

L'analyse des mégadonnées : un nouveau domaine à explorer

Nii Ayi Armah, Analyses de l'économie canadienne

- L'analyse conjoncturelle est extrêmement tributaire des données. Plus celles-ci sont à jour, exactes et pertinentes, meilleure sera l'évaluation de la situation de l'activité économique au moment présent.
- Grâce aux progrès techniques, il est possible d'exploiter les données numériques provenant des transactions commerciales, des médias sociaux et des ordinateurs connectés. Prises ensemble, ces données sont désignées par le terme « mégadonnées ».
- L'analyse des énormes quantités d'informations numériques que renferment les mégadonnées peut apporter un nouvel éclairage au suivi de l'évolution de l'économie et de l'inflation. De plus, la pertinence temporelle des mégadonnées pourrait améliorer la prise de décisions en temps réel dans le cadre de la politique monétaire.
- Les possibilités qu'offrent les mégadonnées sont toutefois limitées par des défis liés à des contraintes méthodologiques, un accès difficile et les préoccupations entourant la protection de la vie privée.

En fournissant une estimation de l'état de l'économie au moment présent, l'analyse conjoncturelle¹ contribue à l'établissement des projections macroéconomiques à long terme de la Banque du Canada, lesquelles éclairent à leur tour les décisions de politique monétaire. Il serait avantageux aux fins de l'analyse conjoncturelle de disposer d'une information complète et immédiatement accessible à propos de chaque transaction économique ou financière effectuée à l'intérieur d'un pays, car cette information faciliterait le calcul d'une mesure exacte et à jour des grands indicateurs macroéconomiques. Malheureusement, un tel ensemble de données idéal n'existe pas. Les données macroéconomiques officielles que préparent les organismes statistiques sont publiées avec un temps de retard et font l'objet de révisions. Ainsi, le produit intérieur brut (PIB), un jeu de données trimestrielles, est rendu public après un délai de deux mois et est révisé pendant les quatre années qui suivent; l'indice des prix à la consommation (IPC), une série mensuelle, n'est quant à lui pas révisé, mais est publié trois semaines après la fin du mois sur lequel il porte.

¹ Voir, dans la présente livraison, Coletti et Kozicki pour obtenir des précisions sur le rôle de l'analyse conjoncturelle à l'appui de la politique monétaire.

Devant ces insuffisances, certains chercheurs ont envisagé de suppléer aux données officielles par des données « officieuses » susceptibles d'être plus actuelles². Un premier essai d'utilisation de ce type de données dans le cadre de l'analyse conjoncturelle a été effectué par Lamont (1997), qui a constaté que le nombre d'occurrences du mot *shortage* (pénurie) dans la presse écrite pouvait être un bon indice pour prévoir l'inflation aux États-Unis. La Banque du Canada se sert aussi de données non officielles pour surveiller l'évolution de l'économie. Par exemple, ses bureaux régionaux recueillent et analysent des renseignements obtenus lors de sondages menés chaque trimestre auprès d'entreprises de différentes régions du pays pour connaître leurs points de vue au sujet de la demande, des pressions s'exerçant sur la capacité de production ainsi que de l'activité économique à venir. Les résultats de ces consultations, qui sont résumés dans *l'Enquête sur les perspectives des entreprises*, fournissent une source d'informations à jour qui enrichit les perspectives tirées des statistiques officielles. À la faveur des progrès techniques, de la prolifération des données numériques et de la diminution des coûts de l'entreposage électronique, un autre type de données non officielles a récemment fait son apparition et prend rapidement de l'expansion : les mégadonnées.

Les mégadonnées sont d'énormes ensembles de données, dont certaines, notamment sur les transactions commerciales, sont consignées depuis longtemps dans les livres de comptes des entreprises sous forme de chiffres de ventes quotidiens ou de volumes des stocks. Depuis un certain temps, les organismes gouvernementaux conservent également une mine de microdonnées administratives. Le coût élevé de l'extraction et de l'organisation de toute cette information a freiné l'exploitation de ces données complémentaires aux fins de l'analyse conjoncturelle. Cependant, grâce à leur numérisation, des sources qui n'existaient autrefois que sur support papier sont maintenant beaucoup plus accessibles et plus faciles à organiser et à analyser. Les données associées aux services d'institutions publiques ou d'organismes d'État offrent une abondance d'informations sur le comportement des citoyens. Par ailleurs, le développement rapide des réseaux informatiques et d'Internet a conduit à l'avènement de nouvelles sources de données comme les médias sociaux et les requêtes sur le Web, ainsi que les paiements électroniques effectués par carte de crédit ou de débit. Du fait de leur omniprésence et de la vitesse à laquelle elles peuvent être collectées, les mégadonnées sont susceptibles de fournir une information plus à jour et détaillée sur les transactions commerciales et financières. Elles pourraient donc être une ressource non officielle supplémentaire et laissent entrevoir des possibilités inexplorées pour faire progresser l'analyse conjoncturelle. Parce qu'elles permettent d'exploiter des volumes gigantesques d'informations numérisées, les mégadonnées offrent de nouvelles connaissances qui présentent un intérêt pour le suivi de l'activité économique et de l'inflation. Leur actualité pourrait enrichir les données officielles afin d'améliorer la prise de décision en temps réel dans le domaine de l'analyse conjoncturelle, en plus de contribuer à l'établissement des statistiques officielles (Daas et van der Loo, 2013).

Le texte qui suit passe en revue les principales caractéristiques des mégadonnées puis présente leurs applications et avantages possibles à l'appui de l'analyse conjoncturelle. Il aborde ensuite les défis qui empêchent de tirer le meilleur parti de ces grands volumes de données, ainsi que les initiatives visant à régler ces difficultés. En conclusion, il donne un aperçu des usages potentiels des mégadonnées aux fins de l'analyse conjoncturelle dans l'avenir.

² En ce qui concerne les données officielles, un compromis doit être fait entre leur publication en temps opportun et leur exactitude.

Les mégadonnées

Le terme « mégadonnées » désigne de grands ensembles hétérogènes de données numériques, issues des transactions commerciales et des échanges dans les médias sociaux effectués par des milliards de personnes dans le monde.

Les quatre V : volume, variété, vélocité, valeur

Les mégadonnées sont caractérisées par quatre grandes variables : le volume, la variété, la vélocité et la valeur. Leur **volume** est généralement beaucoup plus grand que celui des jeux de données classiques. Manyika et autres (2011) expliquent que, du fait de leur taille, ces masses de données dépassent les capacités de saisie, de stockage, de gestion et d'analyse des outils logiciels habituels des bases de données. L'Encadré 1 donne une idée de l'ampleur des mégadonnées.

Les différents types d'informations qui constituent les mégadonnées proviennent de sources **variées**. Environ 10 % seulement sont des données structurées (Gens, 2011), c'est-à-dire qui s'inscrivent parfaitement dans les lignes et colonnes des bases de données relationnelles. Pour pouvoir être traitées par les outils de gestion et les entrepôts de données traditionnels et interprétées utilement par les analystes, les données doivent être structurées, comme le sont, par exemple, les données de transaction recueillies par les entreprises auprès de leurs clients, ou encore les séries chronologiques que les organismes statistiques compilent sur divers indicateurs macroéconomiques et financiers. Les données non structurées, qui forment les 90 % restants, comprennent des courriels, des microtextes, des billets Facebook, des informations sur le trafic routier et des données audiovisuelles. Les entrepôts traditionnels ploient sous la charge de ces données et ne sont généralement pas en mesure de les traiter.

La **vélocité** renvoie au fait que les données produites à partir de quelques sources de mégadonnées comme les médias sociaux, les appareils mobiles, les transactions électroniques et les appareils en réseau sont actualisées très rapidement. Cela donne lieu à une avalanche de flux de données face à laquelle la plupart des appareils et logiciels d'analyse traditionnels sont impuissants. Pour extraire de la valeur en temps réel de données générées à grande vitesse, il faut des compétences et des systèmes d'analyse pointus.

◀ Le terme « mégadonnées » désigne de grands ensembles hétérogènes de données numériques, issues des transactions commerciales et des échanges dans les médias sociaux.

Encadré 1

Une ressource colossale

- Dans une étude réalisée en 2011, International Data Corporation (IDC) affirmait que 1,8 zettaoctet (1 800 milliards de giga-octets) de données serait créé cette année-là (Gantz et Reinsel, 2011), ce qui correspond à la capacité de stockage de 57,5 milliards d'iPads de 32 giga-octets (EMC², 2011).
- Les marques et organisations inscrites à Facebook reçoivent chaque jour 34 722 mentions « J'aime » par minute (Wikibon, 2012).
- IDC estime que, d'ici 2020, le nombre de transactions interentreprises et entre entreprises et consommateurs atteindra 450 milliards par jour (Wikibon, 2012).
- Walmart traite plus d'un million de transactions à l'heure avec ses clients. Celles-ci sont sauvegardées dans des bases de données dont le volume est estimé à plus de 2,5 pétaoctets (2,5 millions de giga-octets). Le volume de ces données représente 167 fois la bibliothèque du Congrès américain (Talend, 2012).
- En 2011 seulement, l'Association canadienne des paiements (ACP) a traité 6,3 milliards de paiements de détail (ACP, 2012).

L'attribut le plus important des grands volumes de données tient au fait qu'ils permettent de mettre à profit des connaissances et de créer une valeur considérable. Combinés à des méthodes d'analyse avancées, ils pourraient apporter un nouvel éclairage sur le comportement des ménages, les attentes des entreprises, la stabilité des marchés financiers et l'activité économique, et ainsi favoriser l'efficacité du processus décisionnel. Grâce à ces méthodes, il est possible, par exemple, d'analyser des tendances au sein d'un réseau social (qui peuvent être interreliées de façons extrêmement complexes) afin de déterminer quelle pourrait être leur influence sur les attentes d'inflation des consommateurs ou sur d'autres variables économiques (on trouvera des précisions à ce sujet dans Einav et Levin, 2013).

◀ *Combinés à des méthodes d'analyse avancées, les grands volumes de données pourraient apporter un nouvel éclairage sur le comportement des ménages, les attentes des entreprises, la stabilité des marchés financiers et l'activité économique.*

Les mégadonnées et l'analyse conjoncturelle : un aperçu de ce que l'avenir nous réserve

Étant donné l'importance que revêt la détention d'informations exactes et à jour sur l'état actuel de l'économie pour les décisions de politique monétaire, les mégadonnées offrent la possibilité d'améliorer l'analyse conjoncturelle en exploitant les données numériques issues des transactions commerciales et en mesurant l'humeur des consommateurs au moyen des médias sociaux et des requêtes sur Internet. Par exemple, en combinant des mégadonnées aux indicateurs mensuels existants, on pourrait anticiper le taux de croissance du PIB avant que les chiffres des comptes nationaux pour un trimestre donné ne soient rendus publics³.

Un avantage de l'utilisation des mégadonnées est qu'elles permettent de construire des mesures qui évoluent rapidement dans le temps. Au Massachusetts Institute of Technology, les économistes Alberto Cavallo et Roberto Rigobon dirigent une initiative, le Billion Prices Project (BPP)⁴, qui porte sur le calcul d'un indice de l'inflation au jour le jour à partir d'un panier de biens continuellement changeant. Les données sur les prix d'un large éventail de produits sont collectées par un logiciel qui parcourt les sites Web des détaillants en ligne⁵. L'indice BPP représente la moyenne des variations de prix individuelles. Cet indice virtuel de l'inflation en temps réel pourrait donner aux responsables des politiques et aux organismes statistiques un aperçu « en direct » des variations de l'inflation au fil du temps. Par exemple, les chiffres recueillis dans le cadre du projet montrent que des entreprises ont commencé à abaisser leurs prix presque immédiatement après la faillite de Lehmann Brothers en septembre 2008, ce qui donne à penser que la demande globale avait fléchi (Surowiecki, 2011). Or la mesure officielle de l'inflation publiée par les organismes statistiques n'a rendu compte de cette pression déflationniste qu'en novembre, au moment de la sortie des données sur l'IPC d'octobre.

Les Canadiens délaissent de plus en plus les moyens de paiement traditionnels comme l'argent comptant ou les chèques au profit de toute une gamme de modes de paiement électronique (Association canadienne des paiements, 2012). L'analyse de ces données électroniques à jour pourrait aider à prévoir l'évolution de l'activité économique et à estimer les révisions dont pourraient faire l'objet les données officielles sur le commerce de détail et la consommation. D'après certains travaux de recherche, il semble que les données des systèmes de paiement se prêtent à l'étude des

³ Binette et Chang (voir l'article page 4) décrivent un outil de prévision qui fait appel à un ensemble de données dont la taille est importante mais inférieure à celle des mégadonnées.

⁴ On trouvera de l'information sur le projet à l'adresse <http://bpp.mit.edu/>.

⁵ Les prix des services ne sont pas compris dans cet ensemble de données.

incidences économiques d'événements occasionnels extrêmes. Galbraith et Tkacz (2013), par exemple, utilisent les données quotidiennes sur les volumes des transactions par carte de débit et les volumes et valeurs des transactions par chèque au Canada pour examiner les répercussions qu'ont eues sur les dépenses de consommation des particuliers les attentats du 11 septembre 2001, l'épidémie du syndrome respiratoire aigu sévère au printemps 2003 et la panne de courant qui a frappé l'Ontario et certaines parties du Nord-Est et du Midwest américains en août 2003. Contrairement à ce que l'on pensait au moment où ces événements se sont produits, les auteurs n'observent que des effets temporaires et peu marqués.

Les mégadonnées pourraient également servir à étudier l'évolution des marchés du travail et du logement. Ces marchés ont été évalués en fonction du nombre de recherches effectuées sur Internet. Choi et Varian (2009) constatent qu'en considérant les requêtes relatives au chômage et au bien-être, on peut améliorer la prévision du nombre de demandes initiales de prestations d'emploi. Askitas et Zimmermann (2009), D'Amuri (2009) ainsi que Suhoy (2009) notent également que les recherches sur Internet peuvent se révéler pertinentes pour prévoir les conditions sur le marché du travail en Allemagne, en Italie et en Israël (respectivement). Selon Choi et Varian (2011) de même que Wu et Brynjolfsson (2009), l'exploitation des requêtes portant sur le domaine du logement permet d'obtenir de meilleurs résultats qu'avec les modèles traditionnels en ce qui concerne la prévision des ventes de maisons aux États-Unis. En outre, les travaux de Webb (2009) font ressortir que la corrélation élevée entre le nombre de requêtes relatives au mot *foreclosure* (saisie immobilière) et le nombre de saisies effectivement réalisées peut jeter les bases d'un système d'alerte avancée pour anticiper les problèmes sur le marché américain du logement.

McLaren et Shanbhogue (2011) évaluent dans quelle mesure les recherches en ligne peuvent aider à prévoir l'activité sur les marchés du travail et du logement au Royaume-Uni. À cette fin, ils spécifient deux modèles distincts dans lesquels la croissance du chômage et celle des prix des maisons sont fonction des taux de croissance enregistrés précédemment. Ils montrent qu'en intégrant les recherches sur Internet à ces modèles, on peut en améliorer les prévisions. Les auteurs soulignent que ces données sont particulièrement utiles pour analyser l'incidence d'événements inattendus comme une fermeture d'usine temporaire, une épidémie ou une grève. Les données d'une enquête sont nécessairement recueillies à partir de questions prédéterminées, tandis que celles qui proviennent des recherches sur Internet sont plus flexibles et se prêtent à l'étude de ces situations particulières.

Enfin, les mégadonnées pourraient concourir à l'élaboration des statistiques officielles. Certains pays européens font d'ailleurs appel aux données scannées au point de vente par les lecteurs de codes-barres pour calculer l'indice des prix à la consommation (IPC). En Norvège, l'organisme statistique utilise ces renseignements pour calculer un sous-indice des prix des aliments et des boissons non alcoolisées (Rodriguez et Haraldsen, 2006). En juin 2002, le Bureau central de la statistique des Pays-Bas a intégré à la préparation de l'IPC les données scannées aux caisses des supermarchés (Schut, 2002), tandis que l'Office fédéral suisse de la statistique a remplacé le relevé traditionnel des prix auprès des détaillants par les données provenant des lecteurs de codes-barres pour calculer ses indices des prix (Müller et autres, 2006).

◀ *Les mégadonnées pourraient concourir à l'élaboration des statistiques officielles.*

Défis et initiatives

Bien que les mégadonnées aient donné lieu jusqu'ici à des innovations concrètes, elles ne sont pas exploitées à leur plein potentiel du fait de plusieurs facteurs, notamment des contraintes méthodologiques, leur difficulté d'accès et certaines préoccupations entourant la protection de la vie privée.

Des contraintes méthodologiques

Malgré les avancées réalisées dans l'élaboration de méthodes permettant de dégager de la valeur des données volumineuses, l'application de ces méthodes à l'analyse conjoncturelle n'en est qu'à ses débuts. En particulier, on ne sait pas encore avec certitude quel est le meilleur moyen de sélectionner, d'organiser et de regrouper des données non structurées de manière à en tirer des indications utiles sur la conjoncture économique, ni quels sont les outils analytiques qu'il faudra développer pour intégrer ces indications aux informations provenant des sources habituelles. En outre, les groupes de population sur lesquels portent les mégadonnées ne sont pas forcément représentatifs de la population visée par les statistiques officielles. Il pourrait être difficile d'évaluer à l'aide des méthodologies standards la représentativité des échantillons de mégadonnées.

◀ *Malgré les avancées réalisées dans l'élaboration de méthodes permettant de dégager de la valeur des données volumineuses, l'application de ces méthodes à l'analyse conjoncturelle n'en est qu'à ses débuts.*

Un accès limité

Actuellement, les mégadonnées sont en grande partie entreposées dans des silos. Pour en exploiter toutes les potentialités, il faudrait d'abord intégrer les ensembles de données fragmentées afin qu'ils soient facilement et rapidement accessibles aux parties intéressées. L'avènement de l'informatique en nuage a permis de créer des centres de données qui abritent des masses colossales de données en un même endroit. Comme il est très important de regrouper les ensembles de données, plusieurs initiatives sont en cours pour améliorer l'accès aux mégadonnées. Par exemple, en collaboration avec IBM et un consortium de sept universités, les autorités canadiennes et ontariennes ont lancé un nouveau projet de 210 millions de dollars qui prévoit la création d'un centre de recherche et de données, situé en Ontario, dont le but est d'aider les chercheurs en économie et dans les universités à faire appel à une infrastructure infonuagique à haute performance afin de mieux tirer profit des mégadonnées. Aux États-Unis, la Bibliothèque du Congrès américain et Twitter ont signé en 2010 un accord prévoyant de mettre 170 milliards de micromessages archivés à la disposition des chercheurs et d'autres parties intéressées dont les travaux portent sur des sujets allant du suivi des taux de vaccination à la prévision de l'activité sur les marchés boursiers (Osterberg, 2013).

Mégadonnées ou cybersurveillance?

Les données individuelles recueillies au niveau des ménages et des entreprises peuvent fournir des informations pertinentes sur la conjoncture économique du moment. Mais en mettant en lumière des liens cachés entre des éléments d'information apparemment disparates, l'analyse des mégadonnées risque de donner accès à des renseignements personnels que certains jugeraient trop sensibles pour être communiqués. Les motifs présidant à la collecte de ces données et les précautions qui s'imposent sont en train de prendre de l'importance dans les débats sur la vie privée et l'usage approprié des données personnelles. Néanmoins, si les institutions et les particuliers doivent être prudents étant donné le caractère potentiellement intrusif de leurs efforts pour obtenir de grosses quantités d'informations, on peut

difficilement nier que l'analyse des mégadonnées est susceptible de livrer des renseignements précieux sur la croissance économique et d'améliorer l'analyse conjoncturelle. Un cadre réglementaire équilibré est par conséquent nécessaire pour répondre efficacement aux préoccupations entourant la protection de la vie privée et l'utilisation des renseignements personnels, tout en bénéficiant des progrès techniques et du dynamisme d'une économie prospère fondée sur les données.

◀ *Un cadre réglementaire équilibré est nécessaire pour répondre efficacement aux préoccupations entourant la protection de la vie privée, tout en bénéficiant des progrès techniques.*

Conclusion

Une information fiable au sujet de la situation économique actuelle est un élément important de la conduite de la politique monétaire et comme sur ce plan, les données constituent la principale ressource de l'analyse conjoncturelle, il est primordial qu'elles soient exactes et à jour. Une évaluation plus précise de la conjoncture, en temps réel, peut donner lieu à une meilleure estimation de la vigueur de l'économie et à des prévisions plus justes de la croissance. La numérisation des informations et l'avènement d'Internet ont entraîné une hausse exponentielle des volumes de données disponibles et aussi créé de nouvelles sources viables. De façon pratique, ces avancées permettent notamment de construire rapidement des mesures de prix à partir des prix de vente au détail affichés en ligne, d'exploiter les données relatives aux paiements électroniques à l'appui des prévisions économiques et d'utiliser les recherches sur Internet pour évaluer la situation des marchés du travail et du logement. Mais tirer parti de toutes les possibilités de cette abondance de données présente des défis. Si certains progrès ont été accomplis pour dégager de la valeur de ces données à l'aide de méthodes d'analyse traditionnelles, l'utilisation des mégadonnées aux fins de l'analyse conjoncturelle ne fait que commencer.

Ouvrages et articles cités

Askitas, N., et K. F. Zimmermann (2009). « Google Econometrics and Unemployment Forecasting », *Applied Economics Quarterly*, vol. 55, n° 2, p. 107-120.

Association canadienne des paiements (2012). *Examen des méthodes de paiement et des tendances des paiements au Canada*, octobre.

Binette, A., et J. Chang (2013). « CSI : un modèle de suivi de la croissance à court terme du PIB réel du Canada », *Revue de la Banque du Canada*, été, p. 4-15.

Choi, H., et H. Varian (2009). *Predicting Initial Claims for Unemployment Benefits*, Google Inc., 5 juillet. Internet : http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/fr//archive/papers/initialclaimsUS.pdf.

——— (2011). *Predicting the Present with Google Trends*, Google Inc., 18 décembre. Internet : <http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf>.

- Coletti, D., et S. Kozicki (2013). « Introduction : les outils utilisés à la Banque du Canada pour l'analyse de la conjoncture », *Revue de la Banque du Canada*, été, p. 1-3.
- Daas, P., et M. van der Loo (2013). *Big Data (and Official Statistics)*, Statistics Netherlands. Document de travail présenté à la réunion sur la gestion des systèmes d'information statistique (MSIS 2013) tenue à Paris et Bangkok, 23-25 avril.
- D'Amuri, F. (2009). *Predicting Unemployment in Short Samples with Internet Job Search Query Data*, document n° 18403, Munich Personal RePEc Archive.
- Einav, L., et J. D. Levin (2013). *The Data Revolution and Economic Analysis*, document de travail n° 19035, National Bureau of Economic Research.
- EMC² (2011). *Le Monde n'est que « Big Data » : les données mondiales vont plus que doubler tous les deux ans et devraient atteindre 1,8 zettaoctets en 2011, selon la nouvelle étude IDC-EMC*, communiqué, 28 juin.
- Galbraith, J. W., et G. Tkacz (2013). « Analyzing Economic Effects of September 11 and Other Extreme Events Using Debit and Payments System Data », *Analyse de politiques*, vol. 39, n° 1, p. 119-134.
- Gantz, J., et D. Reinsel (2011). *Extracting Value from Chaos*, International Data Corporation (IDC) Digital Universe. Internet : <http://www.emc.com/leadership/programs/digital-universe.htm>.
- Gens, F. (2011). *IDC Predictions 2012: Competing for 2020*, IDC Analyze the Future, décembre. Internet : <http://cdn.idc.com/research/Predictions12/Main/downloads/IDCTOP10Predictions2012.pdf>.
- Lamont, O. (1997). « Do “Shortages” Cause Inflation? », *Reducing Inflation: Motivation and Strategy*, sous la direction de C. D. Romer et D. H. Romer, Chicago, University of Chicago Press, p. 281-306.
- Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh et A. Hung Byers (2011). *Big Data: The Next Frontier for Innovation, Competition and Productivity*, McKinsey Global Institute, mai.
- McLaren, N., et R. Shanbhogue (2011). « Using Internet Search Data as Economic Indicators », *Quarterly Bulletin*, Banque d'Angleterre, deuxième trimestre, p. 134-140.
- Müller, R., H. M. Herren, S. Röthlisberger et C. Becker Vermeulen (2006). *Recent Developments in the Swiss CPI: Scanner Data, Telecommunications and Health Price Collection*, Office fédéral de la statistique de la Suisse. Étude présentée à la neuvième réunion du Groupe d'Ottawa, Londres, 14-16 mai.
- Osterberg, G. (2013). *Update on the Twitter Archive at the Library of Congress*, blogue de la Bibliothèque du Congrès des États-Unis, 4 janvier.
- Rodriguez, J., et F. Haraldsen (2006). « The Use of Scanner Data in the Norwegian CPI: The “New” Index for Food and Non-Alcoholic Beverages », *Economic Survey*, n° 4, p. 21-28.

- Schut, C. (dir.) (2002). *Gebruik van Scannerdata van Supermarkten in de Consumentenprijsindex*, Statistics Netherlands, 4 juillet.
- Suhoy, T. (2009). *Query Indices and a 2008 Downturn: Israeli Data*, document d'analyse n° 2009-06, Banque d'Israël.
- Surowiecki, J. (2011). « A Billion Prices Now », *The New Yorker*, 30 mai.
- Talend (2012). *How Big is Big Data Adoption?* Résultats d'enquête.
Internet : http://info.talend.com/rs/talend/images/WP_EN_BD_Talend_SurveyResults_BigDataAdoption.pdf.
- Webb, G. K. (2009). « Internet Search Statistics as a Source of Business Intelligence: Searches on Foreclosure as an Estimate of Actual Home Foreclosures », *Issues in Information Systems*, vol. 10, n° 2, p. 82-87.
- Wikibon (2012). *A Comprehensive List of Big Data Statistics*, blogue, 1^{er} août.
Internet : <http://wikibon.org/blog/big-data-statistics/>.
- Wu, L., et E. Brynjolfsson (2009). *The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales*, Sloan School of Management, Massachusetts Institute of Technology.