://www.memoiredeshommes.sga.defense.gouv.fr/>

<sup>19</sup> <http://www.memoiredeshommes.sga.defense.gouv.fr/fr/article.php?larub=24&titre=morts-pour-la-france-de-la-premiere-guerre-mondiale>

<sup>20</sup> Il existe des archives nées numériques depuis la création des moyens de stockage correspondant. Notre exemple se limite ici à ce qui est en ligne.





## Papiers d'actualité / Current Affairs in Perspective

N°5 | July 2017

<sup>21</sup> Pour une vision approfondie des archives du web: Schafer Valérie et Thierry Benjamin, « L'ogre et la toile. Le rendez-vous de l'histoire et des archives du web », Socio. La nouvelle revue des sciences sociales (4), 25.04.2015. En ligne: [socio.revues.org](http://socio.revues.org), <[http://socio.revues.org/1003?utm\\_source=alert](http://socio.revues.org/1003?utm_source=alert)>, consulté le 22.05.2015.

<sup>22</sup> Dès 2003, Roy Rosenzweig attirait l'attention des historiens et historiennes sur la préservation du web. Une version traduite de cet article est disponible: Rosenzweig Roy, « Pénurie ou abondance ? Préserver le passé à l'ère du numérique », in: OpenEdition Press, OpenEdition Press, 2012, pp. 113-149. En ligne: [press.openedition.org](http://press.openedition.org), <<http://press.openedition.org/265>>, consulté le 01.02.2013. Plus récemment, l'historien canadien Ian Milligan a évoqué un possible digital dark age: Tremonti Anna Maria, « Preserving digital history is imperative to save cultural history », The Current, CBCRadio, 11.05.2015. En ligne: <<http://www.cbc.ca/radio/thecurrent/the-current-for-november-5-2015-1.3305130/preserving-digital-history-is-imperative-to-save-cultural-history-1.3305263>>.

<sup>23</sup> Voir Moretti Franco, *Graphs, Maps, Trees: Abstract Models for Literary History*, Verso, 2007. Ce livre réunit trois articles publiés dans la *New Left Review* en 2003 et 2004.

<sup>24</sup> Moretti Franco, *Distant Reading*, Verso Books, 2013.

<sup>25</sup> Et notamment sur: Reinert Max, « Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte », *Les cahiers de l'analyse des données* 8 (2), 1983, pp. 187-198.

<sup>26</sup> Nous renvoyons à l'un de nos articles, paru récemment: Clavert Frédéric, « Échos du centenaire de la Première Guerre mondiale sur Twitter », *Matériaux pour l'histoire de notre temps* (120), 04.05.2017, pp. 18-25. En ligne: [Cairn.info](http://www.cairn.info), <[http://www.cairn.info/resume.php?ID\\_ARTICLE=MATE\\_121\\_0018](http://www.cairn.info/resume.php?ID_ARTICLE=MATE_121_0018)>.

<sup>27</sup> Guldi Jo et Armitage David, *The History Manifesto*, Cambridge, United Kingdom, Cambridge University Press, 2014.

<sup>28</sup> «La longue durée en débat», *Annales HSS*, 2015/2.

<sup>29</sup> Putnam Lara, « The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast », *The American Historical Review* 121 (2), 01.04.2016, pp. 377-402. En ligne: [ahr.oxfordjournals.org](http://ahr.oxfordjournals.org), DOI: 10.1093/ahr/121.2.377.

<sup>30</sup> Milligan Ian, « Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010 », *Canadian Historical Review* 94 (4), décembre.2013, pp. 540-569. En ligne: [MetaPress](http://www.metapress.com), DOI: 10.3138/chr.694.