

5. Place des Big Data dans l'histoire et l'archéologie

Bertrand Jouve et Florent Hautefeuille

Les sources

Les sources de l'historien ou de l'archéologue ne sont pas constituées de pétaoctets de données normalisées qui arrivent toutes les secondes, mais de documents souvent lacunaires, dégradés et, pour les périodes anciennes, aux graphies difficiles. Documents écrits (figure 1), tessons de poteries (figure 2) ou pièces de métal ne sont pas produits par l'expérimentation, mais préexistent aux recherches scientifiques. Il est difficile d'évaluer la proportion de ces « objets » qui ont

traversé le temps et qui, lorsqu'ils sont découverts, sont fragiles et difficilement déplaçables. Dans leur ensemble, ils existent en volume considérable, mais ne participent pourtant que très peu d'un traitement massif. L'enjeu du numérique, de la dématérialisation des données, est celui d'une meilleure accessibilité à ces sources par le biais de leurs « copies » numérisées. L'historiographie contemporaine doit aussi compter sur les patrimoines numériques d'origine numérique (*born digital*), pour lesquels nous maîtrisons peu les limites en temps et en espace.

Le travail classique de l'historien

L'historien ou l'archéologue procède souvent à des études monographiques. Pour l'historien, il s'agit de répondre à une problématique en mobilisant un ensemble de sources qui contribuera à définir le périmètre et les enjeux de l'étude. Pour l'archéologue, l'unité de base est le « site archéologique », dont découle une série de méta-documents (rapports de fouilles, résultats d'analyses, photo, relevés...), qui constitue le socle de la réflexion. La mise à disposition des sources ou des corpus structurés est un mouvement qui commence à toucher le domaine de l'histoire, mais qui reste marginal dans celui de l'archéologie française. Les rapports de fouilles archéologiques, par exemple, ne sont pas accessibles en ligne en France, alors qu'ils sont produits en version électronique depuis plus de 20 ans.

En histoire sociale, l'enjeu que représente la capacité à évaluer la représentativité des corpus, ou en démographie historique la forme des données, qui sont plus souvent quantitatives et, donc, plus facilement « normalisables », positionnent ces domaines scientifiques comme les rares champs de l'histoire dans lesquels sont recherchés des passages à l'échelle par la



Fig. 1 – Contrats notariaux médiévaux (Arch. Dép. Lot, 48 J 31).
© F. Hautefeuille. ■

création de grandes bases de données. Mais l'historiographie et l'archéologie classiques restent majoritairement le fait d'études éclatées qui rendent les approches globales difficiles.

Le Big Data et ses enjeux

Une approche globale de ces sources, qui soit à même de restituer la multiplicité et la variété de leurs usages, représente un défi pour l'historien, qui n'échappe pas à la construction d'une science mondialisée. La modification des modes de collecte et d'organisation de l'information, et l'amélioration des outils de traitement permettent d'augmenter significativement le volume et la variété des données interoperables. La capacité à traiter à grande échelle photos, livres ou profils de poteries constitue un enjeu majeur pour la recherche en histoire et archéologie.

L'ère du Big Data doit permettre d'avancer dans la construction d'une histoire globale et connectée : embrasser des aires culturelles jusque-là disjointes, œuvrer sur diverses échelles de réalités passées, dans le temps comme dans l'espace, connecter différents travaux de micro-histoire. Des travaux récents montrent que l'ouverture vers une approche plus globale est dès à présent envisageable. La thèse de Nicolas Perreau, par exemple, qui porte sur plus de 500 000 chartes latines, permet de mesurer, par une sorte de rationalisation des comparaisons, les phénomènes de basculement aux XII – XIII^e siècles des dynamiques économiques d'une Europe du sud vers une Europe du nord. Un autre exemple concerne l'utilisation de la documentation fiscale avec plusieurs centaines de millions de descriptions

de parcelles pour les périodes médiévale et moderne (XIII^e – XVIII^e siècles) à l'échelle européenne. La possibilité de rapprocher et d'analyser ces documents permettrait d'éclairer les dynamiques paysagères et démographiques à ces échelles et sur le temps long. Enfin, un dernier exemple concerne la possibilité d'étudier la dynamique de composition des réseaux sociaux anciens par l'analyse de documents notariés, par exemple. La capacité à reconstruire ces réseaux de milliers d'individus autorise de nouvelles approches sur la relation individu – population. En archéologie, la reconstruction numérique 3D permet des interprétations jusqu'ici impossible.

L'infrastructure européenne *Digital Research Infrastructure for the Arts and Humanities* (DARIAH), portée par la France, a pour objectif de faciliter ces passages à l'échelle dans l'ensemble des domaines des SHS. Se positionner à ces échelles permet aussi d'ouvrir la voie à des interactions innovantes avec des domaines scientifiques qui y sont déjà présents, climatologie ou génétique par exemple.

Les verrous, les outils, les équipements

Tout comme les autres sciences, l'histoire ou l'archéologie ont besoin d'outils innovants adaptés aux nouveaux enjeux et aux nouveaux formats de données, ainsi que d'outils spécifiques pour le partage et le traitement de l'interopérabilité* des



Fig. 2 – Photogrammétrie d'une céramique du XIV^e siècle (Flaunac, Lot). © F. Baleux. ■

sources. On conçoit aisément que la question de la place du numérique dans ces sciences, de son effet sur la transformation des pratiques de recherche – les « Humanités Numériques » (*Digital Humanities*) – est centrale. C'est en grande partie le rôle des deux « Très Grandes Infrastructures de Recherche en Sciences Humaines et Sociales » (HUMANUM et PROGEDO), ancrées en région dans les Maisons des Sciences de l'Homme, que de proposer ce type d'outils. Dans cette perspective, les partenariats avec les grands centres de calculs et de données sont travaillés pour être adaptés aux données des SHS. Les initiatives de *crowdsourcing*, que ce soit pour le référencement des sources et la constitution de bases de données ou pour le dépouillement et l'annotation, se multiplient (cf. V.14).

L'incursion des historiens et archéologues dans les problématiques du Big Data n'en est qu'à ses balbutiements, mais on peut parier que ces nouvelles approches modifieront la question épistémologique de la preuve dans ces disciplines.

Références bibliographiques

- J. A. BARCELO et I. BOGDANOVIC – *Mathematics and Archaeology*, CRC Press, 2015.
- P. MANNING – *Big Data in History*, Palgrave Macmillan, 2013.
- A. SOBOCINSKI – *Des papyrus antiques carbonisés déchiffrés à la lumière des rayons X*, Le Monde, 20 janvier 2015.