



HAL
open science

L'archive à l'épreuve de la mise en ligne, le chercheur à l'épreuve de l'archive numérisée

Anthony Nelzin-Santos

► To cite this version:

Anthony Nelzin-Santos. L'archive à l'épreuve de la mise en ligne, le chercheur à l'épreuve de l'archive numérisée : L'exemple des recensements canadiens de 1871 à 1916. Dans les dédales du web, Éditions de la Sorbonne, 2019, 979-1035103002. halshs-02164046

HAL Id: halshs-02164046

<https://shs.hal.science/halshs-02164046>

Submitted on 24 Jun 2019

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.

Copyright

.....

L'archive à l'épreuve de la mise en ligne, le chercheur à l'épreuve de l'archive numérisée

L'exemple des recensements canadiens de 1871 à 1916

Anthony Nelzin-Santos

CRHNA - université Paris 1 Panthéon-Sorbonne

Le goût de l'archive passe par ce geste artisan, lent et peu rentable, où l'on recopie les textes, morceaux après morceaux, sans en transformer ni la forme, ni l'orthographe, ni même la ponctuation. Sans trop même y penser. En y pensant continûment. Comme si la main, ce faisant, permettait à l'esprit d'être simultanément complice et étranger au temps et à ces femmes et hommes en train de se dire. [...] Ce geste d'approche s'est imposé au point de ne jamais se distinguer du reste du travail. L'archive recopiée à la main, sur une page blanche, est un morceau de temps apprivoisé ; plus tard, on découpera les thèmes, on formulera des interprétations¹.

« L'art se fait avec les mains² », disait Henri Focillon. « L'histoire aussi », semble dire Arlette Farge. Lorsque la main est remplacée par l'ordinateur, est-il encore possible d'avoir ce « goût de l'archive » ? Lorsque le papier jauni et froissé est remplacé par un PDF, peut-on encore nouer un lien intime avec les sources ? Lorsque les cartons pleins à ras bord sont remplacés par un moteur de recherche, peut-on encore découvrir de nouvelles pistes par surprise ?

Autrement dit : la numérisation n'a-t-elle pas « trop » facilité l'accès aux archives, au point de menacer un pan entier du métier d'historien ? Une question ancienne, à laquelle peu répondent par un « oui » ferme ou un « non » définitif, mais beaucoup par un certain désintérêt ou un désintérêt certain³. Une question cruciale pourtant, notamment pour la démographie historique ou l'histoire de l'immigration, dont les sources favorites ont été très tôt numérisées et mises en ligne.

.....

1. Arlette FARGE, *Le goût de l'archive*, Paris, Seuil, 1989, p. 25.

2. Henri FOCILLON, « Éloge de la main », dans *Vie des formes, suivi de Éloge de la main*, Paris, PUF, 1943, p. 101-128.

3. Nicolas DELALANDE, Julien VINCENT, « Portrait de l'historien-ne en cyborg », *Revue d'histoire moderne et contemporaine*, 58/4bis, 2011, p. 5-29.

C'est le cas des recensements canadiens de 1870 à 1916, qui forment l'ossature de mon étude de l'immigration francophone dans les Prairies canadiennes⁴. Leur mise en ligne, entamée au début des années 2000⁵, ouvre des perspectives de recherche inédites. Au-delà des échantillons et des prosopographies à la portée forcément réduite, on peut aujourd'hui espérer aborder les recensements à une autre échelle grâce à de nouveaux traitements largement informatisés.

À ceci près que, loin de faciliter l'accès aux archives, la numérisation et la mise en ligne sont sources de difficultés supplémentaires, qui repoussent le temps du découpage des thèmes et de la formulation des interprétations. Travaillées par des enjeux mémoriels et économiques, façonnées par l'homme et la machine, elles font barrage entre le chercheur et les documents, dont elles ne sont pas une représentation fidèle.

Mutabilité synchronique, mutabilité diachronique

Les recensements canadiens⁶ ne sont pas des sources homogènes, mais une collection de fragments dont la qualité varie selon les particularités de chaque sous-district⁷ et les compétences et idiosyncrasies de chaque « énumérateur⁸ ». Ils sont d'autant plus difficiles à manier qu'ils échappent littéralement à l'emprise de l'historien : en 1955, le Comité des documents publics a autorisé le Bureau fédéral de la statistique à détruire les souches des recensements de 1871 à 1916, à condition qu'elles soient microfilmées⁹.

.....

4. Anthony NELZIN-SANTOS, *Les Français des Prairies canadiennes (1870-1915). Étude de cas : les Battlefords*, mémoire de master en sciences humaines et sociales (mention histoire), sous la direction d'Annick Foucrier, université Paris 1 Panthéon-Sorbonne, 2011, <https://dumas.ccsd.cnrs.fr/dumas-01330887/document>.
5. Quoique de manière très partielle à l'époque, comme le montre une archive du site Ancestry datant d'août 2000, <http://web.archive.org/web/20000815053016/http://www.ancestry.com/search/rectype/alldblist.asp>, consulté le 17 décembre 2016.
6. Sauf indication contraire, les prochaines utilisations du mot « recensement » feront référence aux recensements canadiens.
7. La plus petite unité de découpage du territoire dans le cadre du recensement de la population, dont l'étendue peut varier du quartier d'une grande ville au bassin d'une rivière. Le recensement de 1871, le premier réalisé à l'échelle nationale, compte ainsi 1 701 sous-districts regroupés en 206 districts.
8. Le terme de l'époque pour « recenseur ». Le recensement de 1871 a mobilisé 2 789 agents.
9. Ou, du moins, que le tableau de dénombrement des « vivants » soit microfilmé. Les tableaux de dénombrement des morts, des propriétés mobilières et immobilières, des établissements industriels, des terres cultivées et de leurs produits, du bétail et des pêcheries, et bien d'autres encore, sont irrémédiablement perdus.

L'archive à l'épreuve de la mise en ligne

Or, quoi qu'en dise Bibliothèque et Archives Canada¹⁰, les microfilms n'en fournissent pas « une copie exacte¹¹ » : certains sont surexposés, d'autres trop sous-exposés, quelques-uns sont flous, beaucoup sont illisibles. À son tour, la transcription disponible sur le site web des archives canadiennes n'est pas « une copie exacte » des microfilms : aux erreurs de transcription bien compréhensibles s'ajoutent des fautes d'orthographe plus frustrantes, et l'absence de nombreuses données de la source originale n'est pas compensée par l'ajout de notes marginales exclusives à la copie.

Doublement dissocié de son support, le contenu de l'archive est doublement altéré, « rendu autre » par les copies successives. Cette archive numérisée n'est pas loin d'être une archive « nativement » numérique, tant elle a été façonnée par les artefacts de la numérisation. Comme les souches papier avant elle, elle subit les outrages du temps et des agents chargés de son traitement, et continue donc à évoluer en raison de sa double mutabilité.

Une mutabilité synchronique, d'abord : à un instant *t*, la présentation des données d'un recensement n'est pas identique à la présentation des données d'un autre recensement. La transcription du recensement de 1881¹² reprend la quasi-totalité des champs du tableau de « dénombrement des vivants » – ne manquent que les champs relatifs à la scolarité des enfants et à la nature des éventuels handicapés, que l'on peut toutefois retrouver dans la copie numérisée du microfilm. Mais la transcription du recensement suivant, celui de 1891¹³, se contente de l'état civil.

Une mutabilité diachronique, ensuite : sur le temps long d'une recherche, la présentation des données d'un recensement est susceptible d'évoluer, comme lors de la refonte du site de Bibliothèque et Archives Canada en 2012. Pire, les données elles-mêmes peuvent changer :

-
10. L'institution chargée « de constituer et de préserver le patrimoine documentaire » et d'être le « dépositaire permanent des publications et des institutions fédérales » canadiennes, selon la loi sur la Bibliothèque et les Archives du Canada de 2004. Comme son nom l'indique, elle cumule les fonctions de bibliothèque nationale et d'archives fédérales, un pôle unique conçu pour faciliter « la concertation des divers milieux intéressés à l'acquisition, à la préservation et à la diffusion du savoir ».
 11. « À propos du Recensement de 1906 », *Bibliothèque et Archives Canada*, <http://www.bac-lac.gc.ca/fra/recensements/1906/Pages/propos-recensement.aspx>, consulté le 12 mars 2016.
 12. « Recensement de 1881 », *ibid.*, <http://www.bac-lac.gc.ca/fra/recensements/1881/Pages/propos-recensement.aspx>, consulté le 12 mars 2016.
 13. « Recensement de 1891 », *ibid.*, <http://www.bac-lac.gc.ca/fra/recensements/1891/Pages/propos-recensement.aspx>, consulté le 12 mars 2016.

- ↔ le moteur de recherche des bases de données des recensements de 1851¹⁴, 1906¹⁵, 1911¹⁶, et 1916¹⁷ a été modifié en 2013 ;
- ↔ certaines entrées des bases de données des recensements de 1871¹⁸, 1881¹⁹, 1891²⁰, et 1901²¹ ont été corrigées la même année ;
- ↔ et plus de 130 000 entrées de la base de données du recensement de 1861 ont été mises à jour en 2014²².

Ces changements ont été annoncés, mais pas détaillés, si bien qu'il est impossible de les reporter sans un long travail de vérification manuelle. Et si d'autres changements ont eu lieu depuis 2014, Bibliothèque et Archives Canada ne les a pas divulgués. Tel qu'il est disponible en ligne, le recensement est donc une copie inexacte d'une copie infidèle à la présentation inconstante et au contenu mouvant... Voilà qui douche les espoirs d'une exploitation « facile ».

.....

14. « Lancement d'une nouvelle version de la base de données du Recensement de 1851 », *Le blogue de Bibliothèque et Archives Canada*, <https://ledecoublogue.com/2013/07/25/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-de-1851/>, consulté le 6 janvier 2017.
15. « Lancement d'une nouvelle version de la base de données du Recensement des provinces du Nord-Ouest (1906) », *ibid.*, <https://ledecoublogue.com/2013/01/08/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-des-provinces-du-nord-ouest-1906/>, consulté le 6 janvier 2017.
16. « Lancement d'une nouvelle version de la base de données du Recensement du Canada (1911) », *ibid.*, <https://ledecoublogue.com/2013/02/26/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-du-canada-1911/>, consulté le 6 janvier 2017.
17. « Lancement d'une nouvelle version de la base de données du Recensement des provinces des prairies (1916) », *ibid.*, <https://ledecoublogue.com/2013/01/29/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-des-provinces-des-prairies-1916/>, consulté le 6 janvier 2017.
18. « Lancement d'une nouvelle version de la base de données du Recensement du Canada (1871) », *ibid.*, <https://ledecoublogue.com/2013/07/04/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-du-canada-pour-lontario-1871/>, consulté le 6 janvier 2017.
19. « Lancement d'une nouvelle version de la base de données du Recensement du Canada (1881) », *ibid.*, <https://ledecoublogue.com/2013/04/30/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-du-canada-1881/>, consulté le 6 janvier 2017.
20. « Lancement d'une nouvelle version de la base de données du Recensement du Canada (1891) », *ibid.*, <https://ledecoublogue.com/2013/06/05/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-du-canada-1891/>, consulté le 6 janvier 2017.
21. « Lancement d'une nouvelle version de la base de données du Recensement du Canada (1901) », *ibid.*, <https://ledecoublogue.com/2013/05/23/lancement-dune-nouvelle-version-de-la-base-de-donnees-du-recensement-du-canada-1901/>, consulté le 6 janvier 2017.
22. « Importante mise à jour de la base de données du Recensement de 1861 au Canada », *ibid.*, <https://ledecoublogue.com/2014/08/12/importante-mise-a-jour-de-la-base-de-donnees-du-recensement-de-1861-au-canada/>, consulté le 6 janvier 2017.

Enjeux religieux, enjeux mémoriels, enjeux commerciaux

Bibliothèque et Archives Canada communique peu sur les recensements : elle leur a consacré moins de trente messages depuis qu'elle s'est inscrite sur Twitter en mai 2011, soit moins d'un demi-pourcent du nombre total de ses tweets, préférant mettre l'accent sur les sources photographiques²³. « Vous voulez des noms ? », demande un des premiers messages sur le sujet ; « trouvez vos ancêtres », enjoint un autre relatif aux recensements des provinces des Prairies. Si un message évoque « les chercheurs » et que trois autres mentionnent une « base de données », onze contiennent le mot-clé « généalogie » et huit parlent d'« index nominatif »... car il s'agit bien de cela.

Les recensements de 1906 et de 1916 offrent une mine d'informations sur le développement agricole des Prairies. Ces données ne sont pas transcrites sur le site de Bibliothèque et Archives Canada, restant « enfermées » dans les copies numérisées des microfilms. Qu'importe ! Il ne s'agit pas de constituer une véritable base de données permettant d'étudier différents aspects du « boom du blé²⁴ », mais de proposer un moteur de recherche permettant de trouver des noms. C'est que les généalogistes ont activement participé à la mise en ligne des recensements, au premier rang desquels les généalogistes mormons.

Les membres de l'Église de Jésus-Christ des saints des derniers jours croient en un « plan de salut » qui s'achève par l'exaltation au royaume céleste de ceux qui ont reçu les sacrements au temple et respecté les doctrines et alliances. La généalogie joue un rôle central dans l'exercice de leur foi²⁵ : chaque individu se replace dans une longue chaîne qui doit remonter jusqu'à Adam, et doit pourvoir au salut de l'ensemble de sa famille, notamment par le baptême des défunts. La Société généalogique d'Utah coordonne cet effort de recensement des ancêtres des mormons : fondée en 1894, elle possède aujourd'hui plus de 2,5 millions de microfilms de sources collectées dans plus de cent pays²⁶, dont le Canada et la France²⁷.

Cette entreprise religieuse est aussi une entreprise économique : Ancestry, le pendant commercial de la Société généalogique d'Utah, a réalisé un chiffre d'affaires

-
23. Voir le compte Twitter @BiblioArchives, <https://twitter.com/BiblioArchives>, consulté le 17 mai 2016.
 24. Kris INWOOD, Thanasis STENGOS, « Discontinuities in Canadian Economic Growth, 1870-1985 », *Explorations in Economic History*, 28/3, 1991, p. 274-286.
 25. Bernadette RIGAL-CELLARD, *La religion des mormons*, Paris, Albin Michel, 2012.
 26. « Volunteers Transcribe 250 Million Historical Records », *The Church of Jesus Christ of Latter-Day Saints*, <http://www.mormonnewsroom.org/article/volunteers-transcribe-250-million-historical-records>, consulté le 6 janvier 2017.
 27. Ministère de la Culture et de la Communication, arrêté du 29 septembre 1987 approuvant l'avenant à l'accord du 28 octobre 1960 entre la direction des Archives de France et la Société généalogique de Salt Lake City, <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000322495>, consulté le 30 avril 2016.

de plus de 683 millions de dollars en 2015²⁸. En vertu de la loi sur la statistique de 1985, le recensement de 1921 est ouvert à la consultation depuis 2013... mais pas sur le site de Bibliothèque et Archives Canada. Numérisé par les mormons, il n'est disponible que sur le site des mormons, qui exigent la création d'un compte utilisateur²⁹, et interdisent tout dépouillement automatisé³⁰.

Le moteur de recherche du site d'Ancestry permet pourtant de dépasser l'index nominatif, grâce à de nombreux champs et quelques opérateurs, qui permettent de construire de véritables requêtes. Celui du site de Bibliothèque et Archives Canada est beaucoup plus fruste : au-delà des critères géographiques, il se limite à l'état civil (nom, prénom et âge). Le champ « mots-clés » agit bien sur d'autres champs, mais sans opérateurs³¹ ni possibilité de préciser son périmètre d'action, il est bien difficile de l'exploiter de manière répétée et consistante.

En s'imposant comme la seule voie d'accès aux sources, ces moteurs agissent comme une grille dont les clés ne sont pas livrées au chercheur. Est-ce que les critères de recherche sont manipulés par des algorithmes de « recherche approximative », comme le code soundex³² ou la distance de Damerau-Levenshtein³³ ? Comment le champ « mots-clés » indistinct fonctionne-t-il ? Dans quel ordre la liste des résultats est-elle assemblée ?

Ce dernier point peut sembler plus évident que les autres, la logique nominative favorisant l'ordre alphabétique. Mais son utilisation est loin d'être systématique, comme celle du « numéro d'item », la référence unique de chaque ligne d'un recensement. Si un score ou un poids est attribué à chaque résultat, alors ce paramètre sort d'une boîte noire, une nouvelle source d'altération de l'archive numérisée.

De l'index généalogique à la base de données historique

Quand bien même il serait documenté, le moteur de recherche continuerait à imposer une manière d'aborder le recensement, prévue pour des usages ponctuels et linéaires plutôt que systématiques et transversaux, sous l'effet d'enjeux mémoriels et d'intérêts

-
28. « Ancestry.com LLC Reports Fourth Quarter and Full Year 2015 Financial Results », *Ancestry*, <http://www.ancestry.com/corporate/newsroom/press-releases/ancestrycom-llc-reports-fourth-quarter-and-full-year-2015-financial-results>, consulté le 6 janvier 2017.
 29. L'accès est gratuit depuis les locaux de Bibliothèque et Archives Canada et dans les bibliothèques abonnées, mais payant partout ailleurs.
 30. « Conditions d'utilisation », *Ancestry*, <http://www.ancestry.ca/cs/legal/termsandconditions>, consulté le 30 avril 2016.
 31. Qui ont disparu avec la refonte du site de 2012.
 32. Voir notamment Donald E. KNUTH, *The Art of Computer Programming*, 3, *Sorting and Searching*, Reading, Addison-Wesley, 1998.
 33. Vladimir LEVENSHTAIN, « Binary Codes Capable of Correcting Deletions, Insertions and Reversals », *Soviet Physics Dokl*, 10, 1966, p. 707-710.

économiques. Sauf à jeter l'éponge, il faut trouver un autre angle d'« attaque » de l'archive numérique, voire de l'objet de recherche.

Lorsque le moteur de recherche empêche d'isoler les seuls Français, il invite à adopter une démarche comparative, selon des modalités qui ne sont pas nouvelles (la mise en ligne ne crée pas nécessairement les conditions d'une rupture épistémologique), mais peuvent être réalisées dans un temps plus ramassé ou à une échelle plus étendue (la mise en ligne crée les conditions d'une rupture pratique). L'impératif révisionniste peut trouver matière dans les contraintes informatiques.

Reste, et ce n'est pas une mince affaire, à surmonter ces contraintes pour enfin parvenir à l'analyse historique. Un outil de *scraping* comme Scrapy est capable de naviguer d'une page web à l'autre selon un itinéraire précis, et d'en extraire les données selon des paramètres précis, autrement dit d'accéder au recensement de manière systématique et transversale³⁴. Dans ce cadre, il est intéressant de noter que la structure même des pages web du site de Bibliothèque et Archives Canada « résiste » au *scraping*, en présentant les données du recensement sous la forme d'un bloc de texte plutôt que sous la forme d'un tableau ou d'une liste.

Le *scraping* fige l'archive numérisée : la mutabilité diachronique s'envole dès lors que la base de données est reconstituée sur l'ordinateur du chercheur. La mutabilité synchronique demeure, mais elle encombre l'analyse historique (en imposant un retour au microfilm pour retrouver des données manquantes dans la transcription) plus que le traitement informatique (la correspondance du champ « nom » de deux recensements ne pose pas de problème majeur).

Bien sûr, il n'est pas exclu que Bibliothèque et Archives Canada modifie ses données dans le futur, comme elle l'a fait par le passé. Faudrait-il alors mettre à jour la base de données ? Si 130 000 entrées sont concernées, il faudra peut-être relancer une session de *scraping*, en fonction du temps et des objectifs de la recherche. Un outil de comparaison comme MySQLDBCompare³⁵ serait alors capable de fournir une liste exhaustive des différences entre l'ancienne et la nouvelle base de données collectées, et résoudrait définitivement le problème de la mutabilité diachronique.

Si seulement quelques dizaines d'entrées sont touchées, on peut sans doute s'épargner cette peine, les problèmes intrinsèques à ce type de source étant nettement plus importants. Les noms français sont souvent écorchés par les énumérateurs anglophones, et la complexité des sociétés amérindiennes et métis n'est pas toujours saisie par le cadre strict du recensement. Aux doutes de l'énumérateur s'ajoutent ceux du transcritteur, qui ne maîtrise pas toujours le jeu complexe d'abréviations (« B.C. » signifie-t-il « Bas-Canada », ou « Colombie-Britannique » ?), et ajoute parfois des erreurs dans les noms et les lieux (« Poranteau » pour « Paranteau », par exemple).

.....
34. Ce qui explique sans doute que l'utilisation de ce genre d'outils soit interdite par les conditions d'utilisation d'Ancestry.

35. « mysqlldbcompare — Compare Two Databases and Identify Differences », *MySQL Utilities 1.4 Manual*, <http://dev.mysql.com/doc/mysql-utilities/1.4/en/mysqlldbcompare.html>, consulté le 29 mai 2016.

Le « dépouillement informatisé » ne trouve donc pas sa résolution dans le *scraping*: il convient de nettoyer les données, de tenter de corriger les erreurs de copie, de restaurer de la cohérence partout où cela est possible, sans jamais « surcorriger » et introduire de nouvelles erreurs. Cette opération est longue – les spécialistes aiment à dire qu'elle prend 80 % du temps de l'analyse des données³⁶ – et pénible, mais un outil comme OpenRefine la facilite énormément.

Il agit essentiellement comme une interface à différents outils, notamment la distance de Damerau-Levenshtein et divers algorithmes phonétiques, qui permettent de corriger les données par passes successives. La distance de Damerau-Levenshtein donne le nombre minimal d'opérations pour transformer une chaîne de caractères en une autre par insertion, suppression, substitution ou transposition. En cherchant les termes « séparés » par une distance égale à 1, on élimine rapidement les inversions de caractères et les coquilles les plus simples.

Ce travail à l'échelle du caractère ne permet pas de repérer les erreurs de transcription, contrairement au travail à l'échelle du phonème réalisé par le soundex, un algorithme phonétique créé il y a bientôt cent ans pour les besoins du recensement américain. Deux mots écrits différemment mais prononcés de manière identique ou semblable sont encodés de la même manière, comme ils ont pu avoir été perçus de la même manière par un énumérateur ne parlant pas la même langue que la personne qu'il interroge ou par un transcripateur peinant à reconstituer une rature.

Le soundex n'est pas infallible, mais on peut employer d'autres algorithmes phonétiques, voire calculer la distance de Damerau-Levenshtein de différents encodages pour comparer des termes plus éloignés. C'est ainsi que l'on peut retrouver une certaine Eulalie qui apparaît ici sous la forme « Edalie » (une faute de transcription) et là sous la forme « Helena » (le témoin d'une intégration par l'anglicisation du prénom). On atteint toutefois rapidement les limites de ces méthodes, qui ne sont d'aucun secours si les transcriptions sont complètement erronées.

Avant d'accepter qu'il s'agisse là d'une limite du recensement, et qu'il faille parfois revenir au microfilm, on peut se tourner vers des méthodes plus « floues ». Si le nom ne permet pas de retrouver un individu entre deux recensements, on peut utiliser une combinaison du pays et de l'année de naissance – à condition de ne jamais oublier que ces informations peuvent être erronées. Il n'y a rien dans la perfection mécanique de l'informatique qui puisse corriger la faillibilité corporelle de l'histoire.

Des difficultés (in)surmontables

Les solutions exposées de manière très sommaire dans cet article se reposent sur des outils librement disponibles et relativement accessibles, et si ma manière de les utiliser est intimement liée à mon objet de recherche, elle est loin d'être inédite et

.....

36. Tamrapani DASU, Theodore JOHNSON, *Exploratory Data Mining and Data Cleaning*, Hoboken, John Wiley & Sons, 2003.

particulièrement innovante. Elle reste pourtant rare : alors même que les paragraphes qui ont précédé passent sous silence les détails de la sélection, de la programmation, et du déploiement des outils évoqués, ils ne manqueront pas d'effrayer certains étudiants et chercheurs à la recherche d'inspiration pour leurs propres solutions.

On tient là un enjeu de taille pour l'historien, qui s'il est aujourd'hui largement familier des outils informatiques, n'est sans doute pas devenu le programmeur dont rêvait Emmanuel Le Roy Ladurie³⁷. Philippe Rygiel écrivait il y a quinze ans des mots qui pourraient avoir été écrits hier :

La page-écran à laquelle nous accédons incorpore des choix faits par les opérateurs chargés de la numérisation et de la mise à disposition du document, choix qui ne sont techniques qu'en apparence puisqu'ils enchâssent des représentations des usages légitimes ou souhaités des documents disponibles, ainsi que des compétences possédées par les utilisateurs et déterminent les usages possibles. [...] nous sommes en permanence confrontés à un écrit-écran dont les caractéristiques ont été fixées par d'autres, qui sont rarement des historiens et prennent rarement en compte les besoins des historiens, d'ailleurs très divers d'une recherche à l'autre. En d'autres termes, il faut, le plus souvent, ou retravailler un matériau qui n'a été produit ni par ni pour les historiens, ou accepter que celui-ci ne réponde qu'à certaines de nos attentes et ne dispense pas toujours de retourner à la source de l'objet examiné³⁸.

La numérisation a peut-être facilité l'accès aux archives, mais elle n'a sans doute pas facilité leur dépouillement, du moins pas d'une manière qui irait au-delà de la simple dématérialisation des pratiques ancestrales. Absent du processus, qui met en jeu des demandes sociales et des intérêts économiques qui le dépassent, l'historien bute sur des dispositifs qui n'ont pas été conçus pour ses besoins. Les outils informatiques permettent de surmonter cet écueil, mais à tant cliquer et coder, il est parfois difficile de ne pas perdre de vue l'analyse historique... et le goût de l'archive numérisée.

.....
37. Emmanuel LE ROY LADURIE, *Le territoire de l'historien*, Paris, Gallimard, 1973, p. 14.

38. Philippe RYGIEL, « Les sources de l'historien à l'heure d'Internet », *Hypothèses*, 1, 2004, p. 341-354.

S. Lamassé, G. Bonnot (dir.), Dans les dédales du web,
Paris, Éditions de la Sorbonne, 2019