

Appel à contributions pour le séminaire de l'école doctorale d'histoire du 27 avril 2024 : Doctorantes et doctorants face à leurs données : collecte, traitement, valorisation

L'idée de cette demi-journée de séminaire « Doctorantes et doctorants face à leurs données : collecte, traitement, valorisation » est née d'une question lancinante : quel est le plus petit dénominateur commun entre tous les doctorantes et les doctorants en histoire à Paris 1 ? D'aucun répondrait sans doute la caféine, l'accumulation déraisonnable de livres plus ou moins poussiéreux ou les regards inquiets de chaque nouvelle connaissance après quinze minutes d'explications (très claires au demeurant) de notre sujet de thèse. Plus prosaïquement, il semble que ce qui réunit doctorants et doctorantes, c'est bien leur rapport aux données.

Les données, tentative de définition historique

Le terme « données » est depuis une vingtaine d'années très connoté avec un traitement informatique et quantitatif, notamment via la construction de bases de données et l'émergence récente du *big data*¹. En effet, pour l'informatique, qu'est-ce qu'une donnée, si ce n'est un assemblage de 1 et de 0, une « description élémentaire, typiquement numérique, d'une réalité »² ? Cette vision est cependant restrictive, comme le fait remarquer Frédéric Clavert, qui note que la « mise en données de l'histoire » n'implique « ni informatique, ni numérique », citant les grandes thèses d'histoire quantitative des années 1960, et que cette dernière peut aussi bien concerner des données quantitatives que qualitatives³. Dans le cadre de ce séminaire, il a été choisi d'adopter une définition très large de « données », comme **des informations obtenues à partir de sources de toutes natures et subissant diverses opérations subséquentes : collecte, exploitation, restitution**. Cette définition est inspirée par les remarques de Jean-Philippe Genet, qui affirme dans un article bien connu que « Le travail de l'historien sur ses sources (collecte, interprétation, réalisation des potentialités) transforme le "réel historique" en une collection de données scientifiquement construites »⁴.

Un thème dans l'air du temps

Denis Woronoff n'écrivait-il pas, dès les premières pages du premier volume *d'Hypothèses*, qu'il « reste nécessaire d'organiser une initiation à des outils de recherche communs à la discipline historique »⁵ ? Nombreux sont depuis les séminaires et journées d'études de l'ED à avoir intégré une réflexion sur les sources⁶, les méthodes propres à certains champs de recherches⁷, ou la construction de bases de données⁸. En effet ces questions sont dans l'air du temps, dans un contexte de généralisation du recours à l'ordinateur en histoire et d'accumulation vertigineuse des données auxquelles nous avons accès. Les années 2010 ont ainsi vu émerger la notion « d'humanités numériques », désignant les « usages réflexifs des technologies numériques dans la recherche et l'enseignement en sciences humaines et sociales »⁹. Cependant la réflexion autour des méthodes

¹ Du Mouza, Cédric, Lamassé, Stéphane et Philippe Rygiel, « De l'histoire numérique à l'histoire données ? », *Les Cahiers de Framespa*, no. 42, 2023, mis en ligne le 04 juillet 2023.

² Abiteboul, Serge. *Sciences des données : de la logique du premier ordre à la Toile : Leçon inaugurale prononcée le jeudi 8 mars 2012*. Paris, Collège de France, 2012, p.10

³ Clavert, Frédéric. « Vers de nouveaux modes de lecture des sources » In : Olivier Le Deuff (dir.) *Le temps des humanités digitales*, Limoges, FYP éditions, 2014.

⁴ Genet, Jean-Philippe. « Histoire, Informatique, Mesure », *Histoire & Mesure* 1, no. 1, 1986, pp.7-18

⁵ Woronoff, Denis. « Préface », *Hypothèses*, vol. 1, no. 1, 1998, pp. 5-6.

⁶ Voir le compte rendu de la journée d'études « L'historien et ses sources » organisée en 2003 par Joseph Morsel : Morsel, Joseph. « Les sources sont-elles « le pain de l'historien » ? », *Hypothèses*, vol. 7, no. 1, 2004, pp. 271-286.

⁷ L'étude des professions : objets et méthodes (2022) - Retrouver le paysage. Intérêts, sources et méthodes (2018) -

Analyser des réseaux : Pourquoi ? Comment ? (2015) La prosopographie, une ressource pour l'histoire sociale (2014)

⁸ « Structurer l'information historique : de la source à la métasource » Journée d'étude de l'école doctorale, 23 mai 2011

⁹ « Éditorial. Donner à lire les humanités numériques francophones (1) », *Humanités numériques*, no.1, 2020, mis en ligne le 01 janvier 2020.

de traitement des données a trouvé une place, notamment à Paris 1, dès les années 1970-1980¹⁰, menant en 1999 à la création du PIREH¹¹ qui propose tous les ans des formations doctorales autour de la lexicométrie, la gestion de bases de données, la cartographie... Plus largement, les chercheuses et chercheurs en sciences sociales se sont emparés du web pour mettre à disposition de nombreuses ressources et tutoriels¹², quand ils et elles n'ont pas créé leurs propres outils de traitement des données¹³.

De plus en plus, la difficulté ne vient pas forcément du manque d'informations, mais bien du trop-plein : quelles méthodes choisir quand elles sont toutes plus ou moins accessibles avec un effort réduit ? Comment résister à la tentation d'accumuler des données, notamment face à l'accroissement permanent des capacités de stockage ? Comment organiser et stocker ces données de natures diverses, afin de minimiser la perte d'information ? Comment les rendre intelligibles, voire exploitables par d'autres ?

La méthode de l'historien.ne à l'épreuve de la reproductibilité technique : axes de réflexion

Les propositions pourront s'inscrire dans trois axes suivant le parcours des données de façon chronologique :

1) Récupérer et organiser des données : La pratique de la photographie d'archive s'est largement répandue dans les dépôts qui le permettent, quand ces derniers ne numérisent pas les archives eux-mêmes. Cela pose la question de la conservation de la matérialité de l'archive dans les données ou métadonnées numériques, des risques (mais aussi des bénéfices) de différer ainsi l'analyse de documents devenus données, du recours à des pratiques non numériques quand l'objet de recherche n'est pas photographiable, mais aussi plus simplement du classement et du stockage des éléments récupérés. En parallèle, les moyens d'acquérir de la donnée se multiplient, aussi bien des solutions techniques permettant l'extraction d'informations (OCR, *web scrapping*, reconnaissance vocale) que la mise en ligne de bases de données bibliographiques ou non (*OpenData*). Les avantages, mais aussi les limites de ces outils pourront sans doute donner lieu à des échanges fructueux.

2) Structurer et exploiter des données : Il s'agit là de l'axe qui sera sans doute le plus familier aux participantes et participants : de l'humble tableau Excel à la base de données relationnelle, du SIG à la lexicométrie en passant par les traitements statistiques les plus divers, les méthodes de traitement des données et les logiciels associés ne manquent pas. Cet axe sera l'occasion de mettre en valeur des pratiques et des outils qui vous paraissent être utiles à d'autres, d'explicitier les étapes de formation de vos corpus et votre raisonnement derrière le choix de tel ou tel outil, telle ou telle pratique.

3) Valoriser et transmettre des données : Une fois les données collectées et structurées, comment les rendre intelligibles au plus grand nombre ? Comment assurer leur conservation et leur utilisation par d'autres chercheurs et chercheuses ? Plus simplement, comment les restituer par écrit en minimisant la perte d'informations ? De l'épineux problème des annexes et de la bibliographie à la production de cartes ou d'un « plan de gestion des données », en passant par la communication sur les blogs et les réseaux sociaux, ce dernier axe s'intéresse au devenir de vos données et à leur valorisation.

¹⁰ Notamment par l'intermédiaire d'Antoine Prost et son cours « d'introduction aux méthodes historiques » qui débouchera en 1996 sur son *Douze leçons sur l'histoire*, bien connu des étudiantes et étudiants de licence.

¹¹ Dumont, Léo, Julien, Octave et Stéphane Lamassé, « Articuler histoire et informatique, enseignement et recherche : le PIREH de l'université Panthéon-Sorbonne », *Humanités numériques*, no.1, 2020, mis en ligne le 01 janvier 2020.

¹² Entre autres <http://www.quantihmc.ens.fr/> lié au séminaire « L'histoire et l'historien face au quantitatif » lancé en 2005 à l'ENS par Claire Lemercier et Claire Zalc, ou le [tutoriel](#) de QGIS proposé par l'UMR Passage

¹³ Le vétéran Tri-deux, créé par Philippe Cibois pour des usages statistiques adaptés aux sciences sociales et aujourd'hui disponible dans une [version web](#), les [outils développés par le Médialab de Sciences Po](#)

Le but de ce séminaire est donc clairement méthodologique, orienté vers la pratique, les aspects concrets du processus de recherche. **Qu'est-ce qui, dans votre usage des données, vous semble généralisable, quels outils, quels processus, quelles méthodes pourraient être utiles à d'autres chercheurs et chercheuses, même très éloignés de votre période ou de votre aire culturelle de prédilection ?** L'objectif de cette demi-journée n'est bien entendu pas de résumer l'entièreté de la diversité des pratiques des doctorantes et doctorants, mais d'en offrir un aperçu permettant aux participantes et participants de jeter un regard réflexif sur leurs propres pratiques, de les modifier, de les enrichir, de permettre peut-être, en déplaçant le regard, en changeant de méthode, de résoudre des problèmes qui paraissaient jusque-là insolubles... **Présentation de logiciels, retour réflexif sur la résolution d'un problème particulier posé par vos données, processus de structuration de vos corpus...** Vos propositions de communication et questions éventuelles sont à envoyer à l'adresse Romane.Carballo@univ-paris1.fr jusqu'au 31 décembre 2023.

La demi-journée de séminaire se tiendra le 27 avril 2024, et laissera une large place aux échanges entre participants et participantes ainsi qu'aux démonstrations concrètes de ce qui se passe dans « l'atelier de l'historien »¹⁴, bien que des retours d'expérience plus théoriques ou épistémologiques soient aussi envisageables.

Romane Carballo - Centre d'histoire sociale des mondes contemporains (UMR 8058).

¹⁴ Référence au blog tenu par Franziska Heimburger & Émilien Ruiz depuis 2009 (<https://boiteaoutils.info/>) faisant suite à une initiative lancée comme doctorant.e à l'EHESS : Heimburger, Franziska, et Émilien Ruiz. « Faire de l'histoire à l'ère numérique : retours d'expériences », *Revue d'histoire moderne & contemporaine*, vol. 58-4, no. 5, 2011, pp. 70-89.