Du *Psautier de Mayence* aux zettaoctets – Quel environnement juridique pour le *Big Data* ?

« Big data is like teenage sex : everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it. »
Dan Ariely, professeur de psychologie et d'économie comportementale, Duke University, Caroline du Nord, janvier 2013.

EMMANUEL JOUFFIN

Docteur en droit
Responsable juridique de banque

XAVIER LEMARTELEUR

Responsable juridique Technologies de l'information

Le *Big Data* est-il une révolution ? D'un point de vue technique, il s'agit beaucoup plus d'une évolution que d'une révolution ; d'un point de vue juridique, cette évolution technique pourrait conduire à une révolution juridique. Les principes de base induits par le *Big Data* – vaste collecte de données, le profilage et le principe de sérendipité – sont en nette opposition avec les concepts fondamentaux régissant la régulation de la vie privée en France et dans l'Union européenne. Le *Big Data* pose également la question de la propriété des données : toutes ces informations collectées sont-elles la propriété de l'entreprise ? Ces données personnelles appartiennent-elles à des sociétés ou à des individus ?

1. Dans une étude de 2011, Mc Kinsey[1] estimait que toute entreprise de plus de 1 000 salariés stockait en moyenne 200 téraoctets (10^{12}) de données, soit l'équivalent de la bibliothèque du Congrès américain, réputée être la plus grande du monde. Que de chemin parcouru depuis l'impression du premier ouvrage, le *Psautier de Mayence*, en 1457 ! Entre cette date et 1500, 8 millions de livres ont été imprimés, soit l'équivalent estimé de ce qu'avaient jusqu'alors produit les scribes depuis le III^e^ siècle après Jésus-Christ[2]. Depuis, le phénomène n'a fait que croître et embellir. Une phrase le résume : « *Prenez toutes les informations produites par l'humanité depuis l'aube des temps jusqu'en 2003 ; maintenant, nous produisons la même quantité en tout juste deux jours*[3]. » En termes de chiffres, depuis les origines de l'humanité et jusqu'en 2003, on estime la production de données à 5 exaoctets (10^{18}), soit 5 milliards de milliards d'octets. Pour la seule année 2013, on estime que 4,4 zettaoctets (10^{21}) de données ont été générés[4]. Face à l'augmentation du volume de données à traiter, les systèmes d'information évoluent ; la dernière révolution est dénommée « *Big Data* ».

2. Selon les archives de la bibliothèque numérique de l'ACM (Association for Computing Machinery), l'expression « *Big Data* » serait apparue en octobre 1997 dans des articles scientifiques sur les défis technologiques à relever pour visualiser les « *grands ensembles de données* »[5]. Ce concept est susceptible de divers niveaux de lecture. En premier lieu, le *Big Data* est avant tout une innovation *marketing*, comme l'avaient été auparavant le web 2.0, le *cloud computing* et ses variantes SaaS[6], IaaS[7] et PaaS[8] et comme l'est déjà l'IoT (*internet of things* ou Internet des objets).

1. *Big Data : the Next New Frontier for Innovation, Competition and Productivity*: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.
2. Elizabeth L. Eisenstein, *La Révolution de l'imprimé. À l'aube de l'Europe moderne*, Paris, La Découverte, mai 1991.
3. Neelie Kroes, vice-présidente de la Commission européenne en charge de l'agenda numérique, 7 novembre 2013.
4. Si l'on s'en tient toutefois au strict volume de données et non à leur qualité.
5. https://fr.wikipedia.org/wiki/Big_data.
6. *Software as a Service* : modalité d'exploitation de logiciels installés sur des serveurs distants plutôt qu'en local sur la machine de l'utilisateur.
7. *Infrastructure as a Service* : également une typologie de *cloudcomputing*.
8. *Platform as a Service* : type de *cloudcomputing* destiné aux entreprises.

Si les technologies de l'information ne font qu'évoluer au rythme de la loi de Moore[9], le marketing ne connaît que révolutions après révolutions. Le *Big Data* est donc plus une révolution marketing qu'une révolution technologique.

3. Sur le plan technologique, le *Big Data* s'inscrit dans la continuité des développements de l'informatique qui nous ont menés des premiers ordinateurs apparus au cours de la deuxième guerre mondiale aux supercalculateurs[10]. En effet, le *Big Data* n'est pas apparu *ex nihilo* : il trouve sa filiation dans les entrepôts de données (*data warehouses*) qui ont déjà « révolutionné » les années 2000 et dont le concept remonte à la fin des années 1980[11].

4. Sur le plan juridique cette évolution technique pourrait bien toutefois conduire à une réelle révolution, remettant en cause une partie des grands principes sur lesquels sont fondées certaines branches importantes du « *droit des technologies de l'information* », comme la protection des données à caractère personnel. Nous nous attacherons, dans une première partie, à présenter le *Big Data* (I.) en évoquant ses caractéristiques et enjeux (1.) puis, en nous préoccupant de sa cohérence avec la protection des données à caractère personnel, tant du point de vue de la loi informatique et libertés, que du point de vue du futur règlement européen (2.).

Dans une deuxième partie, nous nous interrogerons sur le fait de savoir si, du fait même des contraintes juridiques, les données ne sont pas, tout à la fois, la raison d'être du *Big Data*, mais peut-être aussi sa plus grande faiblesse (II.). Nous aborderons ce sujet sous deux aspects particuliers : les bases de données tout d'abord (1.), puis sous l'angle de l'existence d'un droit de propriété sur les données à caractère personnel (2.).

I. PRÉSENTATION GÉNÉRALE DU *BIG DATA*

1. Caractéristiques et enjeux du *Big Data*

« *Demandez à n'importe quel* Chief Data Officer *de définir Big Data et il va se mettre à regarder ses chaussures. En réalité, il y a de forte chance pour que vous obteniez autant de définitions différentes que le nombre de personnes auxquelles vous poserez la question* »[12].

1.1. Esquisse de definition technique

a. Une évolution et non une révolution technique

5. Si à la fin des années 2000, les *data warehouses* révolutionnaient le stockage de données, c'est en s'appuyant notamment sur des formats permettant de stocker et de traiter des données non structurées (XML), de son côté, le web 2.0[13] se démarquait par un usage important d'AJAX[14]. Ce qui caractérise les grandes ruptures est souvent le socle technique qui les supporte.

Si l'on en vient au *Big Data*, une première tentative de définition peut être faite en fonction des technologies utilisées. Ainsi, toute personne confrontée à un projet *Big Data* verra nécessairement apparaître des termes tels que « Hadoop[15] », « NoSQL[16] » ou encore « Mapreduce[17] ». Il n'est pas question ici de détailler le fonctionnement technique de ces divers éléments, mais on retiendra que ces socles technologiques sont généralement libres[18] et qu'ils ont vocation à permettre le traitement d'un énorme volume de données. Ce dernier aspect, le volume des données traitées, est un des éléments caractéristiques du *Big Data* constituant la première brique d'une définition fonctionnelle dite « par les 3 V ».

b. Une définition traditionnelle par les « 3 V »

6. Un cadre de définition a été proposé dès 2001 dans un rapport de recherche du cabinet Gartner, ce dernier a dégagé trois critères cumulatifs permettant de définir le *Big Data* au travers des « 3V » (volume, vélocité et variété, ces éléments étant cumulatifs). Le volume est le premier critère permettant de déterminer l'existence d'un *Big Data*. Pour autant, se pose la question de la masse critique à atteindre. La rapidité de traitement des données peut s'avérer décisive encore qu'elle ne soit pas nécessairement prépondérante. En effet, si l'instantanéité, ou tout du moins une très grande rapidité de traitement peut être utile pour « scorer » en temps réel un client, riposter à une fraude ou bien encore piloter du trading à haute fréquence, elle peut être moins cruciale pour des traitements nécessitant une intervention humaine en vue d'une prise de décision.

7. Le troisième critère, celui de la variété des données traitées, manifeste un enjeu majeur du *Big Data*. L'agréation des informations présentes sous des formats divers (fichiers textes, audio, vidéo) issus de médias variés (réseaux sociaux, bases de données clients, objets connectés) est un des défis majeurs si l'on souhaite avoir d'une personne un profil le plus complet possible.

À ces trois critères, peuvent s'ajouter plusieurs autres « V ». Un quatrième « V » est relatif à la véracité des données collectées, critère prenant en considération la dimension qualitative de l'information. À ce critère, s'ajoute parfois celui de la Visibilité[19].

9. La Loi de Gordon Moore (on parle aussi de conjectures de Moore) est apparue dans Electronics Magazine en 1965. Cette loi postule que le nombre de transistors par circuit intégré doit doubler, à prix constants, tous les 18 mois et ce jusqu'en 2017.

10. Ces supercalculateurs sont devenus des enjeux de prestige national, se matérialisant par une course à la puissance de calcul. Le record est de 1 000 petaflops (un milliard de milliards de calculs par seconde).

11. B.A. Devlin et P.T. Murphy, « An architecture for a business and information system », *IBM Systems Journal*, vol. 27, n° 1. 1988 accessible à l'adresse suivante : http://altaplana.com/ibmsj2701G.pdf. On peut par ailleurs noter que ces *data warehouses* n'ont fait que succéder aux « infocentres » des années 1960-1970.

12. « Big Data Gets Persona », *MIT Review*, octobre 2013.

13. Évolutions des techniques, fonctionnalités et usages du web permettant un usage plus simple et interactif.

14. *Asynchronous JavaScript and* XML.

15. *Framework* sous Java libre permettant de faciliter la création d'applications aptes à gérer un volume très important de données.

16. *Not Only SQL* : systèmes de gestion de base de données de grande ampleur.

17. Architecture de développement informatique inventée par Google permettant le traitement de données très volumineuses, supérieures à 1 téraoctet.

18. La grande majorité des logiciels et autres éléments nécessaires aux plates-formes *Big Data* sont distribués sous une licence logiciel libre.

19. Ce critère rend des divers types de données qui peuvent être collectées. Diversité liée au format (texte ou image) de leur nature (données publiques, statistiques, géo-démographiques ou bien encore privées).

c. Définition légale ?

8. Il n'existe pas de définition légale *stricto sensu* du *Big Data*, la commission de terminologie substitue le terme français « mégadonnées » à l'expression anglo-saxonne *Big Data*, dans le même temps, elle y associe une ébauche de définition. Le *Big Data*, ou plutôt les « mégadonnées », serait des « *données structurées ou non dont le très grand volume requiert des outils d'analyse adaptés* »[20].

Le G29[21] précise la notion en définissant le *Big Data* comme « *[…] la croissance exponentielle dans la disponibilité et le traitement automatisé de l'information : [le Big Data] se réfère à des jeux gigantesques de données détenus par des entreprises, des gouvernements et d'autres organisations, qui sont alors largement analysés à l'aide d'algorithmes*[22] *informatiques. Le Big Data s'appuie sur les capacités croissantes des technologies à collecter et stocker de grandes quantités de données, mais aussi de les analyser, d'en comprendre le sens afin de prendre avantage de la valeur de la donnée […]. Les attentes liées au Big Data sont qu'il pourrait en fin de compte conduire à des décisions meilleures et plus opportunes* »[23]. Malgré ces quelques ébauches de définition, on ne peut toutefois que constater la défaillance du droit à définir précisément la notion de *Big Data*.

9. En synthèse, on peut retenir toutefois que le *Big Data* repose sur des socles logiciels permettant de traiter un volume de données important et hétérogène en quasi-temps réel et qu'il s'appuie sur le principe de « sérendipité »[24] : peu importent les données traitées, leur qualité et leur nature, le fait de multiplier les sources par le truchement d'algorithmes doit permettre d'obtenir des informations utiles *in fine*, c'est-à-dire économiquement exploitables. En effet, le développement du *Big Data* s'inscrit dans de nouveaux modèles économiques construits sur la donnée.

1.2. Quels enjeux économiques ?

10. Pour le G29[25] : « *Il existe de nombreuses applications pour les grands volumes de données dans divers secteurs, y compris les soins de santé, les communications mobiles, la distribution d'énergie, la gestion du trafic, la détection des fraudes, la commercialisation et la vente au détail, à la fois en et hors ligne. Le* Big Data *peut être utilisé pour identifier les tendances générales et des corrélations, mais son traitement peut également affecter directement les personnes. Par exemple, dans le domaine du marketing et de la publicité, le* Big Data *peut être utilisé pour analyser prévoir des préférences personnelles, les comportements et les attitudes des clients individuels et induiront ensuite des mesures ou décisions à l'égard de ces clients, tels que des rabais personnalisés, des offres spéciales et des publicités ciblées en fonction du profil du client* »[26].

Le *Big Data* permet d'entrevoir un marketing d'un nouveau type ainsi résumé : « *nous sommes en train de passer d'un modèle classique de segmentation à un modèle de caractérisation comportementale. [ ] Le* profiling *des clients apporte sans aucun doute une valeur ajoutée à l'entreprise qui peut alors affiner et personnaliser ses produits et ses offres* »[27]. Une telle évolution est à rapprocher des enjeux économiques que représente le *Big Data* dont le chiffre d'affaires était estimé à 6,3 milliards de dollars en 2012, pour atteindre 8,9 milliards de dollars en 2014 et 24,6 milliards de dollars attendus en 2016[28]. En ce qui concerne la croissance des revenus de ce nouveau marché, elle est estimée à plus de 40 % par an en moyenne, soit 100 millions de dollars estimés en 2009 et 50 milliards de dollars en 2018[29].

Si le *Big Data* n'est pas une révolution technologique, il n'en demeure pas moins qu'il ouvre de nouvelles perspectives en termes marketing notamment, permettant une connaissance client accrue. Au plan juridique, le *Big Data* suscite d'importantes interrogations concernant sa compatibilité avec les principes fondamentaux du droit des données personnelles.

2. Le *Big Data* est-il soluble dans les données à caractère personnel ?

11. Le *Big Data*, en raison de son fonctionnement même, remet en cause certains des piliers qui soutiennent la protection des données personnelles depuis 1978. Ainsi, des principes fondamentaux tels que celui de « *finalité* » du traitement, de « *proportionnalité* » ou encore, la notion même de donnée à caractère personnel sont questionnés par le *Big Data*, que ces principes soient énoncés par la loi Informatique et Libertés ou bien encore, par le règlement général sur la protection des données[30] qui vise à

20. Journal officiel du 22 août 2014 : « Avis : vocabulaire de l'informatique ».

21. Le G29 ou Groupe de travail « Article 29 » sur la protection des données est un organe consultatif européen indépendant sur la protection des données et de la vie privée. Son organisation et ses missions sont définies par les articles 29 et 30 de la directive 95/46/CE, dont il tire sa dénomination, et par l'article 14 de la directive 97/66/CE. Il regroupe des représentants des différentes autorités de protection des données de l'Union.

22. Un algorithme (du nom du mathématicien perse Al-Khawarizmi, latinisé au moyen âge en Algoritmi) est une suite d'opérations ou d'instructions à appliquer dans un ordre déterminé et permettant de résoudre un problème ou d'obtenir un résultat donné.

23. Article 29 Data Protection Working Party, 00569/13/EN, WP 203, Opinion 03/2013 on Purpose Limitation adopted on 2 April 2013. Traduction libre de l'anglais : « *Big data refers to the exponential growth in availability and automated use of information: it refers to gigantic digital datasets held by corporations, governments and other large organisations, which are then extensively analysed using computer algorithms. Big data relies on the increasing ability of technology to support the collection and storage of large amounts of data, but also to analyse, understand and take advantage of the full value of data (in particular using analytics applications). The expectation from Big Data is that it may ultimately lead to better and more informed decisions.* »

24. De l'anglais « *serendipity* ». Le terme aurait été inventé par Horace Walpole (1717-1797) dans une lettre du 28 janvier 1754 à son ami Horace Mann : « *[…] cette découverte est presque de l'espèce que j'appelle serendipity, un mot très expressif que je vais m'efforcer, faute d'avoir mieux à vous narrer, de vous expliquer : vous le comprendrez mieux par l'origine que par la définition. J'ai lu autrefois un conte de fées saugrenu, intitulé Les Trois Princes de Serendip : tandis que leurs altesses voyageaient, elles faisaient toute sorte de découvertes, par accident et sagacité, de choses qu'elles ne cherchaient pas du tout : par exemple, l'un des princes découvre qu'un chameau borgne de l'œil droit vient de parcourir cette route, parce que l'herbe n'a été broutée que sur le côté gauche, où elle est moins belle qu'à droite - maintenant saisissez-vous le sens de* serendipity *? L'un des exemples les plus remarquables de cette sagacité accidentelle […].* »

25. Article 29 Data Protection Working Party, 00569/13/EN, WP 203, Opinion 03/2013 on Purpose Limitation adopted on 2 April 2013, Annex 2: Big Data and Open Data, p. 45.

26. Traduction libre. Texte original : « *There are numerous applications of Big Data in various sectors, including healthcare, mobile communications, smart grid, traffic management, fraud detection, marketing and retail, both on and offline. Big data can be used to identify general trends and correlations but its processing can also directly affect individuals. For example, in the field of marketing and advertisement, Big Data can be used to analyse or predict the personal preferences, behaviour and attitudes of individual customers and subsequently inform 'measures or decisions' that are taken with regard to those customers such as personalised discounts, special offers and targeted advertisements based on the customer's profile.* »

27. « *Les données numériques : un enjeu d'éducation et de citoyenneté* », avis du Conseil économique, social et environnemental présenté par M. Éric Peres, rapporteur au nom de la section de l'éducation, de la culture et de la communication, spéc. p. 17.

28. Transparency Market Research, « Big Data Market Global Scenario, Trends, Industry Analysis, Size, Share and Forecast 2012-2018 ».

29. La Cour de cassation a posé des limites et rappelé qu'un fichier de clientèle non déclaré à la CNIL ne pouvait être vendu, étant du fait même de cette non-déclaration, chose hors du commerce : Cass. com. 25 juin 2013, *Revue des contrats*, 1er mars 2014, n° 1, p. 119.

30. Proposition de règlement du Parlement européen et du Conseil relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. La première proposition de texte a été publiée le

adapter le cadre légal en matière de privacy aux derniers développements technologiques.

2.1. Extension de la notion de donnée à caractère personnel

12. Le traitement massif de données hétérogènes qu'induit le *Big Data* conduit quasi nécessairement au traitement de données à caractère personnel. La notion de données à caractère personnel répond à une définition large énoncée par l'article 2 de la loi Informatique et Libertés disposant que « *constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres* »[31].

Ainsi, toute donnée faisant apparaître directement l'identité d'une personne physique entre dans le champ d'application de la loi de 1978, mais aussi indirectement toute information rattachée à une personne physique identifiée ou identifiable[32].

13. Ce même article vient limiter ce que l'on doit entendre par « *identifiable* » en y adjoignant une limite : « *pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne* ».

C'est notamment cette limite que le *Big Data* vient repousser. En effet, les ressources colossales en termes de collecte, stockage et traitement de l'information conduisent à rendre les individus de plus en plus aisément identifiables. En effet, la multiplication de collectes de données non nominatives *ab initio* et la mise en œuvre d'algorithmes de profilage permettent désormais une identification des personnes, du moins indirectement.

14. Le *Big Data* marque-t-il la fin de l'anonymat ? Tout dépend de ce que l'on entend par anonymat. Si l'on retient la conception exposée par la CNIL selon laquelle « *pour qu'une solution d'anonymisation soit efficace, elle doit empêcher toutes les parties d'isoler un individu dans un ensemble de données, de relier entre eux deux enregistrements dans un ensemble de données (ou dans deux ensembles de données séparés) et de déduire des informations de cet ensemble de données* »[33], il est certain que l'anonymat est grandement mis à mal par le *Big Data* et les possibilités en termes de profilage par recoupements qu'il permet.

15. Cette disparition de l'anonymat est encore renforcée par l'extension prétorienne de la définition de la notion de donnée à caractère personnel. La CNIL a pu ainsi considérer dans le cadre d'un traitement en *Big Data* que s'« *il est exact que les données collectées à l'occasion de la navigation d'utilisateurs non authentifiés (actifs ou passifs) ne peuvent être automatiquement considérées comme directement identifiantes dès lors que, prises isolément, elles ne se rapportent qu'au seul terminal (fixe ou portable) ou au seul navigateur de l'utilisateur, non identifié.* [...] *La formation restreinte relève, à cet égard, que le seul et unique objectif poursuivi par la société consiste à recueillir un maximum d'éléments sur des personnes singularisées afin d'optimiser la valorisation de leurs profils sur le plan publicitaire. Son modèle économique ne requiert donc pas de connaître les nom, prénom, adresse ou autres éléments directement identifiants sur les personnes, qui ne lui sont pas nécessaires pour les reconnaître lors de chaque nouvel usage qu'elles feront de ses services.*

Cependant, les éléments qu'elle collecte à cette fin, qui peuvent être combinés entre eux [...], *lui permettent de cerner avec une précision extrême le comportement d'une machine et, derrière celle-ci, de son utilisateur, à laquelle elle est capable in fine d'affecter les caractéristiques de son activité quotidienne, ses interactions avec autrui, ses centres d'intérêt, des éléments liés à sa personnalité, ses choix de vie, etc.*

En d'autres termes, l'accumulation de données qu'elle détient sur une seule et même personne lui permet de la singulariser *à partir d'un ou de plusieurs éléments qui lui sont propres. Ces données doivent, en tant que telles, être considérées comme identifiantes et non comme anonymes* »[34].

Cela n'est pas neutre sur le plan juridique, le traitement de données à caractère personnel implique le respect de l'ensemble des dispositions sur la protection des données. La première des contraintes issue de la loi de 1978 est celle de la licéité du traitement et la nécessité d'en déterminer la finalité. Cette obligation ne va pas sans quelques difficultés d'application dans le contexte du *Big Data*.

2.2. Les données à caractère personnel face à la finalité versus sérendipité

16. Consubstantiellement, le *Big Data* ne répond pas à une finalité précisément déterminée ou déterminable dès l'origine de la collecte des données comme précédemment évoqué. Il repose en grande partie sur le principe de sérendipité qui postule que l'utilité de la donnée, sa valeur et donc sa finalité, naît de son croisement, de son traitement et de sa qualification au travers d'algorithmes. En conséquence, le stockage des données dans le *Big Data* ne répond pas à l'origine à une finalité particulière. En ce sens, le *Big Data* marque une profonde antinomie de principe avec les obligations issues de la réglementation sur la protection des données.

En effet, l'article 6 de la loi Informatique et Libertés dresse les conditions de licéité des traitements, au 2° de ce même article, le texte impose que les données à caractère personnel soient « *collectées pour des finalités déterminées, explicites et légitimes et ne* [soient] *pas traitées ultérieurement de manière incompatible avec ces finalités* ».

25 janvier 2012 : European Commission, Brussels, 25.1.2012, COM(2012) 11 final, 2012/0011 (COD), Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), {SEC(2012) 72 final} / {SEC(2012) 73 final}.

31. Pour plus de précisions sur la définition de « donnée à caractère personnel », voir : Groupe de travail « article 29 » sur la protection des données, Avis 4/2007 sur le concept de données à caractère personnel, 20 juin 2007 ; 01248/07/FR, WP 136.

32. On notera les hésitations relatives à la qualification de l'adresse IP en tant que donnée à caractère personnel.

33. CNIL, délibération n° 2015-255 du 16 juillet 2015 refusant la mise en œuvre par la société JC Decaux d'un traitement automatisé de données à caractère personnel ayant pour finalité de tester une méthodologie d'estimation quantitative des flux piétons sur la dalle de La Défense (demande d'autorisation n° 1833589).

34. CNIL, délibération n° 2013-420 du 3 janvier 2014 de la formation restreinte prononçant une sanction pécuniaire à l'encontre de la société Google Inc.

Ce principe de finalité n'est pas remis en cause par la proposition de règlement général sur la protection des données[35] dont le considérant 25 énonce, de manière très claire, que « *le consentement donné devrait valoir pour toutes les activités de traitement ayant la même finalité. Lorsque le traitement a plusieurs finalités, un consentement sans ambiguïté devrait être donné pour l'ensemble des finalités du traitement.* » L'article 5 de cette même proposition est tout aussi clair sur ce sujet et fixe le principe selon lequel les données ne peuvent être collectées qu'en vue de « *finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement de manière incompatible avec ces finalités ; le traitement ultérieur des données à caractère personnel à des fins d'archivage dans l'intérêt public ou à des fins scientifiques, statistiques ou historiques n'est pas considéré, conformément à l'article 83, comme incompatible avec les finalités initiales* ».

17. De son côté, le G29, dans une déclaration du 16 septembre 2014[36], rappelle, quant à lui, qu'il n'y a pas de raison de croire que les principes de protection des données de l'Union, tels qu'ils sont actuellement inscrits dans la directive 95-46, ne sont plus valables et appropriés pour le développement du *Big Data*, sous réserve de futures améliorations afin de les rendre plus efficaces en pratique. Il précise, par ailleurs, que ces principes sont applicables à toutes les opérations de traitement, en commençant par la collecte et que, leur respect, est essentiel afin d'assurer une concurrence équitable et efficace entre les acteurs économiques sur les marchés concernés.

18. Ce même G29, dans une opinion consacrée cette fois-ci à la limitation de la finalité des traitements[37], après avoir rappelé[38] que la finalité s'entend d'amont en aval, de la collecte en vue de finalités spécifiées, explicites et légitimes jusqu'aux traitements ultérieurs compatibles avec les finalités exprimées en amont, poursuit en précisant qu'un traitement dans un but différent ne signifie pas nécessairement qu'il y ait incompatibilité. Le G29 se livre alors à une analyse casuistique de compatibilité, laquelle doit tenir compte des éléments suivants :

la relation entre les buts pour lesquels les données personnelles ont été recueillies et les finalités des réutilisations ultérieures ;

le contexte dans lequel les données personnelles ont été collectées et les attentes raisonnables des personnes concernées quant à leur utilisation ultérieure ;

la nature des données personnelles et de l'impact de leur utilisation ultérieure ;

les garanties relatives à un traitement équitable et afin d'éviter tout impact excessif sur les personnes concernées.

Cette analyse casuistique est le point central de tout projet *Big Data*. En raison de la difficulté à déterminer une finalité précise en amont, il conviendra de s'appuyer sur une finalité globale permettant de couvrir de manière suffisamment large les traitements subséquents découverts au travers des profilages opérés par le biais des algorithmes mis en œuvre dans le cadre du *Big Data*.

19. L'utilisation d'algorithmes en vue de réaliser un profilage des individus n'est d'ailleurs pas sans conséquences du point de vue des risques d'atteinte à leur vie privée. Pour Vinton Cerf, considéré comme l'un des pères fondateurs de l'internet, la cause est entendue et la vie privée n'est guère qu'un incident de parcours, une notion révolue, liée à la société industrielle du XIXe siècle. La norme est désormais la « *publitude* »[39], fondée sur le principe « *si vous n'avez rien à cacher, pourquoi être inquiet ?* ». Rêve de bien des dictatures, ce *big brother* planétaire est en train de se bâtir de façon non seulement volontaire mais même enthousiaste, au travers des réseaux sociaux et des objets connectés[40].

Nous sommes bien loin des principes défendus par l'arrêt de la CEDH *Rotaru*[41] pour qui le simple fait de mémoriser des données relatives à la vie privée d'un individu constitue une ingérence au sens de l'article 8 de la Convention européenne des Droits de l'Homme laquelle garantit le droit au respect de la vie privée et familiale, du domicile et de la correspondance, indépendamment du fait de savoir si ces informations mémorisées sont ou non utilisées par la suite. Bien loin aussi des préoccupations de l'arrêt *Schrems*[42] à l'égard des pouvoirs détenus par les agences de renseignement US, préoccupation ayant conduit à l'invalidation pure et simple de la décision de la Commission du 26 juillet 2000 qui avait considéré les principes du « *Safe Harbor* » comme assurant un niveau de protection des données personnelles adéquat s'agissant des transferts de données personnelles de l'Europe à destination des États-Unis.

20. Le Conseil d'État, dans son rapport annuel 2014[43], s'inquiète lui aussi du recours à des algorithmes risquant d'enfermer les internautes dans un « *déterministe numérique* » conduisant à ce qu'une personne se réduise à la somme du traitement algorithmique des informations la concernant.

De son côté, la proposition de règlement général sur la protection des données aborde directement la question du profilage[44], énonçant à cette occasion divers prin-

35. Dans sa rédaction issue de l'orientation générale du 15 juin 2015.

36. WP 29, Statement on Statement of the WP29 on the impact of the development of Big Data on the protection of individuals with regard to the processing of their personal data in the EU, spéc. p. 2.

37. WP29, Opinion 03/2013 on purpose limitation, 2 avril 2013.

38. *Ibid.*, p. 3 et p. 38.

39. « Les données numériques : un enjeu d'éducation et de citoyenneté », Avis du Conseil économique, social et environnemental préc., spéc. p. 47.

40. Selon un rapport d'avril 2015 de l'Institut Montaigne intitulé « Big Data et objets connectés. Faire de la France un champion de la révolution numérique », spéc. p. 14 et s. : « *L'Internet des objets contribuerait à doubler la taille de l'univers numérique tous les 2 ans, lequel pourrait représenter 44 000 milliards de gigaoctets en 2020, soit 10 fois plus qu'en 2013 [...] Le potentiel de "choses" qui pourraient être connectées d'ici 2020 est estimé entre 30 et 212 milliards selon les sources retenues.* »

41. CEDH 4 mai 2000, *Rotaru c/ Roumanie*, Req. n° 28341/95. Outre cet arrêt, on citera principalement : CEDH, *Amann c/ Suisse* du 16 février 2000, req. n° 27798/95 et *Segerstedt-Wiberg et autres c/ Suède*, 6 juin 2006, req. n° 62332/00. *Idem* : arrêt *Perry c/ Royaume-Uni* du 17 juillet 2003, n° 63737/00, § 36. La Cour EDH estime à cette occasion : « *Il existe donc une zone d'interaction entre l'individu et autrui qui, même dans un contexte public, peut relever de la "vie privée".* »

42. CJUE, gde ch., 6 oct. 2015, n° C-362/14, *Maximillian Schrems c/ Data Protection Commissionner* ; CCE n° 12, décembre 2015, étude 21, « Arrêt *Schrems* : cour(s) magistral(e) de droit à la protection des données personnelles », étude par R. Perray et J. Uzan-Naulin. J.-L. Sauron, GP 29 octobre 2015, n° 302, p. 7.

43. Rapport annuel du 9 septembre 2014 : Le numérique et les droits fondamentaux : spéc. p. 237 et s. : http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/144000541.pdf.

44. Article 20 du projet de règlement.

cipes. Tout d'abord, toute personne physique doit être informée, de façon claire, de l'existence de son droit à s'opposer au profilage[45]. Par ailleurs, le profilage ne peut être un instrument de discrimination et doit nécessairement inclure une appréciation humaine dès lors que ce profilage produit des effets juridiques ou affecte de manière sensible la personne concernée.

Ces nouvelles dispositions, si elles ne constituent pas en tant que telles des obstacles insurmontables aux projets *Big Data*, devront tout de même être prises en compte et pourront constituer des limites pour ces projets. Par ailleurs, au regard de l'intérêt et la vigilance que suscite le profilage des individus dans nos sociétés, on ne peut exclure que, dans l'avenir, ce type de pratiques soit plus sévèrement encadré.

Cependant il ne s'agit pas là des seules difficultés auxquelles se confronte le *Big Data*, une autre pierre angulaire de la réglementation applicable à la protection des données vient encore restreindre les conditions dans lesquelles les données à caractère personnel peuvent être traitées dans un contexte *Big Data*, il s'agit du principe de proportionnalité.

2.3. Proportionnalité et *Big Data*

21. Seules les données nécessaires à la poursuite de la finalité du traitement peuvent être traitées par l'entreprise, c'est ce qui ressort des dispositions du paragraphe 3 de l'article 6 de la loi Informatique et Libertés. Ce principe vient en opposition avec celui qui anime les projets autour du *Big Data*. Comme évoqué précédemment, le fonctionnement intrinsèque de ce dernier repose sur l'accumulation de données variées afin, au travers d'algorithmes, de faire apparaître des schémas qui pourront être exploités par l'entreprise. Les données qui seront déversées dans le « *data lake* » qui alimente le *Big Data*, ne sont pas, par nature, nécessaires ou « non excessives ». Au contraire, la finalité du *Big Data* étant souvent indéterminée, toute donnée, indépendamment de son intérêt ou de sa valeur, a vocation à être conservée et traitée.

22. Si la notion de donnée « non excessive » laisse place à une certaine marge d'interprétation, on notera que, dans sa première version, le projet de règlement européen sur la protection des données adoptait une attitude plus stricte en imposant un principe de minimisation. Ainsi l'article 5 c proposé par la Commission disposait que « *les données à caractère personnel doivent être [...] adéquates, pertinentes et limitées au minimum nécessaire au regard des finalités pour lesquelles elles sont traitées* ». Un tel principe avait vocation à mettre un point d'arrêt au traitement massif de données personnelles et donc, incidemment, au *Big Data*. Pour l'heure, la version de compromis du texte maintient cette notion tout en la vidant de sa substance. L'article 5 c) précise ainsi que les données doivent être « *adéquates, pertinentes et* limitées à ce qui est nécessaire *au regard des finalités pour lesquelles elles sont traitées (minimisation des données)* ». Une telle rédaction marque un quasi-abandon de la notion initiale de minimisation pour revenir au principe proche de celui de la directive de 95-46 et de la loi Informatique et Libertés, à savoir celui de données « *non excessives* ». En tout état de cause, la nécessité de limiter le traitement aux données « non *excessives* » au regard de la finalité du traitement[46] restreint les possibilités liées au *Big Data* et se place dans une certaine dichotomie avec la notion de sérendipité qui l'anime.

23. Enfin La conservation d'une grande quantité de données dans une architecture *Big Data* n'est envisageable que dans la mesure où sera assurée la sécurité des données agrégées. C'est encore une obligation découlant de la loi Informatique et Libertés qui, si elle n'est pas propre au *Big Data*, constitue tout de même un enjeu particulier dans le contexte d'un traitement de données massif.

2.4. La sécurité des données personnelles au travers des déclarations des failles[47]

24. D'ores et déjà, la loi Informatique et Libertés prévoit une obligation particulière de sécurité s'agissant de la protection des données à caractère personnelle. Son article 34 dispose à cet effet que le « *responsable du traitement est tenu de prendre toutes précautions utiles, au regard de la nature des données et des risques présentés par le traitement, pour préserver la sécurité des données et, notamment, empêcher qu'elles soient déformées, endommagées, ou que des tiers non autorisés y aient accès* ». Cette obligation, pénalement sanctionnée[48], reste néanmoins très générale, le texte ne détaillant aucunement les mesures de sécurité devant être mises en place pour satisfaire à l'obligation. Dans le silence légal, l'entreprise ne pourra que se référer à un état de l'art existant constitué de recommandations et de normalisations[49].

25. Outre ce devoir de vigilance, l'article 34 bis II prévoit, s'agissant de « *fourniture au public de services de communications électroniques sur les réseaux de communications électroniques ouverts au public* » (ou plus simplement dit : pour les opérateurs de télécommunications), que la violation de données à caractère personnelle[50] donne lieu à une information sans délai de la CNIL. Cette information est doublée d'une information de la clientèle dès lors que « [...] *cette violation peut porter atteinte aux données à caractère personnel ou à la vie privée d'un abonné ou d'une autre personne physique, le fournisseur avertit également, sans délai, l'intéressé* » (art. 34 bis II al. 2). Chaque opérateur doit tenir à jour « *un répertoire des inventaires des violations de données à caractère personnel, notamment de leurs modalités, de leur effet et des mesures prises pour y remédier et le conserve à la disposition de la commission* ».

45. Article 14 h) du projet de règlement.

46. Même si cette finalité est formulée de manière large comme évoqué précédemment.

47. Nous n'aborderons pas sujétions issues de la loi n° 2013-1168 du 18 décembre 2013 relative à la programmation militaire pour les années 2014 à 2019 et portant diverses dispositions concernant la défense et la sécurité nationale et de ses textes divers textes d'application.

48. Art. 226-17 du Code pénal.

49. Par exemple, les normes ISO de la famille 27 000 s'agissant de normalisation. En termes de recommandations, on pourra utilement se référer à celles de la CNIL mais surtout aux différents guides publiés par l'ANSSI.

50. Art. 34 bis I : « [ ] *on entend par violation de données à caractère personnel toute violation de la sécurité entraînant accidentellement ou de manière illicite la destruction, la perte, l'altération, la divulgation ou l'accès non autorisé à des données à caractère personnel faisant l'objet d'un traitement dans le cadre de la fourniture au public de services de communications électroniques* ».

La violation de l'obligation de notification d'une violation de données à caractère personnel auprès de la CNIL, ou auprès de l'intéressé, est punie de cinq ans d'emprisonnement et de 300 000 euros d'amende[51]. On retiendra que le projet de règlement européen[52] a vocation à étendre cette obligation, pour l'heure limitée aux opérateurs de télécom, à toutes les entreprises sans distinction sectorielle tout en renforçant les sanctions qui peuvent être prononcées par les régulateurs en cas de manquement.

26. Il est certain que le développement des technologies en matière de traitement de l'information ne va pas sans causer de heurts. En matière de *Big Data*, si la protection des données vient fixer un cadre contraignant parfois difficilement *conciliable* avec les impératifs légaux, ce n'est pas le seul texte qui vient éroder des fondements sur lesquels repose le *Big Data*. L'appropriation même des ressources que constituent les données doit être questionnée. La donnée, carburant du *Big Data*, est-elle librement exploitable ?

II. LES DONNÉES : FORCE ET FAIBLESSE DU *BIG DATA* ?

27. On ne cesse de lire que les données seraient « *le nouvel or noir du XXIe siècle* » ; elles sont perçues comme un actif que les entreprises se doivent de collecter, parfois à grands frais, et sur lesquelles elles disposeraient d'un droit de propriété.

Cependant, la donnée, et tout particulièrement celle qui se rapporte à un individu identifié ou identifiable, peut-elle être l'objet d'une appropriation par l'entreprise ? Appartient-elle à l'individu qui en est à l'origine ou est-elle encore un bien non-disponible ? Les entreprises investiraient-elles dans un actif susceptible de s'évaporer sans laisser la moindre trace ?

1. Le sort des bases données

La question reste complexe et mérite d'être examinée sous deux angles distincts. Le premier concerne la base de données en elle-même, le second s'attache aux données qui la composent.

1.1. Concernant la base de données elle-même

28. Le cadre légal applicable à la protection des bases de données est un cadre ancien et stable qui ne devrait pas être remis en cause par l'avènement du *Big Data* et de ses data lakes. Ce cadre protecteur résulte des dispositions des articles L. 342-1[53] et suivants du Code de la propriété intellectuelle, lesquels établissent un droit *sui generis* au profit du « *producteur de la base de données* ».

Ce droit est accordé au « producteur de la base de données », c'est-à-dire à « *la personne qui prend l'initiative et le risque des investissements correspondants* », la protection accordée concerne le « *contenu de la base lorsque la constitution, la vérification ou la présentation de celui-ci atteste d'un investissement financier, matériel ou humain substantiel* »[54].

Ces mécanismes de protection trouvent pleinement à s'appliquer dans le contexte du *Big Data* et participent donc à reconnaître à l'entreprise, devenue producteur de base de données, un droit de propriété opposable aux tiers.

29. Un autre mécanisme de protection classique des bases de données est fondé sur le droit d'auteur. L'article L. 112-3 du Code de la propriété intellectuelle dispose ainsi que « *les auteurs [...] de recueils d'œuvres ou de données diverses, tels que les bases de données, qui, par le choix ou la disposition des matières, constituent des créations intellectuelles. On entend par base de données un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique, et individuellement accessibles par des moyens électroniques ou par tout autre moyen.* »

C'est ici la structure, la hiérarchisation et l'organisation de la base de données qui, sous réserve de présenter une originalité, sont protégées par le droit d'auteur, protection qui nous semble pouvoir être remise en cause par les développements technologiques et l'avènement des bases de données non structurées dans lesquelles les données ne sont plus organisées dans un agencement strict et rigide[55].

30. Toute base de données n'est cependant que l'accumulation d'informations plus ou moins disparates, elle n'est rien sans la multitude de données individuelles qui la constituent. Au-delà de la protection de la base de données dans son ensemble, il convient donc de s'interroger sur l'appropriation de ce qui fait le *Big Data*, à savoir la « data ».

1.2. Concernant les données contenues dans la base de données

31. Dans le cadre de traitements de données, se pose la question de la propriété de ces dernières., ce qui laisse entendre que les « datas » sont nécessairement l'objet d'un droit de propriété. À ce titre, il semble délicat de pouvoir affirmer l'existence d'un droit de propriété en tant que tel sur les données, du moins en ce qui concerne les données à caractère personnel.

Il n'existe pas en l'état de droit de propriété sur les données reconnu *stricto sensu* au profit de l'entreprise. Les longs errements de la jurisprudence en matière de vol d'information[56] n'en sont qu'une illustration. Il est vrai que la décision récente de la Cour de cassation[57], dans laquelle elle étend la qualification de « vol » au domaine de l'immatériel peut avoir pour conséquence

51. Art. 226-17-1 du Code pénal. Cette sanction ne devrait pas se cumuler avec la sanction pécuniaire que la CNIL peut infliger en application de l'article 47 de la loi informatique et libertés en vertu de principe *non bis in idem*.

52. Art. 31 et 32 de la proposition de règlement.

53. Ces dispositions sont issues de la loi du 1er juillet 1998 relative à la protection des bases de données transposant dans le Code de la propriété intellectuelle la Directive 96/9/CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données.

54. Article 341-1 du Code de la propriété intellectuelle.

55. La protection par le droit d'auteur se trouve désormais déplacée vers les algorithmes mis en œuvre permettant de traiter la donnée, qui peuvent constituer une valeur immatérielle importante et un avantage concurrentiel de premier ordre.

56. Article 341-1 du Code de la propriété intellectuelle.

57. Par un arrêt du 20 mai 2015 (Cass. Crim. 20 mai 2015, 14-81.336) la Cour de cassation a reconnu la qualification de vol s'agissant de la soustraction de données dans le cadre de l'application de l'article 311-1 du Code pénal.

incidente de supposer que la donnée est l'objet d'un droit de propriété. Le projet de directive européenne sur le secret des affaires [58] pourrait sembler s'inscrire dans la même ligne, en reconnaissant la protection des informations entrant dans le champ du secret des affaires, elle paraît implicitement reconnaître un droit de propriété sur ces informations. Cependant la rédaction du texte fait montre d'un exercice délicat d'équilibriste de la part du législateur européen qui prend garde de ne pas recourir à la notion de propriété pour y préférer celle notamment de « *contrôleur* » ou de « *détenteur* ».

32. Si l'entreprise peut difficilement revendiquer un droit sur les données, il en va de même pour les individus auxquels les données se rapportent. Le droit français s'est toujours, du moins pour l'heure, refuser à reconnaître un droit patrimonial sur les données personnelles [59]. Même si on assiste lentement à un glissement vers une reconnaissance d'un droit de propriété de l'individu sur ses données [60], il n'en demeure pas moins que ce droit est inexistant pour le moment.

2. Les données à caractère personnel sont-elles des *res nullius* ?

2.1. Quelle pertinence pour le droit de propriété ?

33. Partant du constat que 70 % [61] des données traitées dans le cadre du *Big Data* sont produites, consciemment ou non, de manière directe ou indirecte par des personnes physiques, se pose la question de savoir si ces données sont des biens dénués de maître. Chacun serait alors libre de se les approprier. Si tel est le cas, n'importe qui pourrait alors faire une exploitation sans limite de ces données à caractère personnel, ce qui ne serait pas sans inconvénients pour les individus concernés [62].

Ainsi posée, la question peut paraître incongrue, tant nous sommes convaincus que nos données à caractère personnel [63], en ce qu'elles nous sont consubstantiellement attachées, sont nôtres et sont donc hors commerce, indisponibles, car produits du corps humain.

34. À y regarder de près, la notion de propriété, en dépit de la représentation que nous nous en faisons intuitivement, n'est pas pertinente. Rappelons que le droit de propriété, tel que défini par l'article 544 du Code civil est le « *[…] droit de jouir et disposer des choses de la manière la plus absolue, pourvu qu'on n'en fasse pas un usage prohibé par les lois ou par les règlements* » [64]. Rappelons également que la propriété se manifeste au travers de trois attributs traditionnels qui sont l'*usus*, le *fructus* et l'*abusus* [65]. Pour n'en retenir qu'un seul, l'*abusus*, est-il concevable que l'on puisse aliéner son identité au travers de la cession à titre onéreux de ses données personnelles telles que celles-ci sont définies par la loi Informatique et Libertés ? Ce mythe de Faust revisité à la « *sauce geek* » ne semble pas réaliste. En résumé, pouvons-nous jouir de nos personnelles de manière absolue au sens de cet article ?

35. Tout d'abord, consacrer une telle vénalité des données personnelles serait en contrariété avec l'article 544 du Code civil lequel, rappelons-le, énonce un droit « *quasi absolu* » puisque limité par le seul fait que l'on « *n'en fasse pas un usage prohibé par les lois ou par les règlements* ». Or, le fait de permettre la cession de données à caractère personnel ne reviendrait-il pas à prêter directement son concours à la commission d'une usurpation d'identité telle que définie par l'article 226-4-1 du Code pénal [66] ? Bien entendu, une convention dont l'objet est la commission d'une infraction, est évidemment nulle pour contrariété à l'ordre public. Deux auteurs ont souligné par ailleurs diverses incongruités auxquelles conduirait la reconnaissance d'un tel droit de propriété [67]. Enfin, cette notion de propriété des données personnelles peut, à certains égards, être néfaste puisqu'elle revient à renvoyer à l'individu « *[…] la responsabilité de gérer et de protéger ses données, au lieu de trouver des réponses collectives à un problème de société […]* » [68].

Par ailleurs, que serions-nous prêts à vendre ? Selon une étude conduite par Havas Media Group [69], si 84 % des personnes interrogées se prétendent inquiètes de l'usage qui peut être fait de leurs données à caractère personnel, 30 % se disent néanmoins prêtes à y don-

58. Proposition de Directive du parlement et du Conseil du 28 novembre 2013 sur la protection des savoir-faire et des informations commerciales non divulgués (secrets d'affaires) contre l'obtention, l'utilisation et la divulgation illicites, COM(2013) 813 final, 2013/0402 (COD).

59. Ce refus a encore été récemment affirmé par le Conseil d'Etat dans son rapport « Le numérique et les droits fondamentaux » : « *Il est parfois proposé de reconnaître aux individus un véritable droit de propriété sur leurs données, en pariant sur leur plus grande implication du fait qu'ils deviendraient financièrement intéressés à une bonne gestion de leurs données. Le Conseil d'Etat ne recommande pas une telle orientation.* » Conseil d'Etat, rapport « Le numérique et les droits fondamentaux », la documentation française, septembre 2014, p. 25.

60. En ce sens, *ibid*., p. 264.

61. S. Grumbach, « *Big Data: The Global Imbalance!* », conférence Lift France 12, 28 septembre 2012. Viviane Reding, vice-présidente de la précédente Commission européenne, estimait que la valeur des données livrées par les citoyens européens en 2011 s'était élevée à environ 315 milliards d'euros. Elle prédisait que cette valeur produite par les données en Europe pourrait représenter 1 000 milliards en 2020, soit 8 % du PIB européen (http://europa.eu/rapid/press-release_SPEECH-13-788_en.htm).

62. Cf. nombre d'ouvrages présentant des dystopies fondées sur une exploitation et un profilage poussé des individus. Cependant l'affirmation d'un droit de propriété sur la donnée n'est pas sans poser d'autres difficultés. Le CNNum, dans son rapport sur la neutralité des plateformes de mai 2014, (spéc. p. 38, http://www.cnnumerique.fr/plateformes/) s'est prononcé contre l'instauration d'un droit de propriété privée sur les données personnelles.

63. Telles que définies par l'article 2 de la loi Informatique et Libertés. Le projet de règlement général y ajoute notamment : les référence « *à un identifiant, par exemple un nom, un numéro d'identification, des données de localisation, ou un identifiant en ligne, ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale* ».

64. On remarquera au passage qu'aucun autre droit n'est affirmé avec autant de force dans le Code civil. Sur les origines de cet article, B. Terrat, « Du régime de la propriété dans le Code civil », p. 329, in *Livre du centenaire du Code Civil*, éditions Dalloz. Cette affirmation avait pour objet tout à la fois de consolider et solder la période révolutionnaire en confirmant les droits des acquéreurs des biens nationaux.

65. *Usus* : droit de détenir et d'utiliser une chose, hors la perception des fruits ; *Fructus* : droit de percevoir les fruits de la chose, lesquels peuvent notamment être obtenus par le commerce ; *Abusus* : droit de disposer de la chose, notamment par la vente.

66. Art. 226-4-1. Le fait d'usurper l'identité d'un tiers ou de faire usage d'une ou plusieurs données de toute nature permettant de l'identifier en vue de troubler sa tranquillité ou celle d'autrui, ou de porter atteinte à son honneur ou à sa considération, est puni d'un an d'emprisonnement et de 15 000 euros d'amende. Cette infraction est punie des mêmes peines lorsqu'elle est commise sur un réseau de communication au public en ligne.

67. F. Mattatai et M. Yaïche, « Être propriétaire de ses données personnelles : peut-on recourir aux régimes traditionnels de propriété ? », Partie I et II, RLDI n° 114, avril 2015, p. 60 et RLDI n° 115, mai 2015, p. 63.

68. V. Peugeot, « Données personnelles : sortir des injonctions contradictoires », Vecam, 13 avril 2014.

69. « Les Français prêts à monnayer leurs données personnelles », publié le 26 septembre 2014 sur le figaro.fr.

ner accès pour 500 euros par an, sans que l'on sache au juste quelles seraient les données concernées. Ces personnes sont-elles prêtes à vendre leurs données de géolocalisation contenues leurs GPS ou bien leurs données bancaires, de santé ou bien encore celles portant sur leurs convictions religieuses[70] ?

36. Le Conseil national du numérique, dans son avis de mai 2014, a très bien résumé les principales failles de cette approche[71] : « *elle renvoie à l'individu la responsabilité de gérer et protéger ses données* » ; elle « *renforce l'individualisme et nie le rapport de force entre consommateurs et entreprises* » ; « *elle ne pourrait que générer des revenus anecdotiques pour les usagers et susciter à l'inverse un marché de la gestion protectrice des données numériques* ». En définitive, « *elle déboucherait sur un renforcement des inégalités entre citoyens en capacité de gérer, protéger et monétiser leurs données et ceux qui, par manque de littéraire, de temps, d'argent ou autre, abandonneraient ces fonctions au marché* ».

Le Conseil d'État, dans son rapport pour l'année 2014, a lui aussi rappelé ces critiques : « *Même si le prix des données de chaque individu est appelé à croître de manière considérable au cours des années à venir (jusqu'à quelques euros), la valeur de l'actif que la reconnaissance du droit de propriété conférerait à chaque individu restera dérisoire* » ; *par ailleurs*, « *les acteurs du numérique rédigeraient leurs contrats comme la fourniture d'un service en échange de la cession de droits d'utilisation des données, ce dont nombre de conditions générales d'utilisation se rapprochent déjà beaucoup ; le rapport de forces entre l'individu, consommateur isolé et l'entreprise, resterait marqué par un déséquilibre structurel* »[72].

37. Une telle situation ne paraît pas souhaitable, elle ne reflète d'ailleurs pas la réalité du cadre juridique applicable au traitement de données personnelles. Si le *Big Data* est un géant aux pieds d'argile ce n'est pas en raison de l'existence ou de l'absence d'un droit de propriété sur les données, mais plutôt des droits qui sont reconnus aux personnes par la législation sur les données personnelles.

2.2. Quelles alternatives ?

38. Les droits reconnus par la loi Informatique et Libertés permettent aux individus de s'opposer au traitement de leurs données[73]. Cette opposition ne nécessite aucune motivation lorsque les données sont utilisées à des fins de prospection commerciale[74]. Les personnes peuvent donc interdire à l'entreprise de traiter leurs données, si ce droit vient à être exercé massivement, c'est la légitimité même du *Big Data* qui peut être remise en cause. Les entreprises ne sont en réalité que les détenteurs à titre précaire des données sur lesquelles elles ne peuvent revendiquer de droits propres.

Cette précarité est appelée à s'accroître dans la mesure où les droits accordés aux personnes sur leurs données devraient prochainement être étendus tant au niveau européen, au travers de l'adoption du règlement européen sur la protection des données, qu'au niveau national, par le biais du projet de loi numérique[75].

Le projet de règlement européen sur la protection des données contient plusieurs dispositions qui sont susceptibles d'être perçues comme une entrave au *Big Data*. Parmi ces mesures on peut évoquer notamment la consécration du droit à l'oubli[76] ou encore l'introduction du concept de portabilité des données[77] permettant aux utilisateurs d'un service de demander à ce que leurs données soient transférées vers un autre prestataire.

39. En France, le projet de loi Numérique, souhaite reprendre au plan national des dispositions issues de la proposition de règlement en discussion à Bruxelles[78]. Il en va ainsi du droit à la portabilité des données dont le contenu est au passage étendu et étoffé. Le texte français innove en introduisant dans la loi Informatique et Libertés le concept d'autodétermination informationnelle, dégagé par la Cour constitutionnelle allemande dès 1983[79] et prôné par le Conseil d'État[80]. L'article 26 du projet de loi pour une République numérique vient ajouter la disposition suivante à l'article 1 de la loi de 1978 : « *Toute personne dispose du droit de décider et de contrôler les usages qui sont faits des données à caractère personnel la concernant, dans les conditions fixées la présente loi.* » Le sujet majeur demeure celui de l'éduction numérique. Comment promouvoir cette autodétermination lorsque des personnes sont prêtes à faire commerce de leurs données personnelles moyennant finance ?

40. Le Conseil d'État[81] résume l'état de la question de façon pertinente : « *Face aux limites du cadre actuel de la protection des données à caractère personnel, il est parfois proposé de donner aux individus un véritable droit de propriété sur leurs données ; le but recherché est notamment de susciter une implication plus active, les individus devenant financièrement intéressés à une bonne gestion de leurs données. Le Conseil d'État ne recommande pas d'emprunter cette voie en dépit de son attrait apparent. S'il convient en effet de*

70. Article 8 de la loi informatiques et libertés : sont sensibles les informations concernant l'origine raciale ou ethnique, les opinions politiques, philosophiques ou religieuses, l'appartenance syndicale, la santé ou la vie sexuelle.

71. CNN, avis 2014-2 sur la neutralité des plates-formes : réunir les conditions d'un environnement numérique ouvert et soutenable, p. 37. V. également V. Peugeot, « Données personnelles : sortir des injonctions contradictoires », Vecam, juillet 2014.

72. CE, Rapport préc., pp. 265-266.

73. Art. 38 de la loi Informatique et Libertés.

74. Le projet de règlement européen sur la protection des données dans son article 19, précisant que les individus n'ont pas à justifier d'un motif légitime dans l'exercice de leur droit d'opposition. Il appartiendra au responsable de traitement de justifier d'un intérêt légitime pour continuer à traiter ces données malgré l'opposition formulée par la personne concernée.

75. Projet de loi pour une République numérique.

76. Les dispositions concernant le droit à l'oubli sont contenues dans l'article 17 du projet de règlement. Elles viennent consacrer la jurisprudence de la CJUE issue de l'arrêt du 13 mai 2014, Google Spain SL et Google Inc. c/ Agencia Española de Protección de Datos (AEPD) et Mario Costeja González.

77. Art. 18 du projet de règlement.

78. On peut légitimement s'interroger sur la pertinence d'une telle démarche dans la mesure où le texte européen n'est pas fixé et que, s'agissant d'un règlement, il ne nécessitera pas de transposition au niveau national.

79. Décision du 15 décembre 1983 au sujet d'une loi sur le recensement. V. Y. Poullet et A. Rouvroy, « Le droit à l'autodétermination informationnelle et la valeur du développement personnel. Une réévaluation de l'importance de la vie privée pour la démocratie », *in* K. Benyekhlef & P. Trudel (dir.), *État de droit et virtualité*, Montréal : Thémis, 2009.

80. Conseil d'État, rapport « Le numérique et les droits fondamentaux », précité, p. 267.

81. Rapport annuel 2014, « Le numérique et les droits fondamentaux » : http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/144000541.pdf.

renforcer la dimension de l'individu acteur dans le droit à la protection des données, c'est en envisageant celui-ci comme un droit à l'autodétermination plutôt que comme un droit de propriété [...]. En l'état du droit, il n'existe pas de droit de propriété de l'individu sur ses données personnelles. »

La « *marchandisation* » de l'humain au travers de la vente de ses données personnelles en dit long sur l'état d'une société dans laquelle tout, ou presque, se vend et s'achète. Ce qui est certain, c'est que le droit de propriété, tel qu'envisagé en 1804 par ses rédacteurs, ne répond pas aux difficultés que soulèvent les données personnelles. La nécessité d'une pédagogie du numérique est plus que jamais nécessaire.

41. Tout ceci pose avec intensité la question de la place qu'occupe la défense des données à caractère personnel dans les dispositifs législatifs européens et nationaux, lesquels manifestent, parfois dans leur intitulé[82], une tension évidente entre cette protection et la promotion d'un marché européen de la donnée. *In fine*, la protection des données personnelles est-elle une fin en soi ou bien, n'est-elle qu'une contingence au service d'un objectif « *transcendant* » qui résiderait dans une concurrence équitable et efficace entre les acteurs économiques grâce à une liberté de traitement des données ? La CNIL elle-même nous invite à ce questionnement en énonçant que la protection des données apparaît « *plus comme une condition du développement économique que comme une réponse aux menaces d'un fichage accru et mieux outillé fait peser sur les autonomies individuelles* »[83].

42. Enfin, s'il convient de se préoccuper de ceux dont les données à caractère personnel transiteront dans le *Big Data*, il faut aussi avoir une pensée pour les personnes touchées par l'« illectronisme »[84]. Ces personnes qui en raison de l'âge, du manque de revenu ou de l'éloignement géographique n'ont que peu ou pas accès aux technologies de l'information risquent de représenter demain une population marginalisée. Peut-être faudra-t-il y ajouter demain ceux qui, afin de préserver leur vie privée, choisiront d'être débranchés et de vivre une existence de « *marginaux numériques* ». Shashi Tharoor, sous-secrétaire général des Nations unies pour les communications et l'information publique, déclarait en 2004 : « *Une chose est sûre : dollar et PNB ne sont plus les seuls éléments qui séparent les nantis des démunis. La révolution industrielle fait partie du passé ; nous vivons l'ère de la révolution de l'information.* » Affirmation ponctuée par le constat qu'il s'agissait là d'une « *révolution avec beaucoup de liberté, un peu de fraternité et aucune égalité* ». ■

82. Exemple flagrant : la directive 95/46/CE du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

83. 11e rapport d'activité de la CNIL pour l'année 1990, p. 48. Sur cette question, on lira avec intérêt la remarquable thèse de N. Ochoa, « Le droit des données personnelles, une police administrative spéciale », soutenue le 8 décembre 2014.

84. Rapport 2011 du CREDOC relatif à « *la diffusion des technologies de l'information et de la communication dans la société française* ».