Patrick Manning, Big Data in History

Frédéric Clavert

▶ **To cite this version:**

Frédéric Clavert. Patrick Manning, Big Data in History. Lectures, 2014, http://lectures.revues.org/13803. halshs-01101198

HAL Id: halshs-01101198
https://shs.hal.science/halshs-01101198

Submitted on 8 Jan 2015

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.

Patrick Manning, *Big Data in History*

Frédéric Clavert

Docteur en histoire contemporaine, Frédéric Clavert a étudié les sciences politiques et l'histoire contemporaine à Strasbourg et à Leeds. Ses recherches s'orientent aujourd'hui vers les relations entre les banquiers centraux et la construction européenne d'une part, sur les sources de l'historien.ne à l'ère numérique d'autre part. Il a été titulaire d'une bourse du DAAD et a obtenu l'aide à la publication du Prix Pierre Grappin. Après cinq ans comme chercheur (Histoire de l'intégration européenne et Humanités numériques) au Centre virtuel de la connaissance sur l'Europe (Luxembourg), il est désormais ingénieur de recherche pour le LabEx "Ecrire une Histoire Nouvelle de l'Europe".

28/02/2014

Faire une histoire globale de la société humaine, sortir les sciences sociales de l'éparpillement grandissant qu'elles connaissent depuis la Seconde Guerre mondiale par la mise en place d'un dépôt unifiant l'ensemble des données historiques dont nous disposons et disposerons, voici le cœur du propos de cet ouvrage. Ce dépôt, ouvert, distribué, permettra aux chercheurs, aux « décideurs », de trouver les tendances de long terme qui touchent à la société humaine, en priorité sur un certain nombre de facteurs quantitatifs (Population, climat, santé…), puis, plus tard, sur d'autres types de données, y compris plus qualitatives. Ce projet, impliquant un vaste effort collaboratif est appelé *Collaborative for Historical Information and Analysis* (CHIA) est le réel sujet de ce livre.

Les écueils pour la mise en place d'un tel dépôt de données sont grands. Ils sont d'un côté techniques, de l'autre côté plus épistémologiques. Manning, dans cet essai, se concentre beaucoup sur le premier type d'écueil, sans ignorer le second. Une large part de ce livre est en effet consacrée à la description technique de l'archive à construire, aux méthodes de gestion du projet de CHIA et à la comparaison avec des recherches relativement proches.

La description, complexe, du CHIA met en avant une volonté d'agréger progressivement des données, en laissant la trace de toutes les opérations qui sont menées, pour se permettre de revenir en arrière (et corriger certaines données si nécessaire), garder le lien avec les « auteurs » - ceux qui ont élaboré et collecté les données - pour pouvoir mener des analyses à tous les niveaux, du local au global. En outre, le CHIA fera appel à des techniques de simulation des données – notamment pour les régions où les données sont manquantes ou rares. Manning est un historien de l'Afrique et a un grand souci de faire sortir l'Histoire académique de son biais occidental. Les requêtes qui seront effectuées sur le dépôt, à des fins d'analyse et de visualisation (les niveaux les plus « élevés » du CHIA), seront elles-mêmes intégrées à la base de données, afin de les exploiter pour la tester. Pour enrichir cette archive d'histoire globale, l'auteur apporte en outre un soin particulier aux moyens qui seront mis à disposition de tout historien pour soumettre son propre jeu de données.

Ce que, finalement, Manning propose est une forme de mise en pratique de la lecture distanciée chère à Franco Moretti[1] à l'histoire globale. Mais cela ne va pas sans certains écueils épistémologiques et méthodologiques et le soin donné aux aspects techniques du CHIA nous rappelle la difficulté de comparer des données qui sont collectées, pour beaucoup, au niveau national. Mais ce soin, précisément, est aussi une réponse aux problèmes épistémologiques, car Manning montre bien à quel point il est conscient des risques d'une telle aventure. Il rappelle notamment que l'un des travaux fondamentaux de l'équipe du CHIA sera de veiller à ce que l'agrégation des données et leur utilisation se fasse dans les meilleures conditions méthodologiques.

Ce livre n'est pas uniquement à destination des historiens. Il est même tout autant à destination de *decision-makers* et des bailleurs de fonds de la recherche. Patrick Manning cherche en conséquence à convaincre et ce, sur tous les fronts : acteurs décisionnels, enseignants, chercheurs, etc., de l'utilité de son projet. Il rappelle les progrès faits dans les sciences naturelles grâce à la mise en place de grandes bases de données, notamment dans les domaines du climat (de la modélisation vers les données) et du génome (des données vers la modélisation) et réfléchit, à haute voix, sur le meilleur hébergeur pour une base de données sur l'histoire humaine. Il espère, dans ce domaine, que l'UNESCO pourra abriter son projet.

Ce ton qui se veut convaincant, qui n'hésite pas à exploiter une certaine redondance, parfois un peu irritante, rappelle énormément ce qu'est en train de devenir la recherche, y compris en histoire : plus collaborative, plus managériale – ce que regretterons certains - mais aussi de plus grande envergure, rassemblant des équipes internationales. Selon Manning, les sciences humaines et sociales rattrapent une forme de retard : elles sont allées vers le plus particulier quand les sciences « dures » allaient vers le plus général. Pour lui, il est temps, aujourd'hui d'aller vers le global, sans perdre contact avec le particulier, le local. De son point de vue, il y a nécessité d'unifier le champ des sciences sociales dans lesquelles il met l'histoire.

Comme le terme *big data* est présent dans le titre, on peut se demander si ce livre n'entre pas dans le point clé de la définition que donnent boyd et Crawford[2] de cette nouvelle tendance : dans leur article « Critical questions for Big Data » publié en 2012, elles rappellent que le Big Data n'est pas qu'une question de volume de données, mais aussi un changement dans la manière de faire de la recherche, dans la mesure où changer les modes de collecte et d'organisation de l'information aboutit à modifier la manière dont nous questionnons cette information. Le livre de Manning est tout-à-fait exemplaire de ce changement dans le mode de recherche. D'ailleurs, l'auteur rappelle qu'il faudra faire appel à des techniques de fouille de données, qui permettent de faire apparaître des tendances, des schémas (*patterns*) dans les données indépendamment d'un questionnement scientifique sur ces données.

---

[1] Moretti Franco, *Graphs, Maps, Trees. Abstract Models for Literary History*, London, Brooklyn, Verso, 2007.

[2] boyd, danah, et Crawford, Kate, « CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon », *Information, Communication & Society* 15(5), juin 2012, P. 662-679.

Les reproches que l'on peut adresser à cet ouvrage existent bel et bien – Manning est très conscient qu'il faudra tout au long du projet interroger les méthodes d'agrégation et d'exploitation des données intégrées au CHIA –, mais on ne peut qu'être frappés par un certain nombre d'éléments que l'on pourra accueillir avec optimisme. En premier lieu, la démarche globale du CHIA est ouverte, dans de nombreux sens du terme (utilisation de logiciels libres, démarche collaborative inclusive, etc). Ensuite, on ne peut en vouloir à Manning, bien au contraire, de désirer rendre aux sciences historiques (et sociales) leur capacité à proposer un récit global de l'histoire, dépassant les récits nationaux sans toutefois les ignorer, et de garder un soin particulier à la pédagogie de l'histoire. Enfin, on applaudira également la nécessité mise en avant par Manning de remettre les processus historiques de long terme dans le débat public, qu'il soit local, national ou global.