

DEBAT - DÉBAT - DEBATE

Autour des Big Data. Enjeux, risques et opportunités

Introduction

Catherine Lanneau

Publié en 2014, l'*History Manifesto* des historiens Jo Guldi et David Armitage a suscité depuis lors nombre de commentaires et de publications. La plupart des articles de blogs, recensions et numéros spéciaux de revues ont adopté un ton très critique à l'égard d'un texte volontairement provocateur qui s'employait, du point de vue anglo-saxon, à dénoncer la supposée disparition des historiens du débat public – ce qui prouverait leur impuissance – et à promouvoir, sans réelle analyse critique, la *digital history* et l'usage des *Big Data* comme le principal remède à la "crise de l'histoire". Grâce au brassage de sources numérisées, il serait possible, selon les auteurs, d'en revenir à une "longue durée" – qu'ils mythifient par ailleurs –, et d'appréhender les grands phénomènes globaux de notre époque, comme le défi environnemental. À cette condition, l'histoire retrouverait sa crédibilité aux yeux des décideurs et son intérêt à ceux du grand public.

Par les nombreuses questions qu'il soulève, par les attendus polémiques et souvent partiaux qu'il impose, l'essai a créé un vaste débat

dans le monde des historiens. Les principales contributions à celui-ci sont évoquées dans le premier texte proposé ci-dessous, à commencer par celles de l'*American Historical Review* et, pour la France, des *Annales*. En Belgique, ce sont surtout des historiens de l'époque médiévale, comme Paul Bertrand (UCL), et des historiens du droit¹ qui se sont saisis du sujet et ont alimenté la discussion. Il nous a semblé essentiel que les contemporanéistes se penchent, eux aussi, sur les enjeux majeurs que ce texte met à l'avant-plan et, plus spécifiquement, sur les opportunités et les risques qu'offrent les "humanités numériques" / "*digital humanities*" dans le champ particulier de l'histoire contemporaine. C'est la raison pour laquelle le *Journal of Belgian History* a proposé à deux équipes engagées dans des projets de recherche impliquant largement la digitalisation de nous offrir leur analyse en la matière.

Le premier texte est signé par Christophe Verbruggen, *Ghent Centre for Digital Humanities*, Ramses Delafontaine, Fien Danniau et Sally Chambers, qui déploient leur expertise spécifique pour le premier dans les rapports entre histoire et droit, pour la deuxième dans les secteurs de l'histoire publique et de la *digital history* et, pour la troisième, comme *Digital Humanities Research Coordinator* à l'UGent.

En réponse à l'*History Manifesto*, les auteurs soulignent l'importance que revêt aussi pour eux l'histoire socialement engagée, même s'ils la conçoivent différemment: l'essentiel

1. Voir par exemple QUENTIN VERREYCKEN, "Quel est le problème de l'History Manifesto?", in *ParenThèses*, publié le 13/04/2015, URL: <http://parenthese.hypotheses.org/713> (consulté le 08/10/2017). Aux Pays-Bas, voir OTTO VERVAART, "Going the long roads: legal history and The History Manifesto", in *Rechtsgeschiedenis Blog. Legal history with a Dutch view*, publié le 21 août 2015, URL: <https://rechtsgeschiedenis.wordpress.com/2015/08/21/going-the-long-roads-legal-history-and-the-history-manifesto/> (consulté le 08/10/2017).

réside, à leurs yeux, dans les opportunités collaboratives nouvelles qu'offre l'intégration de méthodes de recherche digitales dans la recherche historique. Si l'informatisation, l'usage de bases de données et de statistiques ne sont pas des phénomènes récents, la mutation contemporaine réside dans la généralisation de ces pratiques et, plus largement, dans la rapidité et la profondeur du changement induit dans les sciences humaines. Les *Digital Humanities* sont devenues une "transdiscipline" qui conduit chacun à accorder plus d'attention aux questions techniques, qui vont de la manière de collecter les sources jusqu'à celle de partager les résultats acquis.

Selon les auteurs le point fondamental est en réalité celui du passage, dans la recherche en sciences humaines, d'un paradigme individuel à un paradigme collaboratif, qui va bien au-delà de la simple coopération. Cette collaboration, transdisciplinaire et transnationale, passe par des infrastructures de recherche nouvelles qui aiguillent les chercheurs dans la phase de "transition digitale". Les auteurs font alors part de leur propre expérience en matière d'infrastructures de recherche digitales selon deux axes : les défis structurels et l'insertion possible, dans ce modèle, d'un public non académique.

Le premier défi, et non le moindre, est celui de définir le terme "infrastructure de recherche" et de s'interroger sur la possibilité – ou non – d'user, en sciences humaines, d'outils créés par et pour d'autres. Au niveau européen, cinq infrastructures en sciences humaines existent, dont DARIAH (*Digital Research Infrastructure for the Arts and Humanities*), au sein de laquelle Verbruggen est impliqué. Il s'agit à la

fois de soutenir et de développer les fonctions de base en matière de recherche scientifique mais aussi de contribuer à faciliter techniquement le travail collaboratif. Plusieurs exemples sont proposés, dans le domaine de la géolocalisation et des bases de données biographiques, sans faire l'impasse sur les écueils potentiels et les risques d'erreurs d'identification inhérents à la gestion d'une importante masse de données de provenances diverses.

Sur la relation entre *digital* et *public history*, les auteurs évoquent une connexion logique mais soulignent aussi que les plateformes professionnelles de recherche ne seront pas pour autant fréquentées par le grand public. Une réelle *public history* requiert l'acceptation d'une "autorité partagée" entre le chercheur et l'individu *lambda* pour parvenir à une co-création de connaissance, qui est cependant rare : la plupart des projets que les auteurs donnent en exemple constituent davantage une œuvre d'enrichissement des bases de données et des collections par les particuliers qu'une œuvre de recherche collaborative.

En résumé, les auteurs se veulent optimistes sur les apports de la digitalisation aux sciences humaines : la recherche est rendue plus rapide et aisée, les moyens d'en communiquer les résultats se sont étoffés, la collaboration transdisciplinaire progresse, tout comme celle entre professionnels et "amateurs". Pour autant, toute médaille a son revers : à leurs yeux, les infrastructures de recherche digitales ne peuvent être de simples portails d'accès aux collections. Elles doivent amener historiens, archivistes et responsables de musées à mieux collaborer et elles doivent surtout favoriser la réflexion critique sur le contexte de numé-

risation et de production des sources digitalisées. C'est toute une "heuristique digitale" qui reste à construire car les sources numérisées ne peuvent être appréhendées en dehors des documents originaux. Le lien est étroit entre critique historique et *digital literacy*.

Cette question de la *digital literacy* et de la contextualisation des sources est au cœur de la seconde contribution à ce débat, sous la plume de Pieter Lagrou (ULB), Delphine Lauwers (AGR) et Ornella Rovetta (ULB), impliqués dans le projet BRAIN "Jusinbellgium", mené avec les Archives générales du Royaume, la KU Leuven et la Philipps-Universität de Marbourg (*International Research and Documentation Center for War Crimes Trials*). Ce projet consiste à identifier, décrire et digitaliser les sources produites par les juridictions belges lors des procès d'après-guerre, de 1914 à 2014².

Dans leur introduction, les auteurs rappellent qu'à la fin des années 1960, certains historiens, tel Emmanuel Le Roy Ladurie, estimaient que l'histoire quantitative et l'informatique allaient profondément et durablement bouleverser la manière d'écrire l'histoire mais aussi le métier d'historien. L'avenir a montré qu'il en était allé tout autrement. Qu'en sera-t-il du tournant actuel, celui des "humanités numériques"? Comment vont résister les "historiens analogiques"? Soucieux d'offrir d'abord leur définition de la *digital literacy*, les auteurs interrogent ensuite les conséquences de la numérisation sur le rapport aux sources et la méthode des historiens. En termes d'accessibilité et de modalités de recherche textuelle, les avantages mais aussi les risques sont nombreux.

En ce qui concerne les avantages, les auteurs fournissent plusieurs exemples très concrets des bénéfices que chaque chercheur peut tirer de la digitalisation des sources: gain de temps, d'argent, recherche plus aisée dans une somme de documents en accroissement permanent, capacités d'océrisation et d'indexation offrant des possibilités de recherche inédites, etc. De telles avancées, qu'aucun chercheur voire aucun individu ne renierait, sont toutefois étroitement corrélées à une série de risques et d'inquiétudes, dont trois sont plus particulièrement épinglés par les auteurs.

Le premier est celui du choix: comment et sur base de quels critères décide-t-on des sources qui seront ou non numérisées? Indépendamment de l'impact potentiel sur les recherches (ne va-t-on pas se précipiter sur ces sources, au détriment d'autres?), la question posée est celle des valeurs démocratiques: il s'agit de privilégier les sources matériellement en péril du fait de leur consultation fréquente et les sources *a priori* les plus intéressantes pour un large public, au niveau mondial. Dans les deux cas, il revient à l'Etat de fixer les priorités. Le deuxième risque est celui, pointé également par les auteurs du premier texte, de la décontextualisation: il est indispensable de connaître la provenance et l'environnement de production des sources numérisées afin de les appréhender dans toutes leurs dimensions. Il revient donc aux concepteurs d'infrastructures de recherche de viser avant tout le pragmatisme et le réalisme afin de permettre aux historiens du futur, dont nul ne peut connaître les questionnements, d'avoir accès à des collections de sources digitali-

2. Voir le site internet du projet: <http://jusinbell.hypotheses.org/> (consulté le 08/10/2017).

sées “analogues” aux collections originales. Enfin, le troisième risque est bien sûr celui du respect de la vie privée et de l’e-réputation. La numérisation démultiplie les chances qu’un document, notamment judiciaire, soit accessible au plus grand nombre par le biais des bases de données, des recherches en plein texte ou des indexations automatiques. La législation européenne protège les vivants mais les descendants peuvent, eux aussi, être concernés, à charge de prouver alors l’intention diffamatoire.

Dans une deuxième partie de leur texte, P. Lagrou, D. Lauwers et O. Rovetta évoquent plus en détails le projet qui les réunit, Jusinbellgium. Ils le replacent d’abord dans le contexte des divers outils numériques de connaissance de la justice pénale internationale³, visant à faciliter l’accès de tous à des sources souvent peu ou mal connues et difficiles d’accès. Ils soulignent ensuite combien leur projet “place ainsi la digitalisation à l’intersection d’une réflexion sur les modalités techniques, la recherche, les usages et les usagers des archives judiciaires”. Car le cœur du débat réside dans la réflexion sur la “logique de production” des archives, quelles qu’elles soient. Ce n’est pas la base de données mais l’historien qui va s’interroger sur le contexte et le but de production d’un document, ces

informations ayant une particulière acuité en matière d’archives judiciaires. Il importe donc d’établir la “carte d’identité” du document, sa “traçabilité”, en fournissant aussi aux chercheurs une indication sur l’ampleur de chaque fichier.

Pour l’avenir proche, les auteurs épinglent deux défis: le transfert aux Archives générales du Royaume de nombreux fonds d’archives longtemps conservés par leurs producteurs (Affaires étrangères, Victimes de guerre, Justice militaire...), qui ne pourront être sérieusement gérés si l’on opère une rationalisation aveugle ou si l’on perd la mémoire de leur gestion; l’investissement dans des infrastructures de recherche numériques mais des infrastructures durables, réalistes, “dans un esprit conservateur et analogique”.

La question des *digital humanities* se pose donc à tout chercheur du XXI^e siècle mais sans doute n’est-elle ni la panacée universelle, ni la somme de dangers que d’aucuns y voient. Nous espérons que la lecture des deux textes qui suivent suscitera des réactions, des objections, des compléments d’information ou encore des questions qui permettront de poursuivre le débat dans nos colonnes et de renforcer ainsi l’interaction entre la revue et son lectorat.

3. Jusinbellgium s’intègre à la base de données jurisprudentielle de la Cour pénale internationale.