

Vers une géopolitique de la donnée

Par **Thierry BERTHIER**

Maître de conférences en mathématiques, chercheur à la Chaire de Cyberdéfense & Cybersécurité Saint-Cyr-Thales-Sogeti

et **Olivier KEMPF**

Docteur en science politique et chercheur associé à l'IRIS

La donnée est présentée par certains comme une ressource, quand d'autres voient en elle le pré-lude à une nouvelle monnaie. Alors que l'Internet des objets qui s'annonce prévoit des volumes de données démultipliés par rapport à ceux que nous connaissons aujourd'hui, nul ne doute que la donnée sera demain l'élément central de nos sociétés, que ce soit sur le plan économique ou sur le plan politique. Nous nous proposons ici de décrire succinctement ce phénomène pour montrer qu'il est principalement le fait d'acteurs privés, les acteurs publics (en France et en Europe, en particulier) semblant être largement impuissants face à cette nouvelle dynamique. Si la donnée devient un instrument central de la puissance, il nous manque encore une grille d'analyse pour pouvoir mesurer la puissance data-numérique d'une nation.

La data-économie en quelques chiffres

Parfois considérées comme le nouveau pétrole, les données numériques modifient notre perception du monde en nous donnant accès à toute sa complexité. Elles alimentent désormais tous les moteurs d'une économie en mutation qui est devenue en quelques décennies une économie de la connaissance. Estimée en 2013 à plus de 15 000 milliards d'euros [1], la part des produits et services numériques dans l'activité mondiale représente aujourd'hui près d'un sixième du total de l'économie des biens et services traditionnels (elle était évaluée en 2013 à 100 000 milliards d'euros) [2]. En mai 2015, la valorisation boursière cumulée des 15 premiers groupes mondiaux du digital a atteint les 2 400 milliards de dollars, ce qui correspond au PIB de la France en 2015 [3]. On notera que cette valorisation du Top 15 digital cumulé ne s'élevait en 1995 qu'à 16,75 milliards de dollars.

Quel que soit leur cœur de métier, les entreprises industrielles organisent leurs activités autour de systèmes d'information qui conditionnent leur stratégie, leur prospérité et leur pérennité. Bâties sur des modèles entièrement nouveaux, les villes intelligentes optimisent leurs ressources et leurs consommations énergétiques grâce à une intégration, à toutes les échelles, de la donnée et de son traitement. Leurs architectures conjuguent les matériaux et les données numériques pour offrir des espaces d'un genre nouveau où l'information devient ubiquitaire, c'est-à-dire accessible partout, pour tous, et tout le temps.

Avec plus de 1,01 milliard de capteurs et d'objets connectés en 2013 et près de cent milliards de ces objets qui le seront sans doute en 2020, c'est bien à un déluge de

Rang mondial (mai 2015)	Société	Origine	Capitalisation boursière (2015) en milliards de dollars
1	Apple	États-Unis	763,567
2	Google	États-Unis	373,437
3	Alibaba	Chine	232,755
4	Facebook	États-Unis	226,009
5	Amazon.com	États-Unis	199,139
6	Tencent	Chine	190,110
7	eBay	États-Unis	72,549
8	Baidu	Chine	71,581
9	Priceline.com	États-Unis	62,645
10	Salesforce.com	États-Unis	49,173
11	JD.com	Chine	47,711
12	Yahoo!	États-Unis	40,808
13	Netflix	États-Unis	37,700
14	LinkedIn	États-Unis	24,718
15	Twitter	États-Unis	23,965
Total du Top 15			2 415,867

Tableau 1 : Classement par capitalisation boursière décroissante des grands acteurs mondiaux de l'Internet (famille des « GAFA » élargie) en 2015.

données que l'on a affaire et qui engendre de la richesse, des besoins et de nouveaux métiers. En 2015, l'humanité a produit, en une seule minute, 200 millions de méls, 15 millions de SMS, 350 000 tweets, 250 gigaoctets de données sur Facebook et plus de 1 740 000 gigaoctets d'informations numériques au niveau mondial. Google a traité quotidiennement plus de 24 pétaoctets de données (soit 24 millions de milliards d'octets).

Le schéma de la Figure 1 de la page suivante donne une idée de l'échelle des volumes usuels de données.

Cette production, qui augmente aujourd'hui de manière exponentielle, n'est que partiellement exploitée, faute d'infrastructures algorithmiques adaptées. Plus que le



Figure 1 : Les différents supports (naturels et numériques) de conservation de données et leurs capacités de stockage (mesurées en puissances 10 d'octets : Mo = mégaoctet, Go = gigaoctet, To = téraoctet, Po = pétaoctet, Eo = exaoctet).

problème du stockage, c'est bien celui du traitement qui doit mobiliser les industries de la donnée et les recherches dans les data sciences.

Ces quelques chiffres montrent toute la puissance d'impact des données numériques sur l'économie mondiale. Ils ne doivent pas pour autant masquer le fait que cet impact dépasse très largement le seul cadre économique pour investir les champs géopolitique et stratégique. En tant que ressource virtuelle, la donnée donne des pouvoirs à celui qui sait la collecter et l'exploiter. Elle devient ainsi un enjeu prioritaire de puissance pour les nations.

Le pouvoir politique n'a que peu de... pouvoir sur la donnée

Les données sont devenues un enjeu évident de puissance. Celle-ci est d'abord commerciale, puisque de grands acteurs en tirent profit au travers des dispositifs de traitement des *Big Data* (données massives). Il s'agit, à partir du recueil puis de l'analyse statistique de grands volumes de données, de dresser le profil des consommateurs et de déterminer leurs centres d'intérêt. Ce profilage revêt une valeur marchande évidente, puisqu'il permet d'affiner et d'affûter les campagnes publicitaires.

C'est ainsi que les grands groupes qui disposent de données massives peuvent, après traitement, vendre des « profils de consommateur » à d'autres acteurs économiques afin que ceux-ci puissent optimiser leurs dépenses publicitaires en augmentant leurs ventes.

Les sociétés les plus avancées dans ce domaine sont les *GAF A* (Google, Amazon, Facebook et Apple), auxquelles on ajoute de plus en plus Microsoft.

Mais d'autres acteurs ont commencé à mettre en œuvre ce type de démarche, qu'il s'agisse d'acteurs numériques comme les éditeurs de navigateurs, les opérateurs téléphoniques, les fournisseurs d'accès à Internet ou encore de grands sites marchands (SNCF, Le Bon Coin, les grands groupes de médias).

Du côté des administrations, l'enjeu est bien différent, puisqu'il s'agit de mettre à la disposition du public des données ouvertes, ce que l'on désigne sous l'expression d'« *open data* ».

Constatons, toutefois, que les grands services de renseignement (la NSA états-unienne, le GCHQ britannique, la DGSE française) utilisent les mêmes techniques pour observer les différents trafics et rechercher des individus potentiellement dangereux : le *Big Data* est devenu à ce titre une des composantes essentielles de l'espionnage.

La maîtrise de la donnée est aujourd'hui principalement le fait d'acteurs privés. Il n'existe pratiquement aucune régulation publique. On assiste tout au plus aux premiers conflits d'intérêt entre une logique privée et une logique étatique.

À cet égard, l'affaire qui a opposé le *Federal Bureau of Investigation* (FBI) à Apple est révélatrice. L'un des terroristes ayant mené fin 2015 l'attaque de San Bernardino, en Californie, avait utilisé un *iPhone*. Le FBI souhaitait accéder

aux données résidentes de ce téléphone, mais ces dernières étaient chiffrées selon une technologie mise en place par Apple. Cette société refusa de coopérer avec l'agence américaine pour deux raisons : l'une de principe, consistant à rappeler que le service qu'elle apporte à ses clients repose sur la confidentialité absolue de leurs échanges, laquelle est rendue possible par un chiffrement intégré ; et une autre raison, plus technique, tenant à la nécessité de mobiliser de nombreux ingénieurs pour réussir à débloquent ledit téléphone, et donc un coût induit important. Au fond, Apple refusait de créer un précédent au nom des libertés publiques (et accessoirement, en usant d'un argument *marketing* habile, en se posant en défenseur du citoyen et de ses données, face à un « *Big Brother* » étatique). Apple reçut d'ailleurs le soutien unanime des grands acteurs Internet américains (Whatsapp annonça, quelques semaines plus tard, qu'il allait mettre en place un chiffrement de « bout en bout » de son service). La suite est connue : le FBI recourut aux services de *hackers* professionnels pour « casser » le code d'Apple et accéder ainsi aux données tant convoitées.

Cette affaire met en relief une double dimension de la donnée : celle-ci peut être massive, mais elle peut également être individuelle. Les États ne disposent que de moyens indirects pour y accéder soit par l'espionnage technique, soit au moyen d'intrusions ciblées pas toujours faciles à orchestrer. Enfin, la donnée révèle l'opposition entre deux types de puissance, les puissances technico-commerciales et les puissances étatiques. Or, ces dernières n'ont pas encore la maîtrise de la donnée, alors que les premières ont acquis leur force et leur pouvoir grâce à une exploitation habile de celle-ci.

Une autre dimension apparaît, mais de manière sous-jacente : celle de la protection de la donnée. En effet, les autorités ne cessent d'alerter sur les risques de la cybercriminalité exhortant les internautes à observer une hygiène informatique afin de baisser le niveau de leur exposition au risque de piratage. Une des méthodes possibles consiste à procéder au chiffrement de ses données et de ses échanges, ce qui suscite dans le même temps l'ire des autorités, puisque ces techniques peuvent également être utilisées par des criminels et des terroristes. Dès lors, celles-ci prônent la limitation des techniques de chiffrement. Les fournisseurs de services ont beau jeu de déclarer que ce n'est pas parce que les criminels utilisent des voitures qu'il faut interdire celles-ci, et qu'il devrait en aller de même du chiffrement.

Ainsi, les autorités publiques (notamment américaines, puisque les États-Unis sont souvent à la pointe de ces débats) font montre d'une schizophrénie certaine, qui témoigne de leur insatisfaction résultant de leur incapacité de maîtriser un domaine nouveau et en expansion qui apparaît à tous comme la ressource sur laquelle repose aujourd'hui la puissance.

Un retard de la France et de l'Europe

L'essentiel des débats se situe aujourd'hui aux États-Unis, qui sont à la pointe des évolutions technologiques, mais aussi économiques du cyberspace. La France et l'Eu-

rope paraissent absentes, du fait qu'elles font appel très souvent à des mécanismes intellectuels inadaptés.

En France, pays de tradition colbertiste, associer entre elles les notions de donnée et de pouvoir revient souvent à évoquer le concept de souverainisme économique. On lance de grands projets qui n'arrivent pas à convaincre le public ou, plus exactement, le consommateur. Or, une des clefs du succès réside dans la compréhension du fait que la puissance doit d'abord être économique avant de pouvoir se transformer en puissance politique. Pourtant, quelques initiatives ont pu donner des résultats, à l'instar du Minitel qui, au cours des années 1980, fit de la France l'un des pays les plus connectés au monde. Le refus de faire évoluer la technologie pour en garder la maîtrise industrielle, tout comme le refus d'en ouvrir le code source pour contrôler la diffusion de la norme ont conduit, parmi d'autres raisons, à la fin de l'aventure Minitel.

Plus récemment, les autorités ont lancé un *cloud* souverain, qui s'est perdu dans les sables mouvants technologiques. Les seuls acteurs français du *cloud* sont aujourd'hui privés, qu'il s'agisse d'OVH ou de Gandi. De même, on a promu un moteur de recherche, Qwant, qui peine à convaincre. La seule innovation reste l'antivirus Uhuru, dont la diffusion demeure confidentielle malgré d'évidentes qualités. En fait, le succès passe par l'usage, lequel doit répondre aux attentes.

Quant à l'Europe, elle ne constitue pas non plus un échelon efficace, malgré son cadre qui devrait favoriser des solutions de masse : avec une population de 450 millions d'habitants riches, éduqués et connectés, et malgré ses institutions (dont le premier métier consiste à dire le droit et la norme), elle devrait susciter l'émergence d'acteurs forts. Or, il n'en est rien pour l'instant. Tout juste peut-on évoquer le projet Junker, qui entend favoriser des développements économiques, ou encore le récent vote par le Parlement européen d'un texte qui, adoptant le PNR (*Passenger Name Record*), y joint un texte sur la protection des données des utilisateurs. La Commission vient d'annoncer son intention de développer un *cloud* souverain scientifique, appelé à être la première pierre d'un *cloud* européen.

Pourtant, ni la France ni l'Europe n'acquerront de puissance sur la donnée si elles ne se dotent pas de grands acteurs faisant référence dans ce domaine.

Ces brefs constats dressent un tableau contrasté. Toutefois, cela reste insuffisant, car nous manquons d'une connaissance comparative plus fine des déterminants de la puissance data-numérique. Nous examinerons ci-après certains critères qu'il faudrait appréhender pour ce faire.

Sur la puissance data-numérique d'une nation

Nous nous interrogeons à présent sur ce qui fonde la puissance data-numérique d'une nation. De natures hétérogènes, les facteurs déterminant cette puissance demeurent dans leur ensemble interdépendants. Ils sont caractérisés par les niveaux de développement de cinq

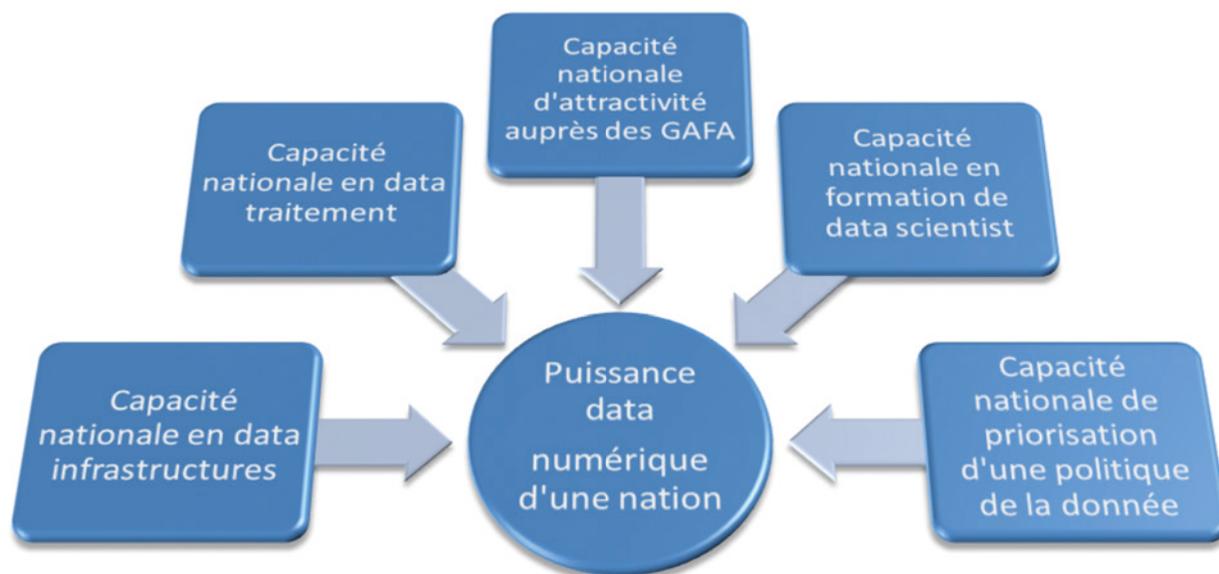


Figure 2 : Les différents facteurs constitutifs de la puissance data-numérique d'une nation.

grandes capacités que la nation considérée a su ou non favoriser au cours des deux dernières décennies : sa capacité en termes de data-infrastructures, sa capacité de data-traitement, sa capacité d'attractivité vis-à-vis des grands acteurs internationaux de la donnée, sa capacité à former des scientifiques de la donnée et, enfin, sa capacité à faire d'une politique numérique nationale ambitieuse une priorité.

La capacité nationale en termes de data-infrastructures

Elle se mesure par la densité et la puissance des infrastructures physiques de stockage et de traitement de la donnée implantées sur le territoire. Son évaluation tient compte du nombre de *datacenters* (par habitant) présents sur le territoire, des capacités maximales de stockage et de traitement de ces *datacenters* (par année de fonctionnement), de leur consommation énergétique annuelle totale, du nombre et du classement (en fonction de leur puissance de calcul) des supercalculateurs (HPC) installés sur le territoire national et de l'activité annuelle (en volume de données traitées) des grands opérateurs (téléphonie, Internet et réseaux) opérant localement.

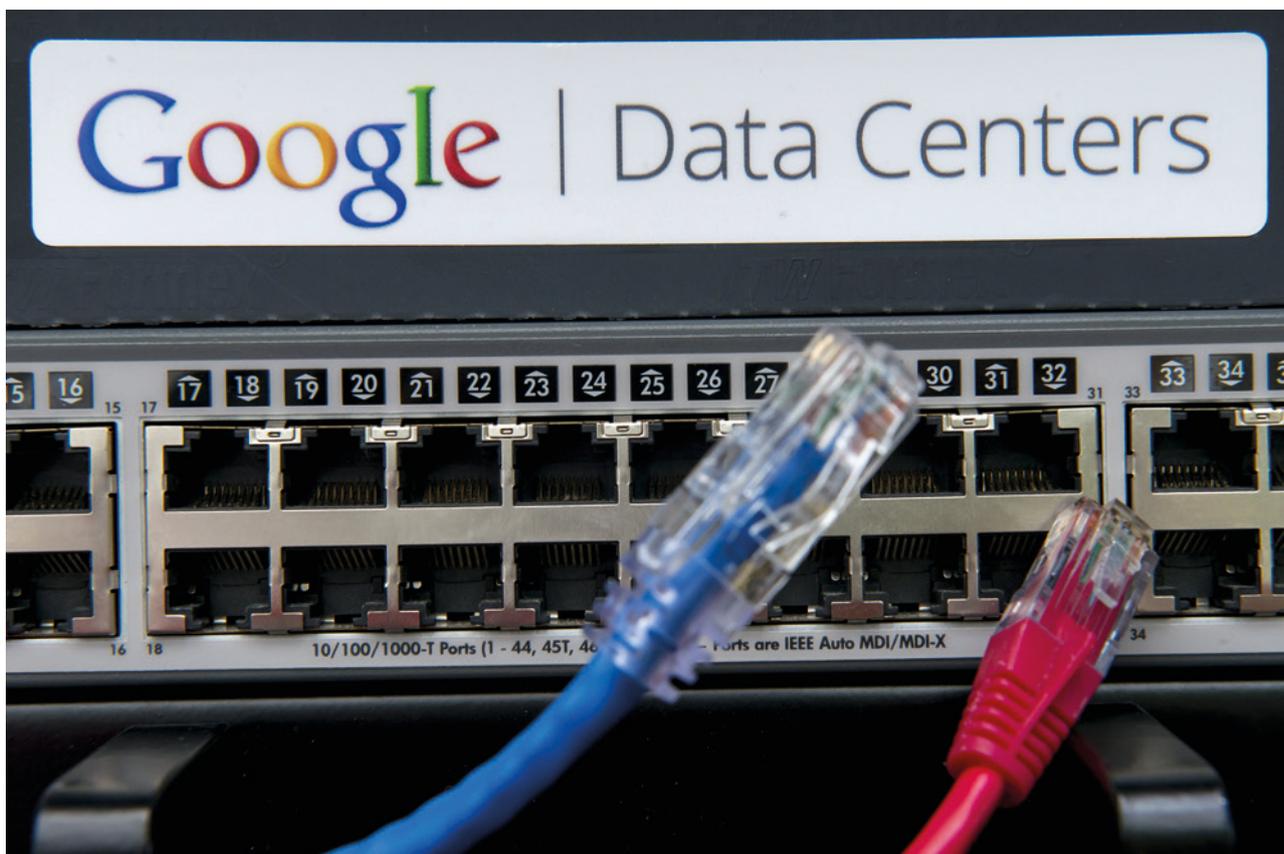
La capacité nationale en termes de data-traitement

Elle reflète le niveau opérationnel des outils de traitement algorithmique de la donnée développés localement et le degré de dépendance de la nation considérée vis-à-vis des infrastructures étrangères (notamment américaines). Cette capacité tient compte de la présence ou du développement d'un *cloud* souverain, sécurisé et pérenne, et de la présence ou du développement d'un moteur de recherche national qui soit capable de concurrencer les grands moteurs de recherche commerciaux (tels que Google, Yahoo, Yandex, Baidu...). Elle tient compte du nombre de laboratoires, de *clusters* de R&D et de chaires univer-

sitaires de recherche en sciences des données présents sur le territoire national, du nombre annuel de brevets déposés en sciences des données par habitant, du nombre annuel de thèses de doctorat soutenues en sciences des données par habitant, du nombre annuel de sociétés (*start-ups* ou grands groupes) classées ou récompensées lors des concours et challenges internationaux de data-innovation. Enfin, elle tient compte du nombre des *start-ups* data-centrées rachetées par de grands acteurs internationaux (les GAFA : Google, Apple, Facebook, Amazon) durant l'année et du nombre des partenariats liant des sociétés locales (*start-ups* ou grands groupes) aux grands acteurs de la donnée (GAFA).

La capacité nationale d'attractivité vis-à-vis des grands acteurs internationaux de la donnée

Elle résulte du niveau d'attractivité que l'écosystème numérique national exerce sur les acteurs mondiaux de la donnée, en particulier auprès des GAFA et du Top15 numérique. Son évaluation s'appuie sur le degré d'implantation des acteurs mondiaux de la donnée sur le territoire national sous la forme de centres de recherche, de *datacenters*, de terminaux commerciaux et de structures de formation. Elle tient compte de l'existence de partenariats de R&D noués annuellement par ces grands acteurs avec des laboratoires ou avec des *clusters* nationaux, du volume annuel des rachats, par ces grands groupes, de *start-ups* « incubées » sur le territoire national et du niveau des participations de ces grands acteurs de la donnée dans des projets nationaux innovants. Elle tient compte également du degré de compatibilité de l'environnement juridique national avec l'activité industrielle des grands acteurs de la donnée et du niveau d'acceptabilité du régime fiscal local pour ces groupes. Enfin, cette capacité tient compte de la perception (positive ou négative) par ces grands acteurs de la donnée des politiques numériques mises en place par le gouvernement du pays



Inauguration par Google de son second centre de traitement des données installé à Baudour (Belgique), juin 2015.

« La capacité d'attractivité d'un pays résulte de celle que l'écosystème numérique national exerce sur les acteurs mondiaux de la donnée, en particulier auprès des GAFAs et du Top15 du numérique. Son évaluation s'appuie sur le degré d'implantation des acteurs mondiaux de la donnée sur le territoire national sous la forme de centres de recherche, de *datacenters*, de terminaux commerciaux et de structures de formation. »

concerné. Pour résumer : ces politiques numériques nationales peuvent-elles les inciter à investir et à s'investir durablement sur le territoire concerné ?

La capacité nationale de formation de scientifiques des données

Elle s'appuie sur la distribution nationale des formations de l'enseignement supérieur en sciences des données (statistiques – *machine learning* – *Big Data*), sur le degré de rayonnement international de ces formations, sur le niveau de sélectivité du recrutement des étudiants suivant ces formations, sur le nombre de partenariats industriels associés à ces formations et sur le taux de création d'entreprises par les diplômés pendant ou à l'issue de leur formation.

La capacité nationale à « prioriser » une politique numérique ambitieuse

Elle est essentiellement fondée sur le niveau de priorité que le gouvernement accorde à sa politique numérique et sur les mesures qu'il met en place pour favoriser le développement d'un écosystème construit sur la donnée. La part du PIB national réinjectée dans la recherche et dé-

veloppement du secteur data-numérique constitue certainement l'agrégat le plus pertinent pour mesurer ce niveau de priorité.

L'approche quantitative de ces cinq grandes capacités passe nécessairement par la définition d'indicateurs numériques formels pour chacun des facteurs listés. Alors qu'elle influence largement les quatre premières, la cinquième capacité semble toutefois la moins facile à évaluer précisément en raison de la forte volatilité du concept de « priorité » pour tout gouvernement en activité.

Pour conclure...

Engagées dans une compétition mondialisée pour la maîtrise de la donnée, la France et l'Europe disposent sur leurs territoires de pôles d'excellence technologique, scientifique, académique et industrielle de très haut niveau. La coordination et la mutualisation de leurs efforts leur permettront de concurrencer efficacement les grands acteurs américains et chinois sur le marché de la collecte et de l'exploitation des données. Il faut que cette coopération européenne se mette en place au plus vite, si les pays membres de l'Union souhaitent réellement maintenir

ou restaurer une part de leur souveraineté nationale en augmentant leur puissance data-numérique. La donnée donne le pouvoir : sachons la conquérir !

Bibliographie

[1] « L'économie des données personnelles : les enjeux d'un business éthique », Rapport CIGREF, 2015.
<http://www.cigref.fr/wp/wp-content/uploads/2015/11/CIGREF-Economie-donnees-perso-Enjeux-business-ethique-2015.pdf>

[2] PÉRES (Éric) – Rapporteur, « Les Données numériques : un enjeu d'éducation et de citoyenneté », Les Avis du Conseil économique, social et environnemental, janvier 2015.

http://www.lecese.fr/sites/default/files/pdf/Avis/2015/2015_01_donnees_numeriques.pdf

[3] KPCB Internet Trends 2015 – Code Conference, Top15 Digital.