



**HAL**  
open science

## VERS LA DATAFICATION DE LA SOCIÉTÉ ?

Fidelia Ibekwe-Sanjuan

► **To cite this version:**

Fidelia Ibekwe-Sanjuan. VERS LA DATAFICATION DE LA SOCIÉTÉ ?. Vincent Meyer. Transition digitale, handicaps et travail social, LEH Editions, pp.31-49, 2017, 978-2-84874-703-3. hal-01898457

**HAL Id: hal-01898457**

**<https://hal.science/hal-01898457>**

Submitted on 18 Oct 2018

**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.

## VERS LA *DATAFICATION* DE LA SOCIÉTÉ ?

FIDELIA IBEKWE-SANJUAN

Professeur des universités

Aix-Marseille Université

Institut méditerranéen des sciences de l'information et de la communication

[fidelia.ibekwe-sanjuan@univ-amu.fr](mailto:fidelia.ibekwe-sanjuan@univ-amu.fr)

Bien que le néologisme de « *datafication* » ne soit pas encore installé dans les usages, une ébauche d'une page wikipédia<sup>1</sup> témoigne de l'existence du terme. Il fait référence à la tendance actuelle qui consiste à fonder le pilotage des activités de nombreux secteurs sur l'exploitation de grandes masses de données générées par nos interactions avec des dispositifs numériques. La *datafication*<sup>2</sup> renvoie également à une croyance répandue que l'exploration de grandes masses de données a la capacité d'accélérer le rythme des découvertes et des innovations dans les secteurs concernés. Avant d'aborder les enjeux sociétaux de la *datafication*, il nous faut clarifier la notion de « donnée » (*datum*) et de « données massives » (*big data*).

### Qu'est-ce qu'une donnée ?

Le sens courant du mot « donnée » laisse entendre qu'il s'agit de quelque chose de neutre ou d'objectif qui serait prêt à l'emploi. Son étymologie latine “*dare*” est également trompeuse car elle signifie l'acte de donner (“*to give*”). Bien qu'il existe des données qui résident dans des éléments naturels (celles dégagées par le temps qu'il fait, par la couleur du ciel ou des feuilles sur les arbres), elles ont besoin d'être interprétées, souvent collectivement, pour construire des significations sur lesquelles nous pouvons être d'accord. Le type de données qui nous préoccupe ici nécessite un effort conscient pour les produire et les mettre en forme. Elles sont donc construites socialement, techniquement, économiquement et situées spatialement et culturellement (comme par exemple, les données sur la consommation d'eau ou d'électricité d'une commune de France). En fait, elles devraient être appelé des « obtenues »<sup>3</sup> ou des « *capta* » (du latin “*capere*”) qui signifie prendre. Comme l'observe Jensen (1950) cité dans Rob Kitchin (2014b : 2) : « C'est un hasard malheureux de l'histoire que *datum* plutôt que *captum* soit venu à symboliser le phénomène basique en science. Car la science traite non pas de “ce qui a été donné par la nature au scientifique” mais de “ce qui a été capté ou sélectionné de la nature par le scientifique ” en accord avec le but de celui-ci ».

---

<sup>1</sup> L'ébauche d'article Wikipedia se trouve à <https://en.wikipedia.org/wiki/Datafication>

<sup>2</sup> Le terme *datafication* apparaît, par exemple, dans un passage de l'ouvrage de Rob Kitchin (2014b : 181) ‘*The Data Revolution. Big Data, Open Data, Data infrastructures & their consequences*’ que nous avons traduit en français: « En effet, (Mattern, 2013) a suggéré que l'urbanisme de *big data* souffre de la “*datafication*”, à savoir cette présomption que tous les flux significatifs et activités peuvent être captés et mesurés ».

<sup>3</sup> Terme suggéré par Bruno Latour lors d'une allocution à une journée d'étude à l'École normale supérieure de Lyon en 2014.

Si on admet ce préalable que la plupart des données sont en fait des « *capta* », se pose de suite la question de quels objets ou phénomènes constituent une donnée. Il y a un consensus chez plusieurs auteurs pour dire qu'une donnée est un signe, un symbole ou un objet qui sert d'appui à l'argumentation. Par exemple, la radio d'un poumon contient des données pour celui qui sait la lire. Une photo de famille des années 1930 est potentiellement une donnée si elle sert de support pour étudier le style d'habillement ou les groupements familiaux de cette époque (Borgman, 2015). Michael Buckland (1991) résumé ainsi cette idée en écrivant que « la donnée est une preuve présumée » ("*Data are alleged evidence*"). Se pose alors la question de savoir ce qui distingue une 'donnée' d'un 'fait' ou d'une 'preuve'. C'est à nouveau vers Rob Kitchin (2014b : 3) que nous nous tournons, citant Daniel Rosenberg (2013), il écrit : « Les faits sont ontologiques, la preuve est épistémologique et les données sont rhétoriques. Une donnée peut également être un fait, de même qu'un fait peut constituer une preuve. L'existence d'une donnée a toujours été indépendante de toute considération de vérité ontologique. Lorsqu'un fait a été démenti, il cesse d'être un fait. En revanche, une fausse donnée reste une donnée malgré tout. Dans des termes rhétoriques, les données sont ce qui existent avant tout argumentation ou interprétation qui les transforment en faits, preuves ou information ».

Cette distinction est en accord avec le caractère indiciel des données : ce sont les matières premières (au sens large du terme) à partir desquelles des faits, des preuves ou des informations peuvent être obtenus. À l'instar de l'information et de la connaissance, une donnée est situationnelle et rhétorique, à savoir que c'est le discours ou l'argumentation porté sur elle qui lui confère tout son sens. Cependant, les données étant des matières premières à partir desquelles nous pouvons déduire de l'information et des connaissances, ceux qui ont accès à des données de grande qualité détiennent un avantage compétitif par rapport à ceux qui n'en ont pas parce qu'ils seront susceptible de faire des découvertes scientifiques ou de générer des valeurs économiques à partir de ces données. Cela nous amène naturellement à considérer le concept très médiatisé de « données massives » (*big data*).

## Que sont les données massives ?

La paternité du terme « *big data* » est attribuée au statisticien Francis Diebold (2000<sup>4</sup>) mais celui-ci en a identifié des occurrences antérieures dans des publicités de la société Silicon Graphics International (SGI). Une présentation de 1998 de l'ancien directeur scientifique de SGI, John Mashey était intitulée « *Big Data and the Next Wave of InfraStress*<sup>5</sup> ». Une autre occurrence du terme était apparue dans un article de Weiss and Indurkha publié en 1998. Sur les plans ontologique et technologique, les données massives désignent des volumes de données dont le traitement nécessite le passage à l'échelle des infrastructures informatiques (architecture hadoop, clusters de machines distribuées) et le recours à une nouvelle génération d'algorithmes d'apprentissage tel le *deep learning*. Dans leur forme, il s'agit de données hétérogènes, pouvant être constituées de textes, de sons et d'images qui sont en constant mouvement, donc volatiles, ce qui explique leur

---

<sup>4</sup> Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting," presented at the Eighth World Congress of the Econometric Society in Seattle in August 2000, and subsequently published as F. Diebold (2003).

<sup>5</sup> The company's overview affirms its chief scientist's claim to paternity of the term « In the late 90s, SGI's Chief Scientist at the time, John R. Mashey coined the term "Big Data." [https://www.sgi.com/company\\_info/overview.html](https://www.sgi.com/company_info/overview.html)

caractérisation par les 3V (Volume, Variété, Vitesse) par Douglas Laney (2001). Cette volatilité soulève d'autres questions, celles de la confiance que l'on peut accorder aux résultats dérivés à partir d'études sur les données massives, d'où le fait que deux autres 'V' (Vérité, Validité) sont venus s'ajouter aux 3V pour donner la caractérisation actuelle du *big data* par les 4 ou 5V.

Alors que le phénomène est désigné par un de ses attributs ontologiques, à savoir le fait d'être massif (*big*), il est paradoxalement rare de trouver un consensus dans la littérature sur la taille minimale à partir de laquelle les données peuvent être qualifiées de « massives ». Au vu de la dématérialisation croissante de l'ensemble de nos activités, il est probable qu'à moyen terme, le concept de « *big data* » finisse par perdre de sa pertinence tant la nature des données numériques sera d'être massives. Pour plusieurs auteurs, le *big data* est moins une question de volume qu'elle n'est de données numériques interconnectées qui posent des défis informatiques en matière de capture, de stockage, d'exploration et de description. Danah Boyd et Kate Crawford (2012 : 664) ont proposé d'appréhender le phénomène du *big data* selon ces trois dimensions :

- la dimension technologique car les données massives nécessitent la montée en charge de la puissance de calcul et de prédiction algorithmique pour pouvoir collecter et gérer de grandes collections de données ;
- la dimension analytique car elles supposent l'identification des motifs dans de grandes collections de données par les algorithmes, ce qui permet de fonder des revendications économique, sociale, technique et juridique ;
- la dimension mythologique car les deux dimensions précédentes suscitent la croyance que les données massives vont nous permettre d'accéder à un niveau plus élevé d'intelligence et de connaissance en nous livrant des éclairages jusqu'alors inatteignables, et qui sont drapés d'une aura de vérité, d'objectivité et d'exactitude.

Ayant rappelé les origines du terme de « *big data* » et tenté de clarifier ce qu'il recouvre, nous pensons utile de rappeler les rapports privilégiés que les États ont toujours eu avec les données. Cela permet d'éclairer la tendance actuelle pour la « datafication » comme le résultant d'une tension historique entre « enclosure » et « disclosure » des données, perçue d'un côté comme un moyen de légitimation de leur pouvoir par les gouvernants, et de l'autre comme un moyen de subjugation ou d'émancipation par les citoyens-gouvernés.

## **Les États et les données : entre « *enclosure* » et « *disclosure* »**

L'adoption actuelle, par de nombreux gouvernements, d'une politique en faveur de l'ouverture des données publiques (*open data*) n'est pas la seule conséquence d'un idéal humaniste visant à garantir la transparence des actions des administrations publiques et des États vis-à-vis de leurs gouvernés. Les premières traces de données recueillies sur des activités publiques sociales et économiques (des recensements de vente de bétail) remontent au XII<sup>e</sup> siècle avant J.C. Yohan Barres (2015) rappelle comment les différentes formes de l'Etat français se sont servies des données comme une arme pour asseoir leur pouvoir, oscillant tantôt entre leur rétention (*enclosure*) et leur divulgation contrôlée (*disclosure*). Les données notamment chiffrées ont toujours été perçues « comme un instrument de l'action publique, une technique de gouvernement<sup>6</sup>, comme support d'une rationalisation de

---

<sup>6</sup> Voir P. Lascoumes et L. Simard, « L'action publique au prisme de ses instruments » Introduction, *Revue française de science politique*, 2011/1 Vol. 61, p. 5-22 ou encore P. Lascoumes et P. Le Galès, « Introduction :

l'exercice du pouvoir ou dans les termes de Michel Foucault comme « *technologie de gouvernementalité* »<sup>7</sup>. L'État chrétien et l'État monarchiste privilégiaient le secret et négociaient la divulgation des données.

La Révolution française a marqué un tournant dans le rapport « État – données », où pour asseoir une nouvelle forme de légitimité, l'État a été contraint de céder aux revendications populaires en mettant en place un « bureau de statistique de la République ». Cela a permis d'aller vers une forme d'État moderne qui émergea à la fin du XVIII<sup>e</sup> siècle pour lequel l'élaboration et la mise en chiffres des réalités sociales sont devenues un moyen « de rendre la société déchiffrable à elle-même. [...] Tous les secteurs de la société sont alors soumis à ce souci de quantifier les comportements de manière à justifier l'action des institutions et à faire exister l'espace public. » (Barres, 2015 : 3).

Apparu au XVIII<sup>e</sup> siècle, le terme de « statistique » serait dérivé des deux termes allemands « *staatskunde* » (études civiques ou sciences politiques) ou « *statistik* » (étude comparative des États). La statistique visait donc à rendre concret pour les gouvernés, la notion d'État, par l'arrangement intelligible des chiffres sur toutes sortes de phénomènes (naissances, taxes, richesses, mortalité, commerce, éducation, salaires...). Cela permettait aux gouvernés d'accéder aux connaissances qui auraient été autrement inaccessibles à leurs expériences individuelles. La statistique fabrique donc des connaissances sans corps, sous forme de données supposées froides et objectives. Cependant, les chiffres et les données demeuraient l'outil du pouvoir notamment avec l'Empire et la Restauration car l'État moderne était plus dans un souci de mise en scène des données publiques dont il conservait jalousement la propriété. Les quelques accès étaient accordés à du personnel et experts dûment patentés, à des fins d'élaboration de statistiques nationales et de recherche.

Le XXI<sup>e</sup> siècle et les politiques incitatives actuelles des États pour l'ouverture des données publiques (*open data*) marquerait selon (Barres, 2015 : 2) un « nouveau moment en rupture avec les conceptions antérieures » de l'État, celui-ci ne se contentant plus de la démonstration et de la mise en scène des données mais incitant les gouvernés à se saisir des données publiques pour créer de nouveaux services et pour innover. Ainsi, après de longs siècles d'*enclosure* et de contrôle exclusif des données publiques vues comme instruments de légitimation de leurs pouvoirs, les États encouragent-ils aujourd'hui la « *disclosure* » des données publiques en tentant de s'arrimer aux mouvements citoyens de l'*open source* et de l'*open access*. Comment expliquer cette *volte face* des États vis-à-vis du contrôle des données publiques ? En réalité, les politiques de l'*open data* apparaissent comme la tentative des États de s'adapter aux bouleversements que le numérique a entraîné dans quasiment tous les secteurs d'activité, entraînant par la même occasion, l'érosion de certaines prérogatives étatiques. Ainsi, « le mouvement d'ouverture des données publiques est le symptôme, sinon le symbole d'une adaptation de l'État aux transformations récentes du capitalisme » (Barres, 2015 : 4).

Porté et défendu par la société civile (chercheurs et associations), le paradigme de l'ouverture des données a gagné de nombreux secteurs avec l'émergence tour à tour de l'*open source*, des *open archives*, de l'*open data*, de l'*open education*, de l'*open science* ou

---

L'action publique saisie par ses instruments », in : P. Lascoumes et P. Le Galès, *Gouverner par les instruments*, Paris, Presses de Sciences Po, 2005. », cité dans Barres (2015).

<sup>7</sup> « Voir Michel Foucault, « La technologie politique des individus » in. *Dits et écrits Tome IV*, Paris, Gallimard, 1994 pp. 813-828 Michel Foucault, Sécurité, Territoire, Population. Cours au Collège de France. 1977-1978, Paris, Gallimard Seuil, 2004 Michel Foucault, « gouvernementalité », *Dits et écrits Tome III*, Paris, Gallimard, 1994, pp.635-657 », cité dans Barres (2015 ; 2).

encore de l'*open innovation*. On pouvait s'attendre à ce que l'ouverture et le libre accès à des biens immatériels signe parallèlement le déclin du capitalisme. Or, c'est le contraire qui s'est produit car comme l'ont observé Gilles Deleuze et Félix Guattari 1987 (cité dans Petersen, 2008), le capitalisme a une "capacité inhérente (...) à se reterritorialiser et à se réinventer". En effet, les mêmes caractéristiques des technologies numériques qui encouragent l'ouverture et la '*disclosure*' des données produites par la multitude –, à savoir une barrière d'entrée très abaissée, l'interconnexion de plusieurs plateformes, la possibilité de mélanger des contenus divers, la gratuité d'accès à des contenus –, facilitent également leur captation donc l'«*enclosure*» et leur exploitation par les sociétés privées du capitalisme. En opérant par la séduction plutôt que par la coercition, ces sociétés privées propriétaires des plates-formes numériques, parviennent à inciter leurs milliards d'utilisateurs à leur fournir gratuitement et en continu, une main d'œuvre cognitive en échange de la promesse d'émancipation, de créativité, d'autonomie individuelle et en pariant avec succès sur la pulsion épistémique de ces milliards de «*producers*» (Heaton, Proulx, 2015). En effet, le paradigme de l'ouverture réunit des acteurs mus par des idéologies et logiques différentes voire contradictoires :

- i) des acteurs de la société civile (individus, associations, chercheurs), partisans du libre accès aux contenus numériques, mus par une idéologie libertaire et par un idéal démocratique ;
- ii) des acteurs de l'économie (ultra-)libérale (sociétés privées fournisseurs de contenus et de services numérique, sociétés du e-commerce) mues par une idéologie ultra-libérale et dont la survie dépend de la captation et l'exploitation des traces numériques laissées par la multitude de «*producers*» ;
- iii) des États et des institutions publiques qui se trouvent en position d'intermédiaires entre libertaires et ultralibéraux, et qui tentent de se ménager des espaces de souveraineté afin de sauvegarder leurs intérêts régaliens tout en encourageant la croissance économique et l'innovation grâce au numérique (donc les ultralibéraux) mais en n'abandonnant pas ses devoirs de service public (donc tendance libertaire).

Le constat de Yohan Barres (2015 : 7) va également dans ce sens lorsqu'il écrit :

« À l'heure où le ressort du capitalisme cognitif repose sur une économie des relations, l'État s'engage dans une forme de gouvernement des relations. Les discours économiques sur l'open data tendent ainsi à construire un idéal interrelationnel, à forger de toute pièce une communauté, supposée propice à la création de richesse. [...] En suivant cette logique on pourrait dire que l'État se pose comme un intermédiaire, à la fois pour valoriser son capital immatériel des données, manifestant par là sa volonté dans le processus d'ouverture de conserver une position dominante quand à leur utilisation potentielle et favoriser les liens entre entreprises à même de monétiser ces données en s'appuyant sur la force d'invention de la multitude ».

La réunion des acteurs aux logiques et idéologies différentes au sein du paradigme de l'ouverture et donc de la «*disclosure*» crée des tensions qui rendent incertaine son évolution. L'idéologie libertaire et altermondialistes peut parvenir, avec ou sans l'aide des États, à fédérer de plus en plus de gens autour des initiatives des «*commons*» et à imposer l'ouverture et le mode collaboratif de production des contenus numériques face à l'«*enclosure*» et à la privatisation des données prônés par les acteurs de l'ultra-libéralisme et du capitalisme. Dans ce cas, la valeur des biens immatériels et intellectuels ne résidera plus dans l'accessibilité au contenu, mais plutôt dans les données collectées sur les utilisateurs des plates-formes numériques qui les hébergent. Cette situation s'est déjà réalisée aujourd'hui. C'est bien grâce aux données amassées sur leurs utilisateurs que les sociétés

privées des plateformes numériques telles AirBnB, Blablacar, Uber ou Amazon ont réussi la désintermédiation d'anciens intermédiaires (agences de publicité, librairies, hôtels, agences de voyages, ...) en s'adressant directement à leurs consommateurs. Ces données permettent à ces plateformes d'avoir une connaissance actualisée et précise des habitudes de leurs usagers qu'elles monnaient auprès des anciens intermédiaires et annonceurs sous forme d'espaces publicitaires et de recommandations personnalisées des produits à acheter ou de contenus numériques à « consommer ».

Dans la suite de ce chapitre, nous montrerons comme la « datafication » a gagné des secteurs de nos activités tels que l'urbanisme et la culture ; secteurs qui impacteront sans doute un développement du travail social et de ses publics.

## **De l'ouverture des données publiques aux « villes intelligentes »**

Dans un article du *New York Times* au titre prophétique de « *A physicist solves the city* » (Un physicien résout la ville), Lehrer (2010) rendait compte des travaux du physicien Geoffrey West qui pensait avoir trouvé les lois fondamentales pouvant expliquer le phénomène des grandes villes, depuis leur émergence en passant par leur croissance jusqu'à leur déclin. Geoffrey West était convaincu qu'il était possible de découvrir les règles qui gouvernent toute chose. En effet, comprendre le fonctionnement des zones urbaines est important pour la planification des villes car selon des estimations, d'ici 2050, 75 % de la population mondiale vivra dans des métropoles. Après deux années passées à analyser des données massives collectées sur plusieurs grandes villes du monde, Geoffrey West annonçait les lois gouvernant les métropoles qu'il a comparées à un énorme organisme vivant. Un des résultats auquel il a abouti est le constat que les grandes villes sont moins consommatrices d'énergie en hiver que les banlieues parce que ces dernières sont moins densément peuplées, donc elles ont besoin de chauffer plus en hiver. Par conséquent, les grandes villes seraient plus écologiques. En suivant ce raisonnement, on aboutit à la conclusion logique qu'il faudrait agrandir les métropoles ce qui paraît contre-intuitif au vu de la densité des infrastructures et appareils électroniques par tête d'habitants dans les grandes villes. Si on accélère la croissance des métropoles, il faut aussi accélérer la vitesse des innovations afin de renouveler les ressources énergétiques trop vite épuisées.

Comme l'a observé le Lehrer (2010), les lois que le physicien Geoffrey West pense avoir trouvées ne sont que des corrélations qui expliquent quelques aspects des grandes villes, mais pas ce phénomène dans son intégralité qui ne saurait être réduit à quelques lois mathématiques. Par exemple, les lois de West n'expliquent pas pourquoi des zones de basse densité de population comme la Silicon Valley produisent énormément de richesses et sont devenues des lieux importants d'innovations aux États-Unis. La modélisation du phénomène des grandes villes à l'aide des données massives est problématique parce qu'elle abstrait des villes ce qui fait leur essence, à savoir les individus et les contextes socio-culturel, historique et social dans lesquels ils vivent. L'étude de Geoffrey West a paradoxalement mis en exergue l'importance des facteurs humains et sociaux dans la croissance des zones urbaines. En effet, il a découvert que la raison principale qui motivait les gens à habiter les métropoles n'était pas la réduction de leur facture d'électricité mais plutôt les besoins du lien social. C'est pour rencontrer d'autres gens ou pour trouver du travail que les gens vont vivre dans des grandes villes. Son propre Institut de recherche interdisciplinaire à Santa Fé l'illustre car Geoffrey West a remarqué que la salle la plus fréquentée était la salle de café.

La ville de New Orleans n'est pas simplement l'archétype des villes de 378,713 habitants<sup>8</sup>. New Orleans « évoque aussitôt l'ouragan Katrina, la cuisine afro-américaine, l'architecture créole française, une population multilingue et multi-culturelle et des problèmes résiduels de ségrégation dans le sud des États-Unis » (Lehrer, 2010). Le fait d'appliquer aux humains et aux organisations sociales, des méthodes et modèles qui fonctionnent pour l'observation de phénomènes macroscopiques comme les épidémies, les astres, la formation des roches, comporte des risques que le sociologue Gabriel Tarde avait identifiés dès le début du XX<sup>e</sup> siècle : c'est comme regarder le monde à 10 000 m d'altitude. Il y a forcément des illusions de perspective. L'historien et sociologue américain Lewis Mumford (1937) observait déjà que le design physique et les fonctions économiques des villes sont secondaires par rapport à leur relation à l'environnement naturel et aux valeurs spirituelles des communautés humaines. Ainsi, la question à se poser, écrivait-il n'était pas « qu'est-ce que la ville ? » (« *what is the city ?* », mais plutôt « qu'est-ce que la ville en tant qu'institution sociale ? (« *what is the city as a social institution ?*). « La ville dans son sens le plus complet, est un plexus géographique, une organisation économique, un processus social, un théâtre d'action sociale et un symbole esthétique d'unité collective » (Mumford, 1937 ; 92).

Grâce aux infrastructures technologiques numériques qui irriguent désormais les métropoles, les villes sont devenues productrices de masses de données que les planificateurs urbains rêvent d'exploiter pour les modéliser en retour. Des projets de « smart city » et d' « intelligent city » se multiplient en France comme à l'étranger. Les objectifs affichés paraissent tous louables. Il s'agit de « rendre la ville plus durable et plus vivable » grâce à « une meilleure gestion de l'énergie [...] favorisant l'usage des énergies renouvelables – une meilleure gestion de la mobilité pour limiter le recours à l'automobile et, ainsi, limiter les émissions des gaz à effet de serre<sup>9</sup> ». Les villes de demain seront donc pilotées par des équipements connectés qui vont relever des données de toutes sortes permettant de connaître, *via* son smartphone, l'état de la circulation, des places de stationnement ou l'horaire de passage des transports en commun. L'état français a inscrit dans la loi de transition énergétique du 18 août 2015, l'installation chez chaque habitant, de nouveaux compteurs électriques connectés et « intelligents », baptisés du nom séduisant de Linky<sup>10</sup>. Linky serait en mesure d'automatiser le relevé des données et la facturation des consommations d'électricité, de réguler la consommation énergétique des logements en fonction des habitudes des habitants, tout ceci sans nécessiter le déplacement des agents humains. Il est possible que ces nouveaux équipements électroniques permettent de faire des économies d'énergie et lutter contre le gaspillage des ressources essentielles comme l'eau et l'électricité mais parallèlement, cette automatisation de la capture des données et la réparation à distance risque de s'accompagner des pertes d'emploi des agents des sociétés de fourniture d'électricité. De plus, ces données individuelles collectées peuvent permettre d'inférer les habitudes quotidiennes de chaque usager. Leur collecte suscite des craintes de la part de la société civile car les modalités de leur cession à des sociétés tiers et leurs usages futurs n'ont pas fait l'objet d'un débat public (Leloup, Eeckhout, 2016). Suivant la logique de la monétisation qui prévaut actuellement, il est fort probable qu'une partie de ces données privées soient cédées à des entreprises tierces qui s'en serviront pour créer de nouveaux produits et services commercialisés qu'elles revendront ensuite aux individus propriétaires de ces mêmes données qui ont servi à élaborer ces produits et services innovants. Les

<sup>8</sup> D'après le dernier recensement datant de 2013.

<sup>9</sup> Pour des villes plus durables, « Les villes intelligentes », *La Recherche*, Avril 2015, n°498, pp. 69-70.

<sup>10</sup> « La loi de transition énergétique du 18 août 2015 prévoit le remplacement de 35 millions de compteurs classiques par des Linky, d'ici 2021 ». D. Leloup et L. Van Eeckhout, *Le Monde*, 7/04/2016.

usagers seront donc à la fois producteurs des données et clients-consommateurs des produits élaborés grâce à leurs propres données. Au-delà des considérations économiques, le concept même de “smart” ou de “intelligent city” relève fondamentalement d’une conception positiviste et physique des villes, vues comme des entités dont on peut calculer les règles de fonctionnement à coup de données, dans le but de modéliser leurs évolutions. Rob Kitchin (2014b : 181<sup>11</sup>) résume ainsi cette conception physique des villes : « Comme le dit, (Hill 2013), le concept de “smart city” reflète une conception technocratique que la ville pourrait être cernée dans ses moindres détails, tel un moteur ou une station nucléaire, si seulement nous avions suffisamment de données, et ainsi la maîtriser par la force brute de la science et de l’ingénierie. En effet, Mattern (2013) suggère que l’urbanisme de *big data* souffre de la “datafication”, à savoir cette présomption que tous les flux significatifs et activités peuvent être captés et mesurés’. Il y a dans cette conception, souvent un présupposé explicite que l’univers est constitué des paramètres connaissables et définissables et que si nous pouvions seulement les mesurer tous, nous garantirions que nous serions en mesure de prédire et de répondre parfaitement en conséquence’ (Haque, 2012). Les solutions technologiques toutes seules ne pourront pas résoudre des problèmes structurels profonds des villes car elles ne s’attaquent pas à leurs causes. Elles ne permettent que la gestion plus efficace des manifestations de ces problèmes. Alors que les technologies de “*smart city*” telle l’analyse en temps réel sont promues comme la panacée pour résoudre des problèmes de gouvernance urbaine, elles ne font que couvrir les fissures au lieu de les réparer [...] à moins de les accompagner par des politiques sociales adaptées à chaque situation locale ».

Observons également que toute modélisation d’un phénomène conduit à gommer ses éléments distinctifs et à élaborer des solutions de type “taille unique” (“*one size fits all smart city*”), implémentées dans des boîtes noires algorithmiques qui ne tiennent pas compte du caractère unique des lieux, des gens, des cultures et de l’histoire. Tout comme les sociétés des plateformes du web ont d’abord capté les internautes avant de monétiser les données collectées sur eux auprès d’autres sociétés privées pour assurer leur survie économique, les projets de “*smart cities*” peuvent à terme créer « une situation d’enfermement technologique (“*technological lock-in*”) qui rend les administrations des villes dépendantes de certaines plateformes technologiques, engendrant par là des situations de monopole (Hill 2013) » et conduisant les villes à rentrer dans « un mode étroit et technocratique de gouvernance (Townsend, 2013) » Kitchin (2014b : 181).

## La physique de la culture

Un autre secteur qui est actuellement sous l’emprise de la « datafication » est celui de la culture. Les données produites par nos activités culturelles intéressent fortement les « *data scientists* » qui voient là une occasion pour appliquer leurs outils de prédiction à un objet qui était jusqu’ici difficile à modéliser. C’est ainsi que la société britannique Epagogix a mis au point un algorithme d’apprentissage par réseaux de neurones qui porte le même nom. Présenté comme un « outil de management de risque » (*risk management tool*), Epagogix est censé éliminer des facteurs humains, passionnels et affectifs qui « obscurcissent » le processus de prise de décisions. Appliqué à l’industrie cinématographique, Epagogix doit fournir aux studios hollywoodiens un moyen objectif pour identifier les paramètres des films à succès de sorte à les introduire en amont dans de nouveaux *scenarii*. Le modèle construit par les réseaux neuronaux d’Epagogix est évolutif, à savoir qu’il s’ajuste en fonction des données rentrées périodiquement dans son corpus

---

<sup>11</sup> C’est nous qui traduisons, l’ouvrage étant écrit en langue anglaise.

d'apprentissage (*training set*). Ainsi, un même scénario rentré dans l'algorithme à un temps  $t$  ne recevra pas nécessairement le même score de probabilité de succès au temps  $t+1$  si de nouvelles données ont été rentrées dans l'algorithme entretemps. Le nom de cette société et de son algorithme n'est pas le fait du hasard. Epagogix fait référence à « epagogue », la vision épistémologique aristotélicienne qui donne la primauté à l'expérience et aux observations comme moyen d'accéder à la connaissance, dans une démarche inductive. Epagogix est donc évocateur de l'épistémologie *empiriciste* promue par Aristote. Faut-il penser que la société cherche par-là à se doter d'une aura scientifique, en débarrassant le processus de sélection des *scenarii* des impuretés des facteurs émotionnels humains, afin que les studios hollywoodiens puissent se concentrer uniquement sur les « *hard facts* » ou les « *cold data* » ?

C'est grâce aux données dont il a été nourri sur la fréquentation des salles de cinéma nord-américaines qu'Epagogix a « découvert » que la présence des effets spéciaux était plus déterminant pour le succès des films que la présence des grandes stars pour le *box office* américain et canadien. Ainsi, un des résultats de cet algorithme serait que c'est le scénario qui fait les stars et non l'inverse. Une des conséquences de cette « découverte » est qu'elle modifiera les rapports de force dans l'industrie cinématographique. Epagogix peut fournir aux grands studios hollywoodiens un levier pour renégocier le cachet des célébrités. La société reconnaît toutefois que son algorithme n'arrive pas (encore) à prédire correctement la probabilité des films musicaux ou des films à petit budget et à petite distribution, à être des films à succès, faute de données suffisantes. Ainsi, la datafication est-elle directement dépendante de la disponibilité des données massives sur les activités d'un secteur donné, ici le cinéma.

Au-delà des considérations économiques pour les grands studios hollywoodiens, le recours aux algorithmes de *big data* dans le processus de production des films a une implication épistémologique inquiétante : elle transfère aux codes informatiques, une expertise qui était jadis celle des humains et introduit une quantification et des probabilités sur des attributs éminemment humains, à savoir l'imagination et la créativité intellectuelle. En fonction de la pondération des attributs des *scenarii* des films, l'algorithme va privilégier certaines histoires au détriment des autres. Le dirigeant d'Epagogix (Nick Meaney) affirmait que la plupart des films qui remportent du succès à Hollywood sont quasiment des histoires de Roméo et de Juliette sur des décors différents (urbain, désert, plage...). Epagogix est emblématique de cette nouvelle génération de sociétés de *data science* dont les domaines d'intervention s'étendent à quasiment tous les secteurs, de l'assurance à la finance en passant par la santé, la culture et la gastronomie.

Cette « datafication » de l'univers cinématographique ne se limite pas à la phase amont de l'écriture et de la sélection des *scenarii*. Elle a aussi gagné la phase aval, à savoir la promotion et la distribution des produits cinématographique. En effet, ce sont à nouveau des algorithmes qui orientent les usagers des plateformes de distribution de films vers la consommation de certains films plutôt que d'autres. La plateforme Netflix avait employé l'algorithme « Pragmatic Chaos »<sup>12</sup> pour prédire des films que les gens étaient susceptibles de regarder et donc placer ces films en tête de leurs fenêtres de Netflix. La société investit actuellement dans le développement d'algorithmes de prédiction qui permettent de

---

<sup>12</sup> Ugur, Candan, « Should Pragmatic Chaos determine the future of human entertainment ? ». Accessible en ligne à <http://ugurcandan.net/cool/future-of-human-entertainment/>. Cet article fait état du prix Netflix de 1 million de \$ remporté par Pragmatic Chaos en 2009 pour son taux de prédiction correcte de films que les gens voudraient regarder.

personnaliser davantage la fenêtre Netflix de chaque usager en fonction de nombreux paramètres<sup>13</sup>. La boucle est donc bouclée : en amont, les algorithmes tel Epagogix oriente les studios hollywoodiens sur le choix des *scenarii* à tourner en films ; en aval, des algorithmes tel Pragmatic Chaos oriente les consommateurs vers les films à regarder (probablement les mêmes promus en amont de la production). Cet encerclement de l'activité culturelle par les algorithmes de big data a été qualifié de « physique de la culture » par Kevin Slavin (2011<sup>14</sup>).

Après l'industrie cinématographique, la « datafication » gagne d'autres institutions culturelles. De nombreux musées aux États-Unis se sont lancés dans la collecte des données de leurs visiteurs pour à la fois élaborer des stratégies publicitaires et promotionnelles et également pour établir la programmation de leurs musées (Gamerman, 2014). Ainsi, le *Dallas Museum* a-t-il offert deux ans d'entrée gratuite à ses visiteurs en échange de la collecte de leurs données personnelles (email, adresse, numéro de téléphone, code postal). De même, le *Los Angeles County Musuem of Art* (Lacma) n'hésite pas à aller chercher les visiteurs là où ils se trouvent. À l'aide de relais électroniques disséminés dans la ville, des gens assis sur une terrasse de café peuvent recevoir des messages sur leurs smartphones leur donnant des informations (non sollicitées) sur une œuvre d'art qui se trouve à proximité. Remarquez que ces nouveaux services numériques fondés sur l'exploitation des données des visiteurs sont souvent l'initiative des « *digital media officers* » et non des conservateurs des musées, des artistes ou des visiteurs eux-mêmes. Comme l'observe Tiffany Jenkins (2013) : « La montée du *big data* est accompagnée par la présomption que tout peut et devrait être mesuré. Même dans mon champ – les arts – les organisations d'arts et les artistes sont de plus en plus soumis à l'injonction de rendre compte de leurs travaux sous forme chiffrée. S'ils ne le font pas, ils ne sont pas subventionnés. Mais la valeur des arts, la qualité d'une pièce de théâtre ou d'une peinture, n'est pas mesurable. Vous pouvez injecter toutes sortes de données dans une machine : dates, couleurs, images, tickets, et aucune de ces données ne peut expliquer c'est que signifie une œuvre, et pourquoi elle est puissante. Cela nécessite l'homme {sic}, pas la machine... Alors que je suis convaincue qu'il y a une place pour le *big data*, elle ne s'applique pas à plusieurs aspects de la vie humaine. Elle n'améliorera pas la condition humaine. L'amour, l'art, la culture et la politique sont tous essentiels à nos vies mais le *big data* nous dira très peu de choses sur ces questions ».

La « datafication » de la culture peut à terme engendrer des mutations profondes dans le travail des professionnels de ce secteur qui seront davantage guidés par des considérations de rentabilité que par des considérations artistiques, historiques ou patrimoniales. L'art et la culture deviendront *in fine* des produits à consommer tels les chaussures ou les vêtements, produits en série mais adaptés à des catégories de consommateurs. Cela n'est pas sans rappeler les critiques formulées à l'encontre de la culture de masse il y a plus d'un demi-siècle par Theodor W. Adorno et Max Horkheimer (1947 ; 13) : « Les distinctions emphatiques établies entre des films de catégories A et B, ou entre des histoires publiées dans des magazines de différents prix ne se fondent pas tant sur leur contenu même que sur la classification, l'organisation des consommateurs qu'ils permettent ainsi d'étiqueter. Il a été prévu quelque chose pour chacun, afin que nul ne puisse y échapper ». Theodor W. Adorno avait précisément refusé de rentrer dans le cadre

---

<sup>13</sup> Lara O'Reilly, Netflix lifted the lid on how the algorithm that recommends you titles to watch actually works. *Business Insider*, 26 Feb. 2016, <http://www.businessinsider.fr/us/how-the-netflix-recommendation-algorithm-works-2016-2/>.

<sup>14</sup> Kevlin Slavin, How Algorithms shape our world, TED Talk, 2011, <https://www.ted.com/talks/kevinslavinhowalgorithmsshapeourworld?language=en>

empirique étrié dans lequel le sociologue fonctionnaliste Paul Lazarsfeld avait enfermé l'étude des effets des médias de masse sur la population américaine. Adorno dira plus tard de sa collaboration avec Lazarsfeld : « Quand j'ai été confronté, à l'exigence de « mesurer la culture », j'ai vu que la culture devait précisément être cette condition qui exclut une mentalité capable de la mesurer » (cité dans Mattelart A. et M., 1995 : 41-42).

## Vers des humanités numériques ?

La datafication a déjà gagné l'activité scientifique elle-même. Les concepts de *e-Science*, de *cyberscholarship* et de *data-driven scholarship* sont révélateurs de cette tendance vers une recherche scientifique orientée par les données massives qui va de pair avec la croyance que la signification peut être indépendante du contexte et des connaissances des domaines concernés (Ibekwe-SanJuan 2014). C'est la dimension mythologique du *big data* évoquée par Danah Boyd et Kate Crawford (2012: 664). Cette croyance a déjà gagné des disciplines des sciences du vivant où des « *data scientists* » ont conquis des terrains autrefois occupés par les spécialistes de ces disciplines en prétendant pouvoir produire des connaissances sans avoir de connaissances sur ces domaines, mais uniquement en faisant ingurgiter des masses de données à leurs algorithmes. Ainsi, les bio-informaticiens qui ont une vision assez étroite de la biologie ont-ils gagné des terrains auparavant occupés par les cliniciens et les biologistes moléculaire ou expérimental. Cela entraîne un repositionnement des acteurs scientifiques et soulève une question épistémologique sur qui est le plus légitime producteur de connaissance, le spécialiste d'un domaine ou le *data scientist* ? (Kitchin 2014a). Cette croyance explique l'invasion des Sciences humaines et sociales (SHS) par des chercheurs issues des sciences dites « dures » (mathématiques, physique, statistique) qui prétendent pouvoir modéliser les phénomènes sociaux, historiques et culturels complexes avec leurs outils de mesure, en les réduisant à une série de formules mathématiques et à des corrélations.

La montée en puissance des projets de « Digital Humanities » (Humanités Numériques) témoigne de cette volonté d'arrimer les méthodes de recherche en SHS sur celles des sciences formelles. Défini comme « l'application du "savoir-faire des technologies de l'information (et de l'informatique/infosciences) aux questions des sciences humaines et sociales"<sup>16</sup>», le terme « humanités numériques » (HN) apparaît comme un oxymore car il sous-entend que des problématiques humaines et sociales peuvent être soumises à un mode de traitement algorithmique. Il est par ailleurs révélateur que la communauté des « digital humanists » considère comme un de ses travaux fondateurs le travail de numérisation et d'alignement de l'œuvre de Thomas d'Aquin par le moine jésuite Roberto Busa en 1949. Devant l'immensité de la tâche, Roberto Busa a fini par solliciter le concours de la société IBM pour accélérer ce travail qui leur a tout de même pris trente ans, compte tenu de la capacité des ordinateurs de l'époque. Les projets des HN des universités prestigieuses souvent cités en exemple comme celui de l'Université de Stanford présente les *Digital Humanities* comme étant « à la croisée de l'informatique et des humanités ». L'objectif clairement affiché est d'outiller les SHS afin de numériser des corpus et favoriser les recherches collaboratives et interdisciplinaires sur ces corpus. À l'Université d'Oxford, l'accent est à nouveau mis sur la numérisation de textes rares et anciens et sur l'outillage

---

<sup>16</sup> Olivier Dessibourg, Venise, machine temporelle, Selon Frédéric Kaplan, professeur en humanités numériques à l'École Polytechnique fédérale de Lausanne, Le Temps, n° 4538, samedi 23 février 2013, pages 17-18, cité dans l'article Wikipédia consacré aux Humanités Numériques : [https://fr.wikipedia.org/wiki/Humanités\\_numériques#cite\\_note-3](https://fr.wikipedia.org/wiki/Humanités_numériques#cite_note-3).

informatique<sup>17</sup>. L'emprise du paradigme computationnel et physique sur le champ des HN explique sans doute pourquoi ce champ est porté par les informaticiens plutôt que par les chercheurs en SHS.

Sur le plan épistémologique, Suzanne Dumouchel (2015) se demandait à juste titre si la simple application des outils numériques à un secteur d'activité suffit à elle seule à justifier de la création d'un nouveau champ scientifique. Le numérique a bouleversé tous les secteurs d'activité dont ceux des disciplines dites de « sciences dures ». Pourtant, elles n'ont pas créé un nouveau champ de « Sciences numériques ». L'auteur se demande donc « que reste-t-il des HN ? Pourquoi faudrait-il un nom à ces nouveaux usages ? En quoi se distinguent-ils d'un usage du numérique pour la recherche en SHS ? » (Dumouchel, 2015).

Actuellement, les HN se trouvent à la croisée de deux chemins. Elles peuvent représenter une opportunité pour avancer l'agenda des SHS en ouvrant de nouvelles perspectives qui compléteront la recherche traditionnelle plutôt qualitative en SHS. Il est indéniable que la mise au format numérique de grandes collections de textes ou d'images rares ou anciennes permet non seulement de les valoriser mais autorise de nouveaux traitements qui multiplient les possibilités de recherches et de découvertes inopinées que l'exploration manuelle de ces corpus par des chercheurs ne pourra permettre. Par exemple, il est tout à fait impossible à un chercheur d'explorer manuellement un corpus de plusieurs millions de tweets ou de vidéos *YouTube*, de textes de romans ou d'articles de journaux. L'exploration automatique permet d'identifier des motifs récurrents ou des traits saillants qui feront l'objet d'analyses plus approfondies par des analystes humains. Plusieurs décennies de recherche sur la linguistique de corpus montrent l'intérêt de l'approche automatique de l'analyse des corpus<sup>19</sup>. Ainsi, l'approche des HN « ouvrirait des possibilités pour des études de surface à grande échelle' de même que, et non 'au lieu' des études approfondies (portant sur un petit nombre d'individus' » (Manovich 2011, cité dans Kitchin 2014b: 146).

Si les machines ont cette capacité d'accélérer les rythmes de découvertes à partir de masses de données, elles peuvent conduire à terme au remplacement des centaines de chercheurs en SHS. Gageons qu'il en sera de même pour d'autres professions ou publics. Les propos de Ramsay (2010) cité dans Kitchin (2014b : 147) ne sont pas de nature à lever les inquiétudes exprimées par plusieurs chercheurs à propos du tournant des HN : « Pour Ramsay (2010) la computation peut assister le processus d'interprétation. Il soutient que la lecture attentive implique les pratiques de sélectionner, d'isoler et d'observer des motifs dans les textes et le code informatique est très efficace dans ces tâches, et de plus, il le fait de manière objective et systématique que le regard subjectif d'un chercheur. Plutôt que de remplacer l'identification de la signification, pour lui, l'emploi des algorithmes facilite un tel processus ».

En effet, pourquoi chercher à éliminer le « regard subjectif du chercheur » alors que le facteur subjectif est précisément ce qui caractérise la recherche en SHS comme dans d'autres secteurs professionnels où la dimension humaine et le rapport direct à l'autre sont une compétence de base ? C'est le regard humain qui permet de faire sens des pratiques, des événements, des faits et des discours situés dans des contextes géographique et historique précis. En quoi l'analyse « objective » par un algorithme est-elle préférable au regard d'un

---

<sup>17</sup> <https://digital.humanities.ox.ac.uk/>.

<sup>19</sup> Une des associations savantes en pointe dans les recherches sur les outils automatiques de traitement de corpus est l'ACL (*Association for Computational Linguistics*) fondée en 1962. Site web : <https://www.aclweb.org/portal/what-is-cl>. Il en existe des dizaines d'autres, souvent par pays.

spécialiste humain si ce n'est qu'elle gomme toutes les spécificités et les contextes ? Par ailleurs, la prétendue objectivité des algorithmes souvent mise en avant dans ce genre de discours est à relativiser. Les méthodes d'analyse automatique des données comportent de nombreux biais méthodologiques. Cette recherche d'objectivité en SHS reflète une pensée positiviste qui présuppose une seule façon de faire une analyse et une seule réponse possible à une question de recherche. Or, les SHS sont caractérisées par la cohabitation d'une pluralité de visions épistémologiques, conceptuelles et méthodologiques au sein d'une même discipline et parfois au sein d'un même champ ou d'une même spécialité (Dousa, Ibekwe-SanJuan 2014). Ce triple pluralisme rend délicat tout essai de modélisation des problématiques des SHS.

Par ailleurs, à trop vouloir rendre la recherche en SHS dépendante des corpus de données et des algorithmes capables de les explorer, le risque est de conditionner les questions de recherche en SHS à ce que les algorithmes estiment comme « remarquable » au sens statistique du terme. Cela reviendra à déléguer aux algorithmes le pouvoir de déterminer les questions de recherche qui méritent d'être investiguées. Et compte tenu que le nombre de corrélations intéressantes trouvées par les algorithmes sur des collections de données massives se compte par centaines de milliers, les analystes humains seront vite dépassés et ne disposeront pas d'indices fiables pour déterminer les corrélations qui valent la peine d'être investiguées davantage. En effet, toute question posée sur le big data risque de retourner de nombreuses corrélations statistiquement significatives.

Il faut donc mettre au point de nouvelles méthodes, en dehors des algorithmes de big data, pour tester les hypothèses induites par l'exploration de grandes masses de données, ce qui suppose de sortir des données pour aller à la rencontre des objets d'étude dans le monde réel.

Les algorithmes de traitement automatique des textes ne peuvent (encore) déchiffrer le contexte extérieur aux textes, déceler l'ironie et le sarcasme ou rendre compte de l'histoire des œuvres, de leurs réceptions et usages. Toutes ces questions qui sont finalement des questions sociétales continueront de nécessiter des éclairages des spécialistes humains. Les problématiques les plus intéressantes en SHS ne se réduisent pas à l'exploration des traits saillants des grandes collections de données, elles résident plus dans la compréhension des phénomènes humains et sociaux. En définitive, il semble important pour les SHS de conserver leur indépendance vis-à-vis du paradigme physique et computationnel, en veillant à ce qu'une proportion non négligeable de ses problématiques ne soit pas contingente à l'existence de grandes masses de données ou que celles-ci ne deviennent pas la seule échelle à laquelle il sera désormais valable de mener une activité de recherche.

## RÉFÉRENCES

Adorno Th. W., Horkheimer M., 1947, « Kulturindustrie, Raisons et mystification des masses », (Traduit de l'allemand par Eliane Kaufholz sous le titre *La production industrielle des biens culturels*, in *Les dialectiques de la raison, Fragments philosophiques*), Paris, Gallimard.

Barres Y., 2015, « De quoi l'Open Data est-il le nom ? » Colloque Big Data - Open Data. Quelles valeurs ? Quels enjeux ?, Texte d'une communication au colloque « Document Numérique et Société », Rabat, 3-4 mai 2015, 9 pages (non publié).

Ibekwe-SanJuan F., (2017), Vers la datafication de la société ?, in Vincent Meyer (eds), *Transition digitale, handicaps et travail social*, LEH Edition, Bordeaux, pp. 31-49.

Borgman Ch., 2015, *Big Data, Little Data, No Data: scholarship in the networked world*, Cambridge Massachusetts, MIT Press.

Boyd D., Crawford K., 2012, « Critical questions for Big Data », *Information, Communication & Society*, 2012, vol. 15(5), pp. 662-679.

Bowker G., 2014, « The Theory/Data Thing. Commentary », *International Journal of Communication*, vol. 8, pp. 1795–1799.

Buckland Michael, « Information as thing », *Journal of the American Society for Information Science*, 1991, vol. 42(5), 351–360.

Dousa T. Ibekwe-SanJuanF., 2014, « Epistemological and Methodological Eclecticism in the Construction of Knowledge Organization Systems (KOSs). The Case of Analytico-synthetic KOSs », pp. 152-159, in : Knowledge organization in the 21st century. 13th International Conference (ISKO 2014), Krakow, Poland, 19-22 mai 2014..

Dumouchel S., Digital Humanities à l'IHA, Hypothèses.org, 3/07/2015, [<http://dhiha.hypotheses.org/1539>]. Consulté le 31/03/2017.

Frické M., « Big Data and Its Epistemology », *Journal of the Association for Information Science and Technology*, 2015, 66(4): 651–661.

Gamerman E., 2014, « When the Art Is Watching You », *Wall Street Journal*, 11 décembre 2014.

Jenkins T., « Don't count on big data for answers », *The Scotsman*, 12 février 2013, <http://www.scotsman.com/the-scotsman/opinion/comment/tiffany-jenkins-don-t-count-on-big-data-for-answers-1-2785890>. Consulté le 31 mars 2017.

Leloup D. Van Eeckhout L., « Faut-il se méfier des compteurs Linky ? », *Le Monde*, 07/04/2016. [http://www.lemonde.fr/planete/article/2016/04/07/faut-il-se-mefier-des-compteurs-linky\\_4898239\\_3244.html#T7GEhFf1W8PsrHWZ.99](http://www.lemonde.fr/planete/article/2016/04/07/faut-il-se-mefier-des-compteurs-linky_4898239_3244.html#T7GEhFf1W8PsrHWZ.99). Consulté le 1/04/2017.

Heaton Lorna et Proulx Serge, « Paradoxical Empowerment: Immaterial Labor Translated in a Web of Affective Connections », *The Information Society: An International Journal*, 2015, 31:1, pp. 28-35, DOI: [10.1080/01972243.2015.977629](https://doi.org/10.1080/01972243.2015.977629).

Ibekwe-SanJuan F., « Big Data, Big machines, Big Science : vers une société sans sujet et sans causalité ? », 19<sup>e</sup> congrès de la SFSIC, Toulon 4-6 juin 2014, 9 pages

Kithin R, «Big Data, new epistemologies and paradigm shifts », *Big Data & Society*, April–June 2014a: 1–12, DOI: [10.1177/2053951714528481](https://doi.org/10.1177/2053951714528481).

Kitchin Rob, 2014b *The Data Revolution. Big Data, Open Data, Data infrastructures & their consequences*, London, Sage.

Lehrer J., 2010, « A Physicist solves the City », *New York Times*, 17 Déc. 2010

Mattelart A. et M., 2004, *Histoire des théories de la communication*, Paris, Éd. La Découverte. Coll. Repères.

Ibekwe-SanJuan F., (2017), Vers la datafication de la société ?, in Vincent Meyer (eds), *Transition digitale, handicaps et travail social*, LEH Edition, Bordeaux, pp. 31-49.

Mumford L., 1937, « What is a city? », *Architectural Record*, pp. 92-96.

Petersen S.-M., 2008, « Loser Generated Content: From Participation to Exploitation », *First Monday*, Critical issues on web 2.0, 2008, vol. 13(3), <http://firstmonday.org/ojs/index.php/fm/article/view/2141>.