



HAL
open science

Big Data : stratégies industrielles et économie de la promesse

Vincent Bullich

► **To cite this version:**

Vincent Bullich. Big Data : stratégies industrielles et économie de la promesse. La comunicazione multipla. Media, piattaforme digitali, Over the Top, Big Data, Magna Carta Edizioni, pp.41-71, 2016. hal-01940630

HAL Id: hal-01940630

<https://hal.science/hal-01940630v1>

Submitted on 26 Aug 2024

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.

***Big Data* : stratégies industrielles et économie de la promesse.**

Vincent Bullich

Maître de conférences en sciences de l'information et de la communication – UFR des sciences de la communication Paris-Nord 13

Laboratoire de rattachement : Labsic – Paris Nord 13 EA 1803

Adresse personnelle : 339, rue Nicolas Parent – 73000 Chambéry

Tel : +33616407681

vincentbullich@hotmail.com

Maître de conférences en sciences de l'information et de la communication, membre du laboratoire des sciences de l'information et de la communication (LabSic - Paris-Nord 13), de l'Observatoire des mutations des industries culturelles de la Maison des Sciences de l'Homme Paris Nord (OMIC – MSH PN) ainsi que du laboratoire d'excellence « Industries culturelles et création artistique » (Labex ICCA). Ses travaux concernent principalement l'analyse socio-économique des industries culturelles et communicationnelles ainsi que l'économie politique de la propriété intellectuelle.

Plan :

Plan :	1
Introduction	2
1. Qu'appelle-t-on « <i>big data</i> » ?	3
1. 1. Perspectives sémantiques	3
1. 2. Du mode d'existence des « <i>big data</i> »	4
2. Usages socio-économiques des <i>big data</i> : procès de production et procès de valorisation	6
2. 1. Le procès de production	6
2. 2. Une double rationalité économique	8
3. Des discours s'inscrivant dans une « économie de la promesse »	10
3. 1. Les <i>big data</i> comme les futures « ressources essentielles »	11
3. 2. La prédictibilité comme solution économique « ultime »	12
Conclusion	14
Bibliographie	14

***Big Data* : stratégies industrielles et économie de la promesse.**

Introduction

Il ne se passe pas un jour sans que des voix en provenance des mondes professionnels, journalistiques voire de la recherche ne nous annoncent une nouvelle « révolution ». Continuellement galvaudé, le terme a en effet intégré une panoplie argumentative et rhétorique systématiquement mobilisée par les auteurs (*lato sensu*) désireux de souligner les caractères « disruptif » et spectaculaire des innovations sociotechniques liées aux réseaux et modes numériques de communication¹. Nous serions donc désormais parvenus à « la révolution des *data* » (Kitchin, 2014) ou des *big data* (Mayer-Schönberger, Cukier, 2014), révolution qui va de pair avec une autre révolution, celles des algorithmes (Cardon, 2015). Une nouvelle fois, on annonce ainsi l'avènement d'un ordre nouveau du monde sous les effets conjugués de stratégies industrielles inédites et d'une technique à même de modifier radicalement le rapport à soi, aux autres et au monde. Si ce type de propos ne manque pas de séduction, ils entretiennent toutefois une relation ambiguë avec la réalité. Il est en effet admis que les *data* prennent une importance croissante dans de nombreux domaines, et tout particulièrement dans ceux qui se rapportent aux activités économiques. Le phénomène n'est pas nouveau mais il s'accroît de façon remarquable depuis près d'une décennie. Néanmoins, et ainsi que cet article se propose de l'exposer, les principaux discours qui se rapportent aux *big data* dans l'espace public ont tendance à manifester une surenchère langagière plutôt qu'à faire montre de rigueur analytique et interprétative. Ces discours, qui contribuent de façon prépondérante à la diffusion de la notion « *big data* », proviennent en grande partie de cabinets d'audit et de conseils (avant d'être repris et par les journalistes et autres observateurs), cabinets qui promeuvent activement le syntagme depuis plus de quinze années (Simon, 2015).

C'est précisément cet aspect que nous allons en partie traiter dans le cadre de cet article. Il s'agit ainsi d'appréhender le phénomène *big data* à partir d'une étude des manifestations concrètes du phénomène, manifestations qui résultent principalement de stratégies d'acteurs industriels, mais également à partir d'une analyse des discours en provenance de ces cabinets, discours qui ont inscrit les *big data* dans « une économie de la promesse ». Nous chercherons, par conséquent, à montrer comment ce phénomène procède tout à la fois « de jeux d'acteurs et de jeux d'arguments » (Chateauraynaud, 2011, p. 81) et viserons ainsi éclairer les conditions par lesquelles il s'impose dans l'espace public et social.

Pour ce faire, cet article se compose de trois temps. Le premier consiste en une investigation sémantique et pragmatique du mot-valise « *big data* ». L'objectif est d'identifier puis d'exposer les traits saillants, les qualités attribuées au syntagme et qui font consensus dans la communauté de chercheurs et d'experts. Dans un second temps, nous nous intéresserons aux manifestations concrètes du phénomène, manifestations abordées par le prisme des actions qui s'y rapportent. Seront ainsi détaillés les principaux usages socio-économiques des *big data* tels que sont susceptibles de produire les acteurs industriels. Dans un troisième temps, nous nous focaliserons sur les discours énoncés par les cabinets de conseils et d'audit. Ce choix est motivé par leur rôle prédominant dans la diffusion de la notion dans les milieux décisionnaires économiques et politiques. Ils apparaissent ainsi comme les principaux promoteurs du phénomène, et nous exposerons notamment dans cette dernière partie les

ressorts de la promesse que leurs rapports et études formulent. En conclusion, nous reviendrons sur les principaux points abordés et sur la nécessité d'une régulation juridique du phénomène.

1. Qu'appelle-t-on « *big data* » ?

Le syntagme « *big data* » est devenu, en l'espace de quelques années, un maître-mot, un élément de langage récurrent dans les paroles d'acteurs industriels comme ceux des observateurs et analystes (journalistes, consultants et universitaires). Toutefois, une revue de littérature (s) - scientifique, professionnelle et journalistique - permet de constater le flou sémantique, l'imprécision qui entoure le syntagme. Cette première partie cherchera donc à stabiliser l'acception du terme. Il ne s'agit pas ici de proposer une définition positive de l'expression, démarche qui aurait comme conséquence d'en figer le sens et d'en réduire ainsi l'amplitude sémantique. Plus modestement, il s'agit, premièrement, d'indiquer les aspects qualifiants du syntagme qui font consensus parmi la communauté de chercheurs en sciences socialesⁱⁱ. Secondement, il s'agit de caractériser le « mode d'existence » des *big data*. Au travers des propositions de cette même communauté, l'objectif est ici d'identifier à la fois les propriétés constitutives de l'objet « *big data* » et le processus par lequel cet objet « s'individualise »ⁱⁱⁱ, c'est-à-dire précisément comment il « réalise » (*rend réel*) ces propriétés.

1. 1. Perspectives sémantiques

Ainsi que l'indique Steve Lohr, journaliste du *New York Times* spécialisé dans les technologies de l'information et de la communication, le terme « *Big Data* » intègre la terminologie des ingénieurs-informaticiens de la Silicon Valley dès le milieu de la décennie 1990. Se rapportant alors exclusivement au traitement des bases des données et à leur visualisation, le terme ne concerne initialement qu'un microcosme professionnel localisé dans le Nord de la Californie (Lohr, 2013). Il faut attendre le début des années 2000, pour que le terme se popularise, s'échappant alors de son cadre technique originel et connaissant, concomitamment, une mutation de son acception. Selon J. P. Simon, consultant spécialisé dans les politiques publiques en direction des médias et télécommunication, la diffusion du terme hors des cercles d'ingénieurs s'amorce dès 2001 à la suite d'un rapport du cabinet d'audit Meta Group (devenu Gartner) sur la gestion des données dans l'entreprise (Simon, 2015). Bien que n'apparaisse à aucun moment le syntagme « *big data* » au sein de ce rapport, il aurait tout de même été à l'origine de la généralisation du terme ainsi que de sa connotation économique. En effet, au cours de la décennie 2000, l'emploi de la notion de « *big data* » s'est déplacé de son pôle technique et « ésotérique » originel (les *big data* comme ensemble de problèmes à solutionner informatiquement) vers un pôle plus économique et plus grand public (les *big data* comme ressources valorisables) au fur et à mesure où le terme investissait les cabinets de conseil et les services des entreprises, avant d'être repris par les journalistes et d'intégrer la sphère des politiques publiques. Par conséquent, à la fin de cette même décennie, le syntagme s'est installé dans la terminologie de nombre d'acteurs sociaux sans toutefois avoir l'objet d'une définition rigoureuse. Par exemple, ainsi que le constatent S. Labelle et J. B. Le Corf dans le cadre de leur étude des politiques françaises d'« *open data* »^{iv}, les notions de « *data* », de « données » et d'« informations » publiques sont généralement interchangeables dans les discours étatiques, même si les deux premières « insistent sur l'importance de l'informatique et sur la numérisation des ensembles informationnels détenus » (Labelle, Le Corf, 2012). Plus globalement, on confond généralement, dans les littératures francophones, le terme « *data* » et celui de « données » (ce qui est somme toute logique d'un point de vue étymologique). L'acception triviale de ces deux termes se rapporte ainsi à l'idée de computable, à la fonction d'indication (« faire voir ») et à l'idée de technique.

Si différence il y a, elle concerne la nature de la technique : l'usage professionnel mobilisant le terme « *data* » afin de désigner presque exclusivement les « *inputs* » et « *outputs* » de procès de productions numériques. On retrouve également cette conception associant *data* et informatique et fortement influencée par la cybernétique sous la plume de nombreux auteurs anglo-saxons^v. Dans le cadre de cet article, les deux termes seront confondus.

Par ailleurs, certains chercheurs défendent l'idée que le terme « *data* » n'est pas équivalent à celui « d'information », ni à celui de « signe » tout en présentant des traits analogues à ceux-ci. En effet, à l'instar de l'information et du signe, les *data* intègrent l'espace de la représentation, du *aliquid stat pro aliquo* (« la chose pour autre chose »), et constituent donc des abstractions et non des faits. Toutefois, la donnée n'a pas de signification « pleine », ne définit pas son propre espace de compréhension ou « d'actualisation » (par analogie avec le stimulus dans un processus énergétique) mais se charge, au contraire, de sens au cours d'un processus appelé « traitement » (Kitchin, 2014). Pour le dire simplement, les informations sont des *data* « toujours déjà » interprétées, c'est-à-dire formatées et ainsi orientées vers un horizon de connaissance. Les *data*, en revanche, sont des informations « en puissance », « virtuelles » qui vont s'actualiser sous l'effet d'un traitement, seul à même de leur conférer du sens, une « valeur de pertinence » (Bougnoux, 1995). D'autre part, on ne saurait considérer non plus les *data* comme des « ressources naturelles », c'est-à-dire des objets qui existeraient en dehors de l'activité humaine. Elles sont foncièrement tributaires des modalités de leur production. Pour cette raison, elles n'existent tout simplement pas sous une forme totalement « brute » mais, à l'inverse, apparaissent toujours comme partiellement *pré-traitées, in-formées* par les conditions de leur production (Gitelman (ed.), 2013). Ces conditions les rapportent ainsi à un ou plusieurs cadres interprétatifs qui leur préexiste nécessairement (bien que leur mobilisation reste virtuelle ainsi qu'indiqué ci-avant). Elle se rapproche en cela du signe qui, d'un point de vue logique, n'existe pas sans interprétant (Peirce, 1978). Elle s'en distingue toutefois en raison de la nature même de l'interprétant qui, en l'occurrence, est indissociable d'une médiation technique. Cette médiation technique implique systématiquement un double processus de décodage : décodage technique (traitement des *bits* informatiques) précédant nécessairement le décodage herméneutique (traduction du langage binaire en sens intelligible).

1. 2. Du mode d'existence des « *big data* »

La démarche qui consiste à qualifier un objet par l'étude de son « mode d'existence » se rapporte moins ici à l'œuvre d'E. Souriau (2009) ou à celle de B. Latour (2012) qu'à celle de G. Simondon (2001) qui envisage l'ontologie comme l'étude de l'indissociable processus d'affirmation de l'être (qu'il nomme « individuation ») et de production d'un « milieu associé » spécifique. C'est précisément ce double processus qui définit un « mode d'existence ». S'il n'est point question ici de perspectives métaphysiques, la référence à la conceptualisation de G. Simondon vise à souligner l'attention portée, dans les paragraphes suivants, aux formes particulières de prédicats proposées par un ensemble de chercheurs qui ont procédé à une analyse conjointe de l'objet « *big data* » et des conditions de leur production et usages.

Trois attributs sont généralement mis en avant pour qualifier ce mode d'existence :

- (1) Les *data* sont numériques. Comme indiqué précédemment, elles sont généralement associées à l'idée de calcul, de computationnel, mais également – et surtout – elles sont définies par distinction à l'analogique. Elles procèdent ainsi par substitution de valeurs discrètes aux valeurs continues (ce qui les différencie également de l'information), sont tributaires d'un code technique spécifique (le système binaire) et d'un paradigme technologique (« l'informatique connectée ») (Doueihy, 2015).

- (2) Elles sont « agrégatives » (Gitelman, Jackson, 2013). Elles intègrent ainsi un processus d'accumulation (ce qui explique que la catégorie du pluriel soit systématiquement employée), sont donc commensurables, comparables, et intègrent un même réseau relationnel. En outre, cette agrégation est indispensable pour leur traitement, tant d'un point de vue technique (phase première du traitement) que d'un point de vue herméneutique (phase seconde) : en d'autres termes, des *data* isolées ou parcellaires sont inutilisables et n'ont aucun sens ni valeur ; ce qui leur confère ces propriétés est précisément leur agrégation.
- (3) Elles sont opératoires. Cela veut dire, premièrement, que les *data* ont une visée fonctionnelle, qu'elles répondent à un besoin (y compris latent) pour celui qui les produit ou pour celui qui les acquiert. Secondement, le caractère opératoire des *data* est lié au fait qu'elles sont potentiellement productrices d'actions, actions qui sont de plus en plus automatisées *via* des dispositifs algorithmiques (voir *infra*).

Toutefois, c'est un quatrième attribut qui apparaît comme prééminent et qui concourt de façon décisive à leur singularisation en tant qu'objet d'étude : les « *big data* » sont « massives ». La tautologie est ici de mise pour souligner une caractéristique essentielle du phénomène, caractéristique que l'on peut rapporter au changement d'échelle de « l'informationnalisation » de la société. Nous empruntons ce concept à B. Miège pour désigner la circulation croissante et accélérée des flux d'informations ainsi que leur contribution accrue à l'ensemble des dimensions de la vie sociale (Miège, 2004). La dynamique est déjà ancienne et fut initialement abordée par le prisme exclusif du rapport de l'information à la production : elle apparaît, par exemple, comme à la fois la cause et la conséquence de l'émergence de la « société de l'information » comme projet industriel et projet de société à partir des années 1970-1980 (Mattelart, 2001). Cependant, on assiste, depuis 2001, à une explosion sans précédent du volume de données médiatisées, en raison d'innovations techniques (notamment la croissance et le perfectionnement des infrastructures-réseaux et, concomitamment, le développement exponentiel des objets connectés) et de l'émergence ou de la mutation de pratiques sociales et d'usages liés à Internet, aux services qui y sont proposés et aux terminaux qui permettent d'y accéder. En 2014, ce sont ainsi 5 Zettaoctets (5×10^{21} octets) de données qui ont été produites et qui peuvent être exploitées par les acteurs sociaux et le rythme de croissance annuelle moyen de ces *data* est estimé à 50 % sur les prochaines années (Aiden, Michel, 2014). Illustrant le phénomène, le néologisme « *data-ification* » s'impose progressivement dans les entreprises afin de désigner l'essor de la production et du traitement des *data* au niveau des services de marketing et des ressources humaines notamment (Mayer-Schönberger, Cukier, 2014 ; Delort, 2015). Selon les spécialistes des systèmes d'information, le changement d'échelle est tel qu'il entraîne un changement de nature du rapport entre les individus et les composantes informationnelles du monde dans lequel ils évoluent (*Ibidem*). Dès lors, la notion de « données massives » (*big data*) prend tout son sens : elle se rapporte à une approche – dans un cadre professionnel ou scientifique – qui n'implique pas seulement d'utiliser des *data*, mais qui implique d'utiliser de très grands volumes de *data*, voire toutes les *data* disponibles sur un objet. Le traitement s'applique ainsi idéalement à des séries de données où $N = \text{Tous}$, parce que c'est précisément l'effet de volume qui permet d'identifier des « signaux faibles », des corrélations inédites et pertinentes. Ce qui est visé est donc un horizon d'exhaustivité, une analyse de la totalité qui se substituerait aux méthodes inéluctablement imparfaites d'échantillonnage et d'extrapolation statistique telles qu'on les trouve à l'heure actuelle dans les différentes formes de « mises en données du monde » (Mayer-Schönberger, Cukier, 2013) et de « gouvernement par les nombres » (Desrosières, 2008). Ainsi, « l'ère des *big data* » consacrerait-elle l'avènement d'une nouvelle heuristique, fondée sur l'industrialisation de l'extraction de données empiriques et de leur traitement, à même de se subroger à la production de théories et aux démarches inductives/déductives

caractérisant la méthode scientifique contemporaine (ainsi que l'affirma C. Anderson dans une tribune provocante de 2008). Bien que de telles assertions apparaissent incontestablement empreintes de futurologie et excessives, elles attirent néanmoins notre attention sur le fait que, pour les tenants des *big data*, la quantité de données disponibles est plus importante que la capacité à les traiter finement, l'obtention de *data* plus décisive que la qualité de l'algorithme de traitement (Mayer-Schönberger, Cukier, 2014). Pour cette raison, le caractère « massif » des volumes constitue donc une composante essentielle du mode d'existence des « *big data* ».

2. Usages socio-économiques des *big data* : procès de production et procès de valorisation

Dans cette deuxième partie, nous allons poursuivre notre investigation sur les propriétés des *big data* en nous intéressant à *ce que l'on fait avec les big data*. Il ne s'agit évidemment pas de dresser ici une liste qui se voudrait exhaustive des utilisations potentielles de ces données par les acteurs mais, plus simplement, de caractériser les principales formes par lesquelles elles intègrent des procès de production et de valorisation. L'approche se veut donc ici principalement socio-économique : nous laisserons de côté les utilisations et dimensions plus politiques^{vi} afin de nous concentrer sur le rapport des acteurs économiques (présentés comme les premiers concernés dans les discours des cabinets d'audit et de conseils) aux *big data*.

2. 1. Le procès de production

Dans l'entreprise, l'emploi du terme *big data* est principalement métonymique et se rapporte généralement au phénomène de *data-ification* (exposé ci-avant), où l'on désigne sous un même terme la production de données et leur exploitation (notamment économique). Le terme est également généralement associé à celui d'algorithmes : on retrouve très fréquemment ce couple dans les discours des professionnels comme des observateurs, les *data* faisant figure de ressources, les algorithmes de procédés permettant un traitement de cette ressource (Cardon, 2015). Ce couple est au cœur d'ensembles socio-techniques plus vastes et complexes, généralement appelés « systèmes d'informations » ou « infrastructures *big data* » et composés de réseaux et de connexions, de serveurs et de *data-centers*, de logiciels de traitement, etc., bref, une architecture qui constitue le « milieu », au sens de G. Simondon (2001), des *big data*. Ces ensembles ont pour fonction l'extraction et l'exploitation d'importantes quantités de données, l'identification des corrélations pertinentes et la mise en œuvre automatique d'actions en vue d'un objectif. Les *outputs* de ces ensembles ne sont donc pas seulement des données, mais généralement des données et des actions^{vii}. L'exploitation de la ressource (au sens technique du terme que les anglophones nomment « *data mining* ») définit un procès de production qui comprend généralement 5 phases (et autant de marchés correspondants) : l'extraction, la transmission, le stockage, le traitement et l'interprétation (qui peut prendre, donc, la forme d'actions automatisées). A l'aval de ce procès, les dispositifs de traitement et d'interprétation algorithmiques sont définis pour répondre à une visée stratégique. Celle-ci se rapporte à un ou plusieurs besoin(s) identifié(s) au sein de l'entreprise. En fonction de ces objectifs, on paramètre un ensemble d'éléments, ce qui configure à la fois la computation et l'action subséquente du dispositif. Parmi ces éléments, on distingue notamment :

- L'objet central du calcul : il s'agit de savoir si le dispositif analyse ce que les individus font de/avec des objets (soit, une analyse des procédures d'attribution des qualités) ou si le dispositif analyse directement les propriétés de ces mêmes objets (analyse de qualités intrinsèques).
- Le type d'analyse impliquée : ce critère dérive du critère précédent et distingue les dispositifs en fonction du type d'analyse des « traces », soient une analyse des actions (ce que les individus ou les objets « font ») ou une analyse sémantique (ce que les individus, voire les objets, « disent »).

- La forme des résultats (« *outputs* ») et le rapport à la normativité : il s'agit d'évaluer si les résultats contribuent à un classement hiérarchisé (procédure de notation) ou à un ordonnancement (distinction des éléments sans hiérarchisation). De plus, il s'agit d'identifier si le dispositif contribue à la construction (*ex ante* ou *ex post*) d'une normalité (sous forme de profils statistiques par exemple) ou s'il se contente de produire une « objectivité a-normative » (Rouvroy, Berns, 2013).
- La nature de la correspondance établie : cette correspondance peut être soit dyadique (correspondance établie entre des individus et des objets ou d'autres individus) ou triadique (analyse du rapport entre individus et objets réalisée *via* l'analyse des rapports d'autres individus et objets : principe du filtrage collaboratif).
- Les modes de pondération : appréciation de la qualité des données qui s'effectue généralement au regard des trois indicateurs que sont « l'autorité », la « notoriété » et la « popularité » des sources.
- La puissance d'extrapolation : ce critère permet d'opérer une distinction entre algorithmes (plutôt) analytiques et algorithmes (plutôt) prédictifs.
- Le niveau de l'apprentissage automatique et le niveau de la « curation » humaine : le premier critère concerne « l'implémentation » (codage et mise en œuvre) de méthodes automatisables d'évolution du dispositif en fonction des variations de son environnement (données récoltées) et/ou des résultats produits. Le second vise à apprécier la quantité de travail de maintenance, d'ingénierie mais également de réécriture requis par le dispositif.
- Le degré d'intervention machinique dans la prise de décision et/ou le produit final.

Il s'agit de bien préciser ici que, pour la plupart de ces items, le paramétrage se réalise suivant une logique de gradation plutôt que d'exclusion. Ainsi, la plupart des dispositifs traitent-ils les données suivant des valeurs qui peuvent être placées sur un continuum entre deux pôles plutôt que selon des choix binaires. Par exemple, la puissance d'extrapolation ou le degré d'intervention machinique s'expriment en termes d'intensité et non en termes d'alternative.

D'autre part, on constate qu'il y a une forte tendance à l'hybridation des dispositifs. Par exemple, les concepteurs cherchent de plus en plus à conjuguer des approches actionnelles (analyse de « ce qui est fait ») et des approches sémantiques (analyse de « ce qui est dit ») dans le traitement des *data* (Bullich, 2015).

Enfin, la performance de ces dispositifs, et plus globalement des ensembles socio-techniques au sein desquels ils s'intègrent, s'apprécie, d'un point de vue de l'ingénierie informatique, au regard de 5 critères (les 5V en français – Delort, 2015) :

- Le volume : désigne la quantité d'informations obtenues et traitées par ces ensembles
- La vitesse : se rapporte aux délais d'actualisation et d'analyse ; les dispositifs les plus performants tendent vers l'obtention et le traitement de données en temps réel.
- La variété : désigne la capacité de l'ensemble à traiter et agréger des *data* de formes distinctes (les différentes traces numériques, des contenus sémantiques, des images, etc.).
- La véracité : désigne la capacité du système à évaluer la qualité des données (ce critère est en partie lié aux modes de pondération exposés ci-avant).
- La valeur : se rapporte à la pertinence des *data* obtenues et de leur traitement en fonction de l'objectif assigné au système.

Avant de clore cette partie sur la production de données et de nous intéresser, dans la partie suivante, à leur destination économique, nous souhaiterions attirer l'attention du lecteur sur le fait que cette production sera, dans un futur proche, principalement le fait de capteurs et d'opérateurs automatiques et requerra de moins en moins d'interventions humaines. En effet,

des applications, *via* les puces RFID notamment, sont à même d'intégrer l'ensemble des phases de production industrielle^{viii} et de générer en continu des *data* censées optimiser l'efficacité de chaque activité, « en remplaçant l'observation et la saisie d'informations par l'intégration même des objets dans le réseau » (Benghozi *et alii*, 2009, p. 26). De très nombreuses filières industrielles sont ainsi susceptibles d'être reconfigurées par de nouveaux services liés aux objets connectés et la production de *data* sera très probablement à l'avenir foncièrement liée à cet « Internet des objets »^{ix}. Cette dénomination manifeste à la fois le fait que l'infrastructure-réseau sera prochainement utilisée de façon prépondérante par des « objets » (physiques ou virtuels) et le fait que ces objets seront à même de communiquer entre eux, dans une autonomie croissante, afin de servir des applications automatisées et continues. Annoncé comme une « une dimension fondamentale de l'Internet de demain » (*Ibidem*, p. 15), l'Internet des objets constitue donc un enjeu majeur pour les industries. Si le champ des possibles apparaît particulièrement vaste (au travers d'applications conduisant notamment à une automatisation des activités de supervision de la production, de contrôle-qualité, de gestion des ressources humaines, de maintenance technique, etc. Xu *et alii*, 2014), les services envisagés ne pourront s'actualiser qu'à la condition d'une adaptation constante et d'une élévation des performances des systèmes d'informations car ce qui est d'ors et déjà sûr, c'est que le phénomène favorise une croissance exponentielle des *data* produites.

2.2. Une double rationalité économique

La rationalité économique du procès de production des *data* est duale. Les *data* sont, en effet, envisagées comme (1) une fin en soi (c'est-à-dire une ressource ayant une valeur propre échangeable sur un marché) ou (2) comme un moyen (c'est-à-dire un facteur permettant la production et la valorisation d'autres biens et services)

- (1) Les *data* comme bien valorisable. Il existe différents marchés de *data* mais le marché principal concerne les données personnelles. Les acteurs centraux de ces marchés sont les courtiers en données (« *data-brokers* »)^x, bien que les acteurs les plus à même de produire et valoriser ces ressources soient, pour le moment, les GAFAX^{xi}.

D'un point de vue comptable, les données sont des actifs incorporels dont la valeur fluctue dans le temps. Cette impermanence génère de grandes difficultés dans les évaluations des *data* qui font l'objet de transactions, difficultés qui expliquent que les contrats de cession ou d'exploitation soient particulièrement sophistiqués. On distingue ainsi généralement deux valeurs pour ces biens. La première est une valeur « immédiate » qui dépend de leur caractère opératoire en fonction d'un objectif qui leur préexiste. La seconde est une valeur latente, une valeur qui ne se révélera qu'avec des utilisations ultérieures, avec de nouveaux objectifs ou alors une fois que les données seront recombinaisonnées (Mayer-Schönberger, Cukier, 2014). Dans les contrats de cession, on retrouve donc un tarif indiqué pour une utilisation première (sous la forme d'une licence d'exploitation) mais également des valeurs d'options (en cas d'utilisations ultérieures).

Les données les plus « liquides » (c'est-à-dire se convertissant le plus facilement en numéraire) sont les profils de consommateurs. Ces profils sont produits – par agrégation, recoupement et corrélation, voire prédiction dans certains cas – à partir de données en provenance de quatre sources principales : les traces Internet (à partir d'un relevé des actions des internautes ou, plus rarement, d'une analyse sémantique des contenus produits par les internautes), les informations relatives aux cartes de fidélités, les données vendues par les administrations (dans certains pays), les données en provenance d'objets connectés (*via* les dispositifs de géolocalisation par exemple). En fonction de la qualité de ces données et de la qualité de leur compilation, ces profils sont échangeables entre 10 centimes et 50 euros (prix pour 1000 profils

pratiqués en Europe occidentale en 2014). Le cabinet Transparency Market Research estime ainsi que le marché global des *big data* avoisine les 8,8 milliards d'euros en 2014 et qu'il va connaître une croissance de 40% en moyenne sur les 3 prochaines années^{xiii}. L'intérêt que porterait l'ensemble du monde économique à ces données pourrait avoir comme incidence de modifier durablement les modèles économiques de nombre d'entreprises. Par exemple, le cabinet EY (anciennement Ernst & Young) préconise que les entreprises culturelles (industries culturelles et médiatiques et entreprises du spectacle vivant) mettent en place des systèmes informationnels « orientés clients » au sein de leur appareil productif afin de valoriser non plus uniquement les produits ou prestations qu'elles proposent mais également les données personnelles acquises auprès de leurs publics (EY, 2013). On peut souligner à ce sujet que ce type d'approche est déjà largement emprunté par les producteurs d'applications pour la téléphonie mobile et les producteurs de jeux-vidéos sur navigateur Web qui, ont depuis des années, adopté un modèle économique fondé sur la gratuité pour le consommateur et la valorisation des données-clients^{xiii}.

- (2) Les *big data* comme facteur de production. La plupart des discours présentant les opportunités économiques liées aux *big data* insiste sur ce second point. La capacité à extraire et traiter de très grands volumes de données y est, en effet, dépeinte comme un levier de développement pour l'ensemble des secteurs économiques. Le « *Data-Driven Marketing* » contribuerait ainsi à l'économie américaine à hauteur de 202 milliards de dollars en 2015^{xiv} et la « *Data-Driven Economy* » apparaît comme le nouveau projet technico-économique global, après et en complément de « l'économie créative »^{xv}. Cette (énième) « révolution » informationnelle de l'économie a, dans un premier temps, principalement été annoncée par les plus grands cabinets d'audit et de conseils mondiaux ainsi que par nombre de services de prospective. Ces acteurs se sont en effet très tôt positionnés comme les hérauts des mutations à venir^{xvi} et ont, ce faisant, tout particulièrement stimulé les activités consistant à proposer des solutions organisationnelles et techniques pour les entreprises désireuses de profiter de la manne nouvelle censément liée aux *big data*. Ainsi, Gartner, EY ou encore Deloitte^{xvii} vont-ils proposer un ensemble de rapports et d'études promouvant notamment la production et la mise en œuvre de « *Decision Support Systems* », ceux-ci étant présentés comme des outils décisifs quant à la conduite de l'entreprise dans le futur.

De manière générale, on peut distinguer deux principales fonctions ou visées stratégiques associées à ces dispositifs et se rapportant au procès de valorisation de la production^{xviii}. La première est relative à une meilleure connaissance de la demande afin d'orienter ses choix. Il s'agit principalement d'extraire des données et de constituer des profils-clients afin de favoriser leurs décisions d'achat sur des produits préexistants. La visée du dispositif est donc ici l'optimisation de l'appariement. Le système est principalement un agent du « *matching* » : il recherche l'adéquation d'une offre et d'une demande en procédant par ordonnancement des éléments informationnels dont il dispose et automatise les propositions commerciales sous formes de recommandations/prescriptions personnalisées. La seconde catégorie de visées stratégiques se rapporte à la conception même des produits. Il s'agit, dans ce cas, de réduire l'incertitude quant à la valorisation des produits en formatant ceux-ci en fonction des données obtenues des consommateurs. Le dispositif intervient dès lors non plus à l'aval, comme précédemment, mais à l'amont de la chaîne de production et constitue un outil d'aide à la décision quant aux choix relatifs au produit même. Il agit pour ce faire, comme une instance de qualification (il identifie et rapporte, par exemple, les propriétés du produit à des catégories d'attentes supposées des consommateurs) qui s'accompagne généralement de procédures de rationalisation

opératoire et d'automatisation (principe de la délégation technique des choix de production).

Au final, on retrouve la bipartition classique en *marketing* entre les stratégies « *push* » et les stratégies « *pull* » : dans le premier cas, il s'agit de réduire les risques concernant la valorisation de la production en utilisant l'ensemble *Big Data*/dispositif de traitement algorithmique comme un outil permettant de « pousser » le produit vers la demande, la computation se portant alors sur l'appariement qui est suscité *via* le système de recommandation/prescription ; dans le second cas, l'incertitude est amoindrie en attirant la demande sur des produits spécifiquement conçus en fonction des besoins tels que formulés par le dispositif.

S'il est indéniable que les *big data* sont effectivement en train d'investir le monde économique, que leur valeur stratégique comme leur valeur de marché tend à s'accroître ainsi que nous l'avons succinctement présenté ici, toujours est-il que l'avènement de « l'ère des *big data* » (EY, 2013) reste pour le moment largement prématuré. Bien que le syntagme soit, sinon omniprésent, très fréquemment mobilisé dans les milieux journalistiques et professionnels depuis quelques années, il ne reflète aucunement le volume des pratiques associées à ce terme, pratiques qui, d'une part, ne concernent encore que peu d'acteurs et, d'autre part, soulèvent nombre d'interrogations qui sont susceptibles de constituer des freins à leur développement. Pour ces raisons, les *big data* font l'objet de discours qui présentent, à l'heure actuelle, un caractère souvent hypothétique, virtuel, et qui, en cela, intègrent une « économie de la promesse » comme nous allons le voir dans la partie suivante.

3. Des discours s'inscrivant dans une « économie de la promesse »

Bien que les applications technico-économiques liées aux *big data* se multiplient, elles restent encore largement exceptionnelles dans le monde de l'entreprise et ne concernent, pour le moment, que certaines firmes qui, précisément, développent leur activité autour de l'exploitation des données. En effet, si la plupart des entreprises ont, depuis des décennies, intégré des systèmes d'informations plus ou moins sophistiqués à leur chaînes de production, l'approche en termes de *big data*, c'est-à-dire l'approche consistant en l'analyse d'un très grand volume de données, ne concerne qu'un nombre limité d'acteurs. Plusieurs obstacles peuvent expliquer cette situation : les coûts des infrastructures techniques et les besoins en moyens humains sont élevés, on trouve généralement un manque de compétences de la part des personnels en postes, il peut exister des difficultés de réorganisation de structures quelques fois sclérosée du fait de leur importance, l'encadrement légal de l'exploitation et de la commercialisation des données reste incertain, etc. On observe donc un net décalage entre les discours et les pratiques, et si les premiers confèrent une grande visibilité au phénomène *big data*, les secondes sont, en revanche, nettement moins répandues.

Dans le cadre de cette troisième et dernière partie, nous allons, par conséquent, envisager le phénomène *big data* comme un projet qui s'inscrit encore très largement dans une « économie de la promesse ». Nous empruntons cette expression à la sociologie des sciences et techniques afin de désigner des modalités de « construction collective du futur », modalités produites à partir de discours et de représentations à même de déterminer les pratiques (Joly, 2010). Ces « promesses » ont donc des effets performatifs et sont susceptibles de contribuer significativement à la « construction sociale de la réalité » (Berger, Luckmann, 1966), en orientant durablement les décisions des acteurs. Parce qu'elle constitue un pari sur l'avenir, parce qu'elle définit immanquablement un horizon d'attente ou « d'espoir » et qu'elle cristallise ainsi une, ou plutôt, un ensemble de conceptions du futur, la technique est particulièrement encline à faire l'objet d'une « économie de la promesse » (Leibing, Tournay, 2010). Scientifiques, ingénieurs comme financeurs ou politiques ont ainsi appris, de longue date, à user d'une rhétorique présentant les « promesses technologiques » dans une « une

logique de l'enrôlement » car, ainsi que le souligne F. Chateauraynaud, « derrière les promesses viennent s'aligner des acteurs, des institutions, des capitaux » (Chateauraynaud, 2011, p. 363). L'enjeu est donc de produire des discours à même de conférer à leurs promoteurs une position dans l'espace public^{xix} et social favorisant « la mobilisation, la circulation et l'accumulation de ressources » (Joly, 2010, p. 205). C'est précisément à ces discours de promotion des *big data* que nous allons nous intéresser dans les paragraphes suivants. A partir d'une étude des principaux rapports à ce sujet proposés par les grands cabinets d'audit et des conseils EY, Deloitte, Gartner et IDC, nous avons ainsi cherché à identifier les principaux ressorts et la nature de la promesse. Comme nous allons l'exposer, celle-ci repose sur (1) l'analogie avec des ressources naturelles et le caractère intarissables de ces ressources (2) la prédictibilité comme solution économique « ultime ».

3. 1. Les *big data* comme les futures « ressources essentielles »

Ainsi que le rappelle Y. Citton dans sa réflexion sur le traitement des données, « futurologues, prospectivistes et autres prophètes des technologies émergentes font de leur mieux pour entrevoir des pertinences aujourd'hui virtuelles, dont ils annoncent avec plus ou moins de succès l'actualisation prochaine » (Citton, 2012, p. 145). Il en va évidemment ainsi avec les *big data*. La propagation impressionnante du terme au cours des dernières années est notamment le fait de campagnes « d'évangélisation »^{xx} menées hors des cercles de spécialistes en informatique par des cabinets d'audit et de conseils. Il n'apparaît pas surprenant que les promoteurs les plus ardents du phénomène soient ces cabinets, puisqu'ils entretiennent un rapport de nécessité avec l'innovation et vivent, en partie, de la précocité de leurs expertises. En effet, une part importante de la valeur des services qu'ils proposent est liée à leur aptitude à s'imposer dans les milieux professionnels comme référents de domaines nouveaux ; une de leurs stratégies consiste ainsi à s'emparer des sujets technico-économiques avant les autres et de proposer des analyses censément exclusives et destinées à pallier la méconnaissance des acteurs économiques. L'enjeu est donc de produire très tôt des horizons d'attentes au sein des milieux décisionnaires à partir de discours plus ou moins fictionnels qui présentent généralement comme inéluctable l'amplification de phénomènes en émergence (sur ce point, nous rejoignons pleinement Y. Citton *supra*). L'expertise se mue alors en *leitmotiv* voire en une injonction subtile et les arguments sont repris, mobilisés, pour légitimer des décisions économiques, des projets industriels et des politiques publiques. En cela, ces discours produits ont incontestablement une performativité sociale élevée voire présentent, quelques fois, un caractère qui s'assimile aux prophéties auto-réalisatrices (Mackenzie *et alii*, 2007). Les rapports que nous avons étudiés^{xxi} présentent les *big data* comme des « ressources essentielles », c'est-à-dire des ressources à même de conférer à celui qui en dispose une position durablement dominante sur les marchés. Cette notion, issue droit de la concurrence, est mobilisée ici pour souligner le caractère présenté comme indispensable, dans un avenir proche, des données au fonctionnement des marchés. Toutefois, ce qui apparaît le plus frappant à la lecture de ces rapports est l'analogie systématiquement faite entre les *big data* et les ressources naturelles. Cette analogie est immédiatement perceptible au travers des univers sémantiques mobilisés : les *big data* sont ainsi généralement assimilées à un minerai ou une huile minérale. Dans les deux cas, l'analogie suggère, d'une part, l'importance de la valeur sous-jacente aux données, qui comme le minerai aurifère ou le pétrole, s'échangeraient facilement sur les marchés et généralement à des cours élevés. D'autre part, les *big data* constituerait une « source d'énergie » pour l'économie. A l'instar du charbon au 19^{ème} siècle ou du pétrole au 20^{ème}, les données seraient l'élément de base du « carburant de l'économie du 21^{ème} siècle ». Le registre lexical est donc foncièrement emprunté à l'industrie minière : « *data-mining* », « extraction », « gisements », « forage », « raffinement », etc. Les data sont en outre présentées comme une ressource facilement accessible, aisément exploitable et très

rentable : le « nouvel or noir » ou « la pépite à transformer en lingots ». Pour cette raison, les *data* constituent un « accélérateur de croissance » à double titre : parce qu'à la base de nouveaux marchés que l'on annonce prolifiques et parce qu'« alimentant » les différents services de l'entreprises et augmentant ainsi sa productivité et son efficacité commerciale.

Outre son caractère présenté comme indispensable au fonctionnement des marchés à l'avenir, la particularité notable de cette ressource est le fait d'être, plus qu'abondante, intarissable. En effet, les discours de ces cabinets mettent en avant, à juste titre, la multiplication exponentielle des sources de *data* : l'essor prodigieux des communications électroniques, l'intensification des pratiques et des usages des terminaux en tout lieu et en tout temps, le développement des infrastructures-réseaux et des objets connectés, tout concoure à ce que la production de *data* soit inépuisable. Afin d'appuyer leurs démonstrations, les cabinets convoquent ce que nous nommons la « rhétorique des grands nombre ». Ainsi que le rappelle O. Ertzscheid, le recours aux « grands nombres » est un élément constitutif de la « mythologie d'Internet »^{xxii}, et les cabinets jouent habilement sur le fait que le web constitue un « formidable écosystème facilitateur et multiplicateur de la moindre interaction, de la moindre navigation, de la moindre publication, de la moindre attention portée » (Ertzscheid, 2011). Ainsi, sont présentés des chiffres tout bonnement astronomiques censés illustrer les potentialités liées au *big data* : 250 milliards de courriels seraient envoyés chaque jour, près de 200000 sms seraient envoyés chaque seconde, il y aurait, chaque année, plus de 60 milliards de « *like* » sur Facebook, 400 millions de tweets postés, 4,5 milliards de recherche Google, etc. Les chiffres ont de quoi impressionner. Cependant, les rapports exposent ces estimations et projections de façon unilatérale : l'abondance y est toujours présentée comme un atout, une chance pour les acteurs économiques et jamais comme un problème à traiter. Or, plus on augmente le volume de *data*, plus les dispositifs de traitement doivent être performants, et donc plus les coûts associés (personnels compétents, infrastructures et matériels appropriés) sont élevés. Par conséquent, les acteurs économiques risquent d'être rapidement confrontés à des problèmes de rendements d'échelle susceptibles d'impliquer des investissements lourds que seuls les plus puissants financièrement pourront durablement supporter.

3. 2. La prédictibilité comme solution économique « ultime »

Le second ressort de la promesse des *big data* avancée par les cabinets de conseils et d'expertise se rapporte à la prédictibilité comme solution « ultime » face aux crises et à l'incertitude économique. Cette perspective, foncièrement empreinte d'un certain type de déterminisme technologique qu'E. Morozov (2013) qualifie de « solutionnisme technologique », répond à l'inquiétude ou à la morosité des acteurs économiques en faisant le pari de l'anticipation, de la connaissance probabiliste du futur. Tels des oracles des temps modernes, les ensembles *big data*/dispositifs algorithmiques satisferont aux questions des décideurs économiques (mais également politiques, voir Rouvroy, Berns, 2013) et dessineront ainsi un avenir où le risque se réduit au fur et à mesure que les capacités de computation augmentent. Ainsi, le « tournant prédictif », suivant l'expression de D. Cardon (2013), qui se profile sous l'effet de la mise en œuvre de ces ensembles sera-t-il à même, selon ces discours, de spectaculairement relancer l'économie. Les prévisions et annonces des cabinets sont, à ce propos, stupéfiantes. Le problème est qu'elles sont souvent divergentes : par exemple le marché des *big data* en 2015 était estimé à 16,9 milliards de dollars pour IDC (prévision de janvier 2014), alors que Transparency Market Research nous annonce à la même époque un chiffre deux fois moins important (8,9 milliards de dollars)^{xxiii} ; de même les prévisions pour 2018 font état d'un marché à hauteur de 48,3 milliards de dollars (pour Transparency Market Research en avril 2015) alors qu'IDC estime à 125 milliards de dollars ce même marché (dans une prévision de décembre 2014)^{xxiv}. Les promesses d'emplois qui accompagnent le développement de « l'écosystème des *big data* » ont, quant à elles, de quoi appâter les

responsables politiques : les cabinets Gartner et EY annonçaient 4,4 millions d'emplois relatifs aux *big data* créés dans le monde entre 2013 et 2015 mais, pour le moment, aucune donnée publique ne confirme ce chiffre^{xxv}. Au-delà du caractère fantaisiste de certaines des estimations présentées ici (comment expliquer autrement de tels écarts ?), elles manifestent une foi dans la technique qui peut être contestée ou, tout du moins, nuancée.

Deux aspects contribuent, en effet, à la dimension outrancière de la promesse. Le premier se rapporte aux mécanismes de conversion des traces Internet, des zettaoctets de données en espèces sonnantes et trébuchantes et en emplois. Cette conversion apparaît comme une « boîte noire » et semble aller de soi dans les rapports étudiés^{xxvi}. Or, des arguments peuvent être avancés pour, précisément, souligner qu'il en va tout autrement : par exemple, et comme nous l'avons indiqué précédemment, le coût d'exploitation des *data* risque d'augmenter au fur et à mesure que le volume disponible augmente (en raison de l'entropie produite par cette croissance fulgurante de données). Ainsi, la conversion *data*/argent pourrait-elle être plus en plus coûteuse, et le taux de conversion de plus en plus bas. Dès lors, la manne *big data* ne concernera que les acteurs les plus aptes techniquement et financièrement à la capter. De même, la plupart des dispositifs de traitement sont conçus pour une automatisation croissante des tâches ; rien n'indique donc que les niveaux d'emplois vont connaître une croissance globale par leur truchement : si un certain nombre d'emplois seront indubitablement créés dans le cadre du développement des stratégies de *big data*, on ne peut dire s'ils compenseront les emplois qui vont disparaître en raison de cette automatisation. A ce sujet, le rapport *The Future of Job* présenté dans le cadre du Forum économique mondial de 2016 prévoit la disparition de 7,1 millions d'emplois dans les 5 prochaines années du fait de l'automatisation, disparition que ne compenseront pas les 2,1 millions d'emplois prévus dans le domaine de l'informatique et des nouvelles technologies (Forum économique mondial, 2016). Le second aspect qui nous apparaît ici critiquable se rapporte à l'enthousiasme excessif dont font montre ces rapports au vu des développements technologiques récents. La portée prédictive du traitement des *big data* est ainsi présentée comme un fait établi alors que les discours des spécialistes d'informatique indiquent, beaucoup plus prudemment, que l'on reste sur des modèles probabilistes et largement imparfaits^{xxvii}. Une des difficultés majeures auxquels sont confrontés les concepteurs de ces systèmes d'informations est liée au fait que les dispositifs ne savent pas « ne pas savoir » et se taire. Ils produisent systématiquement des *outputs* et procèdent, pour ce faire, par réduction de propriétés incommensurables au computationnel. Par conséquent, ces dispositifs produisent des qualités qu'ils attribuent à des objets plus qu'ils ne détectent des propriétés intrinsèques aux objets. En cela, on peut considérer qu'ils « produisent du réel » tout autant qu'ils l'analysent, et c'est précisément cette dimension performative qui, pour le moment, nuit à leur portée prédictive. En effet, ainsi que l'exprime H. Guégen, la fonction prédictive procède en « envisageant l'actuel à partir du possible et réduisant ce dernier à la mesure mathématique du prédictible » (Guégen, 2014, p. 279). Ce faisant, les dispositifs algorithmiques effectuent une amputation du réel, une « forclusion du possible » (*ibidem*) qui manifeste, *ipso facto*, les limites de leur capacité d'anticipation. Par conséquent, la prédictibilité reste, à l'heure actuelle, une promesse que la technique n'est que très partiellement en mesure de tenir ; elle apparaît comme un horizon vers lequel l'ensemble des dispositifs considérés se dirigent à grande vitesse, certes, mais un horizon qui, par définition, est inatteignable. En présentant la capacité à prévoir, et notamment la capacité à prévoir le comportement humain, comme une propriété inhérente aux systèmes couplant *big data* et dispositifs de traitement algorithmique, les discours étudiés apparaissent comme autant de symptômes d'une pensée foncièrement technocentrée, qui réduit un vaste champs de problématiques psychiques, sociales et politiques à des enjeux techniques.

Conclusion

Au terme de cet article, il ne s'agit pas de nier l'importance du phénomène « *big data* » mais de le comprendre comme un phénomène argumentatif et rhétorique tout autant que comme un phénomène relatif à des stratégies industrielles et des avancées techniques. Dans les paragraphes précédents, nous avons donc cherché à montrer le déplacement du centre de gravité illocutoire des discours produits par les cabinets de conseils, d'un pôle « verdictif » vers un pôle « promissif » (selon les catégories de J.L. Autin, 2002) : dans les rapports étudiés, la description et l'examen du phénomène fait, en effet, régulièrement place à la promesse des bienfaits résultant de la « conversion aux *data* ».

Au-delà de ce travail sur les « économies de la promesse », il est important, nous semble-t-il, d'inscrire le phénomène dans le temps long, c'est-à-dire d'appréhender ces nouveaux modes de recueil et d'analyse des données, de computation et d'automatisation dans le cadre de mouvements plus importants temporellement que le « présent perpétuellement en révolution » qui semble être devenu la référence. Il s'agit, pour ce faire, de prendre en considération les travaux historiques sur l'affirmation d'une « raison statistique » (Desrosières, 2010) ou d'une « raison computationnelle » (Bachimont, 2010) qui définissent une perspective plus ample et permettent d'aborder ce phénomène des *big data* comme, par exemple, le dernier avatar en date d'une certaine forme de *libido cognoscendi*, en l'occurrence orientée vers la compréhension du monde et le gouvernement des hommes par le chiffrage et le calcul. A partir de là, la notion « révolution » ou, tout du moins, la question de la rupture peut avoir du sens, précisément parce que ce changement de focale peut éviter à l'analyse, d'une part, la centration sur le très court terme, et, d'autre part, l'écueil du « futur qui colonise le présent » (Giddens *in* Joly, 2014). En effet, le caractère encore largement prophétique de nombre de discours d'accompagnement des *big data* est patent. Ces discours semblent ainsi ignorer la temporalité propre à « l'ancrage social des techniques » qui doivent systématiquement composer avec un ensemble de facteurs sociaux à même de freiner leur diffusion (Miège, 2007). La « révolution des données » n'est donc probablement pas « en marche » contrairement à ce qu'affirment les spécialistes qui promeuvent ainsi leur domaine (Mayer-Schönberger, Cukier, 2014), mais, plus modestement, elle amorce ses premiers pas. Le développement des *big data*, comme procès de « *data-ification* », est incontestable ; en revanche, le développement des *big data* comme paradigme socio-technique et comme nouvelle manne qui irriguerait l'ensemble des domaines économiques est encore au stade embryonnaire.

Enfin, et avant de clore cet article, nous souhaiterions insister sur le fait que cette manne risque d'être captée de façon, sinon exclusive, tout du moins prépondérante par les acteurs de l'économie numérique actuellement en position dominante. Contrairement au point précédent, nous rejoignons sur ce point pleinement les propos de V. Mayer-Schönberger et K. Cukier (2014) qui alertent sur la menace d'une monopolisation des données. Il y a là un enjeu économique et social de première importance qui appelle une réaction de la part du législateur, réaction impliquant de statuer en matière de droit de la concurrence, de respect de la vie privée mais également de « responsabilité technique » et de transparence des choix opérés et imposés aux citoyens par les algorithmes^{xxviii}. Sans cet encadrement juridique, ce qui a été annoncé comme un nouvel *Eldorado* pourrait n'être qu'un *Far West* où le mieux armé fera la loi.

Bibliographie

Aiden, Erez ; Michel, Jean-Baptiste (2014), *Uncharted: Big Data as a Lens on Human Culture*, New York, Riverhead Books.

Anderson, Chris (2008), « The End of Theory: The Data Deluge Makes the Scientific Method Obsolete », *Wired*, [en ligne] : <http://www.wired.com/2008/06/pb-theory/> (page consultée le 10/01/2016).

Austin, John Langshaw (2002), *Quand dire, c'est faire*, Paris, Le Seuil.

Bachimont, Bruno (2010), *Le Sens de la technique : le numérique et le calcul*, Paris, Les Belles Lettres/Encre Marine.

Benghozi, Pierre-Jean ; Bureau, Sylvain ; Massit-Folléa, Françoise (2009), *L'Internet des objets. Quels enjeux pour l'Europe*. Paris, Editions de la Maison des Sciences de l'Homme.

Berger, Peter ; Luckmann, Thomas (1966), *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, New York, Anchor.

Bougnoux, Daniel (1995), *La communication contre l'information*, Paris, Hachette.

Bouquillion, Philippe ; Miège, Bernard ; Moeglin, Pierre (2013), *L'industrialisation des biens symboliques*, Grenoble, Presses Universitaires de Grenoble.

Bullich, Vincent (2015), « Vers une créativité automatisée ? », Communication dans le cadre du Colloque international « *De l'injonction à la créativité à sa mise en œuvre : quel parallèle entre monde de l'art et monde productif ?* », Maison des Sciences de l'Homme de Nantes, 9-10 Avril 2015, Actes en cours de publication.

Cardon, Dominique (2013), « Présentation », *Réseaux*, n° 177, 1/2013, p. 9-21

Cardon, Dominique (2015), *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Paris, Le Seuil.

Carr, Nicholas (2015), *The Glass Cage: Automation and Us*, New York, Norton.

Chateauraynaud, Francis (2011), *Argumenter dans un champ de force - Essai de balistique sociologique*, Paris, Editions Petra.

Citton, Yves (2012), « Traiter les données: entre l'économie de l'attention et le mycélium de la signification », *Multitudes*, n° 49, 2012/2, p.143-149.

Commission Européenne (2015), *Stratégie pour le Marché numérique unique*, [en ligne] : http://ec.europa.eu/news/2015/05/20150506_fr.htm (page consultée le 26/01/2016).

Delort, Pierre (2015), *Le Big Data*, Paris, Presses Universitaires de France.

Desrosières, Alain (2008), *Gouverner par les nombres*, Paris, Presses des Mines ParisTech.

Desrosières, Alain (2010), *La politique des grands nombre. Histoire de la raison statistique*, Paris, La Découverte.

DMA (2016), *The Value of Data 2015: Consequences for Insight, Innovation & Efficiency in the U.S. Economy*, [en ligne] : <http://thedma.org/advocacy/data-driven-marketing-institute/value-of-data/> (page consultée le 09/03/2016).

Doueïhi, Milad (2015), *Qu'est-ce que le numérique ?*, Paris, Presses Universitaires de France.

Evans, David ; Hagi, Andrei ; Schmalensee, Richard (2006), *Invisible Engines*, Cambridge, The MIT Press.

Ertzscheid, Olivier (2011), « l'imaginaire numéraire du numérique », *Affordance.info – Le blog d'un maître de conférences en sciences de l'information. Réflexions, analyses, signalements, opinions.* [en ligne] : http://affordance.typepad.com/mon_weblog/2011/06/imaginaire-numeraire-numerique.html (page consultée le 01/02/2016)

EY (2013), *Comportements culturels et données personnelles au cœur du Big data*, Paris, Ernst & Young et Associés.

Forum économique mondial (2016), *The Future of Job*, [en ligne] : http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf (page consultée le 01/02/2016).

Gitelman, Lisa (ed.) (2013), « *Raw Data* » is an Oximoron, Cambridge, The MIT Press.

Gitelman, Lisa ; Jackson, Virginia (2013), « Introduction », in Gitelman, Lisa (ed.), « *Raw Data* » is an Oximoron, Cambridge, The MIT Press, p. 1-15.

Guéguen, Haud (2014), « La critique et le possible : le rôle de la catégorie du possible dans la critique des TNIC », in George, Eric ; Grangeon, Fabien (dir.), *Critique, sciences sociales et communication*, Paris, Mare & Martin, p. 265-290

Gün Sirer, Emin (2016), *Hacking, Distributed – Personal Blog*, [en ligne] : <http://hackingdistributed.com/> (page consultée le 01/02/2016).

Habermas, Jürgen (2003), *L'espace public*, Paris, Editions Payot.

Hartley, John (ed.)(2005), *Creative Industries*, Victoria, Blackwell Publishing.

Joly, Pierre-Benoit (2010), “On the economics of technoscientific promises”, in Akrich Madelein ; Barthe Yannick ; Muniesa, Fabian ; Mustar, Philippe, *Débordements - Mélanges offerts à Michel Callon*, Paris, Presse des Mines ParisTech, p.203-222.

Joly, Pierre-Benoit (2014), « Pour une approche critique de l'économie des promesses technoscientifiques », *Les dossiers de l'IHES*, [en ligne] : www.ihes.fr/IMG/docx/20140708-UDT-CR-JOLY_Pierre-Benoit.docx (page consultée le 10/01/2016).

Kitchin, Rob (2014), *The Data Revolution*, Londres, Sage.

Labelle, Sarah ; Le Corf, Jean-Baptiste (2012), « Modalités de diffusion et processus documentaires, conditions du « détachement » des informations publiques. Analyse des discours législatifs et des portails *open data* territoriaux », in Bardou Boisnier, Sylvie ; Paillart, Isabelle (coord.), *Information publique : stratégies de production, dispositifs de diffusion et usages sociaux*, *Les Enjeux de l'Information et de la Communication*, n° 13/2, p. 59-71, [en ligne] : <http://lesenjeux.u-grenoble3.fr/2012-dossier/Labelle-LeCorf/index.html> (page consultée le 12/12/2015).

Latour, Bruno (2012), *Enquêtes sur les modes d'existence : Une anthropologie des modernes*, Paris, La Découverte.

Leibing, Annette ; Tournay, Virginie (dir.)(2010), *Les technologies de l'espoir. La fabrique d'une histoire à accomplir*, Québec, Presses de l'Université Laval.

Lohr, Steve (2013), “The Origins of ‘Big Data’: An Etymological Detective Story”, [en ligne] : http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/?_r=0 (page consultée le 01/02/2016).

Mackenzie, Donald ; Muniesa, Fabian ; Siu, Lucia (eds) (2007), *Do Economists Make Markets ? On the Performativity of Economics*, Princeton, Princeton University Press.

Mattelart, Armand (2001), *Histoire de la société de l'information*, Paris, La Découverte.

Mayer-Schönberger, Viktor ; Cukier, Kenneth (2013), « Mise en données du monde, le déluge numérique », *Le Monde Diplomatique*, juillet 2013, [en ligne] : <http://www.monde-diplomatique.fr/2013/07/CUKIER/49318> (page consulté le 10/01/2016)

Mayer-Schönberger, Viktor ; Cukier, Kenneth (2014), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston, Eamon Dolan/Houghton Mifflin Harcourt.

Meta Group (2001), *3D Data Management: Controlling Data Volume, Velocity and Variety*, [en ligne] : <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (page consultée le 10/01/2016).

Miège, Bernard (2004), *L'information-communication, objet de connaissance*, Paris, INA Editions.

Miège, Bernard (2007), *La société conquise par la communication. Tome 3. Les TIC entre innovation technique et ancrage social*, Grenoble, Presses Universitaires de Grenoble.

Morozov, Evgeny (2013), *Save Everything, Click Here : Technology, Solutionism, and the Urge to Fix Problems that Don't Exist*, Londres, Allen Lane.

Pasquale, Franck (2015), *The Black Box Society*, Cambridge, Harvard University Press.

Peirce, Charles S. (1978), *Ecrits sur le signe*, Paris, Le Seuil.

Rouvroy, Antoinette ; Berns, Thomas (2013), « Gouvernamentalité algorithmique et perspectives d'émancipation », *Réseaux*, n° 177, 1/2013, p. 163-196

Seaver, Nick (2014), "Algorithmic Recommendations and Synaptic Functions", *Limn*, n°2, [en ligne] : <http://limn.it/algorithmic-recommendations-and-synaptic-functions/> (page consultée le 10/01/2016).

Simon, Jean-Pierre (2015), « Le big data : un enjeu pour les industries créatives », *InaGlobal - la Revue des industries créatives et des médias*, [en ligne] : <http://www.inaglobal.fr/numerique/article/le-big-data-un-enjeu-pour-les-industries-creatives-8065?qt=7> (page consultée le 01/02/2016)

Simondon, Gilbert (2001), *Du mode d'existence des objets techniques*, Paris, Aubier.

Souriau, Etienne (2009), *Les différents modes d'existence*, Paris, Presses Universitaires de France.

Union Internationale des Télécommunications (2012), *Présentation générale de l'Internet des objets*, [en ligne] : <https://www.itu.int/rec/T-REC-Y.2060-201206-I> (page consultée le 21/03/2016).

Xu, Li Da ; He, Wu ; Li, Shancang, "Internet of Things in Industries: A Survey", *IEEE Transactions on Industrial Informatics*, vol. 10, n° 4, p. 2233-2243.

ⁱ Dans son ouvrage *Save Everything, Click Here*, E. Morozov montre avec érudition comment les outils de communication liés à Internet ont systématiquement été présentés comme « révolutionnaires » par des discours d'entrepreneurs, avant d'être généralement repris par les journalistes et cabinets de conseils. Ce phénomène qu'il nomme « époqualisme » a eu pour principal effet de rendre inaudibles les discours discordants qui visaient, au contraire, à tempérer cet enthousiasme débridé et qui rejetaient le mythe de la génération spontanée promu par les tenants de ces innovations techniques (Morozov, 2013).

ⁱⁱ Plus précisément, les propositions que nous présenterons ci-après proviennent en majeure partie de travaux de chercheurs en sociologie des sciences et techniques et en sciences de l'information et de la communication.

ⁱⁱⁱ Selon, les termes de G. Simondon (2001), voir *infra*.

^{iv} Ensemble de mesures visant à conférer un libre accès aux données publiques.

^v Voir, par exemple, Pasquale (2015) ou Carr (2015).

^{vi} Nous enjoignons le lecteur intéressé par ses thématiques à se reporter aux travaux d'A. Rouvroy et T. Berns (2013).

^{vii} A ce propos, l'anthropologue N. Seaver (2014) qualifie les algorithmes « d'actants » (en référence à la théorie de l'acteur-réseau) afin d'insister sur leur dimension pragmatique, sur le fait qu'ils contribuent de façon immédiate à une chaîne de co-opération qui comprend et affecte tant des machines que des individus.

^{viii} Nous laisserons ici de côté les aspects liés à la vie quotidienne des particuliers (aspects contribuant au désormais fameux « *quantified self* ») et les usages non directement liés à l'industrie (e-santé, domotique, etc.).

^{ix} L'Union Internationale des Télécommunications (UIT) a proposé en 2012 une définition canonique de « l'Internet des Objets » : « infrastructure mondiale pour la société de l'information, qui permet de disposer de services évolués en interconnectant des objets (physiques ou virtuels) grâce aux technologies de l'information et de la communication interopérables existantes ou en évolution. [...] En exploitant les capacités d'identification, de saisie de données, de traitement et de communication, [l'Internet des objets] tire pleinement parti des objets pour offrir des services à toutes sortes d'applications, tout en garantissant le respect des exigences de sécurité et de confidentialité. [...] Dans une optique plus large, [l'Internet des objets] peut être considéré comme un concept ayant des répercussions sur les technologies et la société (UIT, 2012, p. 1).

^x Les firmes dominant ces marchés sont des acteurs internationaux comme Acxiom, Visual Dna, Bluekai ou eXelate.

^{xi} GAFa : acronyme pour Google, Apple, Facebook et Amazon. Selon V. Mayer-Schönberger et K. Cukier (2014), leur position (ultra) dominante sur les marchés et la relation continue qu'ils entretiennent avec leurs consommateurs leur confèreraient un avantage concurrentiel tel, qu'il s'apparenterait à une barrière à l'entrée pour des concurrents potentiels. Parmi ces « léviathans du web », auxquels on peut adjoindre Microsoft et IBM, l'acteur qui serait en possession de la plus grande quantité de *data* et en capacité d'en extraire et d'en traiter le plus est, sans surprise, Google (Mayer-Schönberger, Cukier, 2014).

^{xii} Selon les chiffres présentés dans le Journal du Net, chiffres disponibles sur : <http://www.journaldunet.com/solutions/cloud-computing/marche-big-data.shtml> (page consultée le 14/12/2015)

^{xiii} Le paragon de cette approche est probablement la société Zynga qui a construit son succès non pas sur la vente des jeux-vidéos qu'elle produit mais sur sa capacité à extraire et valoriser les données personnelles des joueurs.

^{xiv} Selon les chiffres de la Direct Marketing Association (DMA - la plus importante association de professionnels du marketing au niveau mondial) disponibles sur : <http://thedma.org/advocacy/data-driven-marketing-institute/value-of-data/> (page consultée le 15/01/2016).

^{xv} Sur « l'économie créative » comme projet industriel et politique, voir : Bouquillion, Miège, Moeglin (2013) et Hartley (2005).

^{xvi} Pour mémoire, le rapport de Meta Group qui, selon J.P. Simon, va largement contribuer à diffuser l'idée d'une importance stratégique de premier plan relative à la capacité à traiter de très grand volume de données date de 2001 (Simon, 2015 ; Meta Group, 2001).

^{xvii} Pour illustrer les campagnes communicationnelles autour des *big data* de ces trois cabinets, voir respectivement : <http://www.gartner.com/technology/topics/big-data.jsp>, <http://www.ey.com/GL/en/Industries/Financial-Services/fso-insights-global-insights-ready-for-take-off>, <http://www2.deloitte.com/global/en/misc/litopicpage.global-topic-tags.bigdata.html> (pages consultées le 17/01/2016).

^{xviii} Nous ne traiterons pas ici de l'utilisation de ces « *Decision Support System* » dans un cadre d'ingénierie ou d'autres formes d'expertises et nous limiterons à leurs principaux usages commerciaux.

^{xix} Espace public au sens habermassien du terme, c'est-à-dire espace d'exposition et de délibération des idées et expressions (Habermas, 2003).

^{xx} Selon l'expression de D. Evans, A. Hagi, R. Schmalensee (2006) qui désignent ainsi le travail de certains acteurs qui cherchent à convaincre des tiers de l'adoption d'une technologie et de pratiques associées. L'objectif est d'atteindre rapidement une masse critique d'utilisateurs précoces (« *early adopters* »), ce qui aura pour effet d'imposer la technologie comme standard technique.

^{xxi} Voir note de fin n°xvii.

^{xxii} Dans son analyse de « l'imaginaire numéraire du numérique », O. Ertzscheid (2011) affirme en effet que « le vertige des grands nombres est constitutif de la statistique du web » et que « la mythologie de l'Internet – au sens des *Mythologies* de Barthes – est construite sur ces chiffres renvoyant à une nouvelle Babel statistique ». Le propos de l'auteur est de mettre en évidence qu'une grande part du succès des principaux sites Internet, et notamment les réseaux socio-numériques, repose sur des métriques et mesures « incommensurables » qui feignent d'attester le plébiscite car, comme le dit l'auteur, « plus il y a de chiffres, moins il y a de réalité ».

^{xxiii} Chiffres repris par des magazines en ligne et disponibles respectivement sur : <http://www.lemondeinformatique.fr/actualites/lire-big-data-un-marche-qui-croit-de-39-par-an-en-moyenne-selon-idc-49433.html> et <http://www.journaldunet.com/solutions/cloud-computing/marche-big-data.shtml> (pages consultées le 28/12/2015)

^{xxiv} Chiffres repris par des magazines en ligne et disponibles respectivement sur :
<http://www.transparencymarketresearch.com/pressrelease/big-data-market.htm> &
<http://www.zdnet.fr/actualites/l-avenir-du-big-data-nouveaux-concepts-et-forte-croissance-due-aux-metiers-39824090.htm> (pages consultées le 28/12/2015)

^{xxv} Chiffres consultables sur respectivement : <http://www.gartner.com/newsroom/id/2207915> (page consultée le 28/12/2015) & (EY, 2013, p. 51)

^{xxvi} Pour illustrer cette assertion, voir, par exemple : EY (2013, p. 50).

^{xxvii} Voir à ce sujet, par exemple, le blog d'Emin Gün Sirer (<http://hackingdistributed.com/>), professeur d'informatique à Cornell University qui critique sèchement les discours qui vantent la perfection technique en matière informatique, *a fortiori* ceux concernant les dispositifs prédictifs.

^{xxviii} Ces aspects sont partiellement pris en compte dans le cadre des 16 initiatives de la Commission Européenne pour une *Stratégie pour le Marché numérique unique* (texte présenté le 6 mai 2015).