

Big data et traçabilité numérique

Les sciences sociales face à la quantification massive des individus

Sous la direction de Pierre-Michel Menger et Simon Paye

Collège de France, coll. « Conférences »



Les traces numériques de l'activité des individus, des entreprises, des administrations, des réseaux sociaux sont devenues un gisement considérable. Comment ces données sont-elles prélevées, stockées, valorisées et vendues ? Et que penser des algorithmes qui convertissent en outil de contrôle et de persuasion l'information sur les comportements, les actes de travail et les échanges ?

Les *big data* sont-elles à notre service ou font-elles de nous les rouages consentants du capitalisme informationnel et relationnel ?

Les sciences sociales enquêtent sur les enjeux sociaux, éthiques, politiques et économiques de ces transformations. Mais elles sont elles aussi de plus en plus consommatrices de données numériques de masse. Cet ouvrage collectif explore l'expansion de la traçabilité numérique dans ces deux dimensions, marchande et scientifique.

Date de publication : 24 octobre 2017

ISBN : 978-2-7226-0466-7

22 €

218 pages

Collège de France

11, place Marcelin-Berthelot
75231 Paris Cedex 05

Contact : publications@college-de-france.fr

Diffusion-distribution : FMSH-diffusion

18 rue Robert-Schuman
94227 Charenton-le-Pont Cedex
Fax : (0033-1) 53 48 20 95

L'ouvrage est dirigé par **Pierre-Michel Menger**, professeur au Collège de France et titulaire de la chaire de « Sociologie du travail créateur », et par **Simon Paye**, maître de conférences à l'université de Lorraine, sociologue du travail et des groupes professionnels.

Avec les contributions de : Jean-Samuel Beuscart, Dominique Boullier, Franck Cochoy, Éric Dagiral, Jérôme Denis, Samuel Goëta, Bernard E. Harcourt, Pierre-Michel Menger, Sylvain Parasie, Simon Paye, David Pontille, Guillaume Tiffon, Didier Tornay, Jean-Sébastien Vayre.



L'édition électronique de cet ouvrage est disponible sur OpenEdition Books : <https://books.openedition.org/cdf/4987> (texte intégral en accès ouvert).

sociologie du Web / sociologie du travail / *big data* / science des données / quantification de l'individu / science ouverte / GAFA / marché des bases de données

La collection « Conférences du Collège de France »

La vie scientifique et intellectuelle du Collège de France s'étend au-delà de l'enseignement qui y est prodigué. De nombreux colloques internationaux, séminaires de recherche et conférences de professeurs étrangers sont organisés chaque année. Et au sein des chaires et des laboratoires, plusieurs centaines de chercheurs engagent des travaux novateurs. La collection « Conférences » a vocation à refléter cette activité.

Nativement numérique, publiée en accès ouvert *freemium* sur [OpenEdition Books](https://books.openedition.org/cdf/4987), elle paraît également désormais sous forme imprimée.

Sommaire

[Introduction](#)

Pierre-Michel Menger

I. Cheminements des *big data* : technologies, marchés, échanges

[Les *big data* à l'assaut du marché des dispositifs marchands : une mise en perspective historique](#)

Franck Cochoy, Jean-Sébastien Vayre

[Gouverner, échanger, sécuriser : les *big data* et la production du savoir numérique](#)

Bernard Harcourt

[La contribution des internautes à la production de données massives : un travail ?](#)

Guillaume Tiffon

II. *Big data* et configurations sociales en mouvement

[La « science des données » à la conquête des mondes sociaux : ce que le « Big Data » doit aux épistémologies locales](#)

Éric Dagiral, Sylvain Parasie

[Infrastructures de données bibliométriques et marché de l'évaluation scientifique](#)

David Pontille, Didier Tornay

[Les facettes de l'Open Data : émergence, fondements et travail en coulisses](#)

Jérôme Denis, Samuel Goeta

III. Données numériques et outils de recherche en sciences sociales

[Des données du Web pour faire de la sociologie... du Web ?](#)

Jean-Samuel Beuscart

[Pour des sciences sociales de troisième génération \(SS3G\)](#)

Dominique Boullier

[Postface : « Un travail de fourmi » \(Entretien avec un *rater*\)](#)

Simon Paye

Les données peuvent être prélevées sur les individus à leur insu. Mais les données émises par chaque individu constituent aussi un bien ou un service qui pourra lui être vendu pour lui permettre de connaître et de contrôler son environnement, sa santé, sa sécurité, ses déplacements, ses relations, ses transactions, ses interactions de sociabilité : c'est le *quantified self*. Or la qualité de ce service marchand de quantification individualisée dépend de l'exploitation à grande échelle des données recueillies [...]. Un ensemble de techniques et d'algorithmes qui nous renseignent sur notre état de santé, sur les qualités de notre nourriture, sur les variations de notre état physique au travail et sur toute autre dimension de notre comportement augmentent la puissance des services marchands qui nous sont vendus, mais ils peuvent alimenter aussi les bases de données qui sont exploitées par la recherche publique à des fins d'intérêt général. Comment établir une distinction stable et efficace entre ce qui sera extrait de nos données personnelles pour améliorer le bien public fondé sur la connaissance, et ce qui, à travers les multiples canaux de diffusion possibles des données, alimentera des usages commerciaux et servira des intérêts particuliers ?

Pierre-Michel Menger, Introduction, p. 14-15.

Le sénateur Rockefeller a tenu des audiences au sein du comité du Sénat des États-Unis sur le Commerce, la Science et le Transport le mercredi 18 décembre 2013, afin d'examiner cette industrie du courtage en données – et de mettre en lumière certaines pratiques assez controversées. Ces audiences ont révélé, par exemple, qu'un courtier en données à Lake Forest, dans l'Illinois, Medbase200, a proposé de vendre à des compagnies pharmaceutiques une liste de « souffrants de viol » au prix de 79 dollars pour 1000 noms. [...] Medbase200 a retiré la base de données des « souffrants de viol » de son site internet après ces révélations, de même que les « listes de victimes de violence domestique, des patients atteints du sida et des malades de la pression exercée par l'entourage qui étaient offertes à la vente ». Mais le nombre et la variété d'autres listes à vendre sont tout simplement stupéfiants.

Bernard E. Harcourt, « Gouverner, échanger, sécuriser », p. 61-62.

Les *raters* constituent une figure à la fois emblématique et peu connue du travail dans l'industrie de la donnée. On ne sait pas combien ils sont, encore moins qui ils sont. Étudiants, *homeworkers*, femmes au foyer, précaires des cinq continents, ils travaillent indirectement pour Google, Amazon ou Microsoft. Ils s'appellent eux-mêmes *human raters* quand le langage plus officiel des intitulés les baptise *internet evaluators* ou *internet assessors*. On connaît peu les entreprises de sous-traitance qui se chargent de rémunérer leurs prestations et d'organiser une vente de travail de masse, comptabilisé à la minute. Enfin, on ne sait pas toujours ce que font précisément ces *raters*, si ce n'est qu'ils travaillent en ligne sur leur propre outil de travail (un ordinateur ou un *smartphone*) et qu'ils sont payés pour compléter, faciliter ou évaluer le « travail » des algorithmes.

Simon Paye, Postface : « Un travail de fourmi », p. 185-186.