

Modèles économiques des données : une relation complexe entre demande et offre

Par Paul BELLEFLAMME

Professeur à Aix-Marseille School of Economics

Le 19 mars 2018, Facebook dégringole en Bourse après la révélation que la société Cambridge Analytica a utilisé les données personnelles de près de 50 millions d'utilisateurs du réseau social sans leur consentement⁽¹⁾. On apprend aussi ce jour-là que les *selfies* des internautes, qui peuvent servir à valider des processus d'identification, se négocient à des prix allant jusqu'à 70 dollars sur les marchés clandestins du « Dark Web ». Quelques mois plus tôt, la société iRobot revenait sur ses déclarations antérieures selon lesquelles elle cherchait à revendre les données collectées par ses robots aspirateurs Roomba, capables de réaliser une carte virtuelle des endroits qu'ils nettoient. La société Uber, quant à elle, annonçait la création d'une plateforme visant à partager gratuitement des données de déplacements de ses chauffeurs et clients avec les planificateurs urbains des quatre cent cinquante villes où elle est active⁽²⁾.

Ces quelques événements récents montrent toute l'importance que les échanges de données occupent dans nos économies. Ils illustrent aussi les différentes formes que peuvent prendre ces échanges, partage librement consenti (cas d'Uber), vol pur et simple (les *selfies* sur le Dark Web), ou encore échanges encadrés par des dispositions contractuelles plus ou moins claires (cas de Facebook et d'iRobot).

L'objectif de cet article est de mieux faire comprendre comment s'organisent les échanges de données. Pour ce faire, nous commençons par décrire le côté de la demande, en étudiant pourquoi, et comment, les données acquièrent de la valeur (Section 1). Nous considérons ensuite le côté de l'offre, en nous demandant d'où viennent les données et qui en contrôle la production et la collecte (Section 2). Il s'agit enfin de comprendre comment l'offre et la demande se rencontrent (Section 3). Nous concluons en réfléchissant aux évolutions que pourraient prendre les échanges de données dans le futur.

Le côté de la demande

La demande de données émane d'entreprises, mais aussi d'organisations non commerciales (des villes par exemple) qui cherchent à améliorer leurs pratiques. Il s'agit d'une demande induite car ce ne sont pas les données en elles-mêmes qui sont recherchées mais bien les informations qui peuvent en être extraites et, finalement, les connaissances que génèrent ces informations et qui contribuent à la prise de décisions⁽³⁾. On comprend donc pourquoi la demande de données est un phénomène récent. En effet, la capacité d'accroître la valeur des données en les transformant en informations a considérablement augmenté ces dernières années sous l'effet conjoint de la numérisation et de la « datafication ». La première tendance est la généralisation du format numérique qui permet de stocker, dupliquer et transmettre les données électroniquement bien plus vite et à un coût énergétique nettement moindre. La seconde tendance est la multiplication des traces

(1) Révélation faite par *The Guardian* <http://bit.ly/2plU1sM> et *The New York Times* (<http://nyti.ms/2u1nLjw>)

(2) Voir, respectivement, <http://bit.ly/2u8ys3W>, <http://bit.ly/2DFoxCv> et <http://bit.ly/2prPK6B>

(3) Thierauf (1999) définit les données comme une collection non structurée de faits et de chiffres, l'information comme des données structurées et la connaissance comme de « l'information à propos de l'information ».

numériques laissées derrière elles par nos activités, que ce soit par nos ordinateurs et smartphones, par les réseaux sociaux, ou les senseurs de nos objets connectés. À cela s'ajoute le développement d'une nouvelle discipline scientifique, la science des données, qui combine outils mathématiques, statistiques et informatiques pour optimiser l'extraction de connaissances à partir d'ensembles de données.

En résumé, la demande de données est en pleine expansion parce que tant les données disponibles que la capacité de les traiter ne cessent de croître. La valeur des données augmente en effet avec ce qu'il est convenu d'appeler les quatre V des données, à savoir leur volume (d'où le terme de *big data*, qui suggère des économies d'échelle), leur variété (c'est-à-dire la diversité de leurs sources, qui suggère des économies d'envergure), leur vélocité (c'est-à-dire la rapidité avec laquelle les flux de données peuvent être traités) et, naturellement, leur véracité (ou leur précision, qui détermine la confiance qu'on peut leur accorder).

Les entreprises sont avides de données parce qu'elles cherchent à améliorer leurs processus de production, à développer des produits et services innovants et à mieux cibler leurs clients avec des offres, des publicités et des prix adaptés. Comme chaque entreprise a pour objectif de surpasser ses concurrentes, une course s'engage à qui utilisera au mieux les données disponibles. En découlent deux conséquences importantes pour la demande de données. D'une part, les entreprises ont une disposition à payer beaucoup plus élevée pour des données auxquelles elles ont un accès exclusif que pour des données qu'elles devraient partager avec leurs concurrents. D'autre part, il est possible qu'au sein d'une industrie, les entreprises concurrentes investissent de manière excessive dans l'acquisition et le traitement de données, avec comme effet que les profits des entreprises finissent par baisser. En d'autres termes, comme dans le célèbre dilemme du prisonnier, des entreprises concurrentes gagneraient à restreindre collectivement leur utilisation de données, mais aucune n'y trouve intérêt individuellement.

Le côté de l'offre

Les données que valorisent les entreprises proviennent de trois sources. Tout d'abord, de nombreuses bases de données sont en accès libre. La plus grosse partie de ces « données ouvertes » est produite par le secteur public (on pense à des données statistiques, scientifiques ou cartographiques) ; des organisations comme les universités ou les organisations non gouvernementales ouvrent également leurs données ; même des entreprises commerciales peuvent y trouver un intérêt (à l'instar d'Uber mentionnée dans l'introduction). Ensuite, les entreprises produisent elles-mêmes énormément de données au fil de leurs activités et par les produits qu'elles vendent⁽⁴⁾. Finalement, vous et moi sommes sans doute les plus gros pourvoyeurs des données qui intéressent les entreprises. C'est le phénomène de « datafication » que nous évoquions plus haut : nous produisons des données, soit directement par nos activités (les photos ou les « likes » que nous postons sur les réseaux sociaux, les sites que nous visitons, les mails que nous envoyons, etc.), soit indirectement par les machines ou équipements que nous utilisons (un smartphone dont la géolocalisation est activée ou une montre connectée par exemple). Ces données sont précieuses pour les entreprises dans la mesure où elles indiquent nos goûts, nos habitudes de consommation, nos interactions sociales, etc.

Pour la suite de notre analyse, il est important de déterminer dans quelle mesure on peut parler d'une offre de données. Pour qu'une offre existe, il faut que l'accès aux données puisse être contrôlé, de sorte que le producteur puisse fixer les termes d'une éventuelle transaction. Pour les deux

(4) On estime par exemple qu'une voiture autonome génère jusqu'à 100 gigabytes de données par seconde (soit l'équivalent de plus de 5 millions de pages de texte).

premières sources de données, les producteurs – organismes publics, entreprises – sont largement en mesure de déterminer les conditions d'accès à leurs données : l'accès est délibérément rendu public pour les données ouvertes ; pour les données d'entreprises, nous verrons dans la section suivante que l'accès est le plus souvent fermé ou encadré par des dispositions contractuelles.

Qu'en est-il des données que nous produisons en tant qu'individus ? Pouvons-nous en contrôler l'accès ? En théorie, oui. Les sites web que nous visitons, ou les objets connectés que nous utilisons, nous invitent à signifier notre accord avec leurs conditions d'utilisation. Même si la possibilité nous est laissée de refuser que nos données soient collectées, nous n'exerçons pas, ou très peu, cette option (qu'on appelle en anglais « opt out »). Pourquoi ? Une première raison est que nous jugeons trop coûteux (en temps et en effort) de prendre connaissance des conditions d'utilisation ou d'appliquer des mesures pour limiter la collecte de nos données⁽⁵⁾. Une seconde raison, qui justifie partiellement la première, est que nous acceptons d'obtenir, en échange de nos données, des services moins chers (souvent gratuits), mieux adaptés à nos besoins (comme des offres ciblées) et potentiellement de meilleure qualité⁽⁶⁾. Cela revient à dire que nous associons un « prix virtuel » à nos données et donc à notre vie privée.

Il arrive que ce prix virtuel devienne un prix réel. C'est le cas quand des entreprises (par exemple des fournisseurs d'accès à Internet) différencient leurs services en proposant aux consommateurs de payer plus cher pour éviter de voir leurs données collectées ou de recevoir des publicités ciblées. En choisissant ce genre d'offres, les consommateurs révèlent leur volonté de payer pour protéger leur vie privée. Il s'agit toujours ici d'un système « opt out », puisque c'est au consommateur de payer pour fermer l'accès à ces données. Que se passe-t-il si, à l'inverse, c'est à l'entreprise de payer le consommateur pour qu'il ouvre l'accès à ses données (système « opt in ») ? On a envie de penser que rien ne devrait changer pour un même montant monétaire (à payer ou à recevoir) et une même variation (à la hausse ou à la baisse) du degré de protection des données. Mais des études montrent qu'en général, les consommateurs demandent en échange d'une érosion de leur vie privée un montant monétaire plus élevé que celui qu'ils sont prêts à payer pour protéger leur vie privée dans une même mesure⁽⁷⁾. Les consommateurs semblent donc attacher une valeur plus importante à leurs données quand leur consentement est nécessaire pour l'utilisation (« opt in »), plutôt que pour l'absence d'utilisation (« opt out ») de celles-ci.

La mise en relation de l'offre et de la demande

Williamson (1991) distingue trois façons d'organiser les transactions économiques : la « hiérarchie » organise les transactions au sein d'une entreprise intégrée, le « marché » utilise le mécanisme des prix pour coordonner offre et demande et, entre ces deux extrêmes, les « formes hybrides » reposent sur des contrats spécifiques.

Actuellement, les transactions sur les données s'organisent essentiellement par le mode hiérarchique ou par des formes hybrides. Dans le premier cas, les entreprises collectent directement, ou produisent elles-mêmes, les données dont elles ont besoin. Quand il s'agit de données personnelles, nous avons vu plus haut que la collecte s'appuie sur des contrats de type « opt out » : pour dire les choses crûment, les entreprises se servent tant que les consommateurs ne les en empêchent pas. Nous avons montré aussi que l'intégration verticale se justifie dès lors que les données permettent

(5) Il faudrait 76 jours pour lire l'intégralité des conditions d'utilisation qu'un Américain moyen accepte de signer en un an (GRALLET *et al.*, 2018). Pour limiter l'accès à ses données, il est possible, par exemple, d'effacer les cookies de son navigateur ou de passer par des serveurs proxy.

(6) Par exemple, un compteur communicant n'a véritablement de valeur ajoutée que s'il peut mesurer de manière précise notre consommation d'eau ou d'électricité.

(7) Voir ACQUISTI *et al.* (2013) ; voir aussi SCHOLZ (2014).

d'obtenir un avantage concurrentiel (les entreprises n'ont en effet aucun intérêt à partager les données qu'elles récoltent et moins encore l'information et la connaissance qu'elles en extraient). Les dispositions légales limitant le partage de données personnelles viennent renforcer cette tendance à l'intégration verticale.

Il arrive toutefois que des entreprises trouvent profitable de partager leurs données, afin de mieux coordonner leurs activités ⁽⁸⁾. Les transactions se basent alors sur des contrats multilatéraux de long terme. Une autre forme de gouvernance hybride est le recours à des intermédiaires spécialisés dans la collecte et le traitement de données. On les appelle « courtiers en données » (*data brokers* en anglais). Les services sur mesure que proposent ces courtiers sont particulièrement prisés par les entreprises qui ne peuvent pas collecter de données par elles-mêmes. En raison des économies d'échelle et d'envergure évoquées plus haut, l'industrie des courtiers en données est dominée par quelques entreprises, en majorité américaines, qui rassemblent des données diverses sur des centaines de millions de consommateurs de par le monde ; citons Acxiom (marketing), Equifax (assurance), Experian (crédit), Corelogic (immobilier) ou Datalogix (finance).

À ce jour, il n'existe pas de « marché des données » à proprement parler. On trouve certes quelques plateformes d'échange de données mais celles-ci sont limitées à une industrie particulière et restreignent considérablement les transactions qui peuvent être effectuées. Ceci s'explique par le paradoxe suivant : comme les données sont stratégiques, la disposition à payer pour des données non exclusives est généralement faible, voire nulle ; mais parce que les données sont non rivales (la consommation par l'un ne réduit pas les possibilités de consommation par l'autre), l'exclusivité est difficile à garantir, singulièrement dans un mécanisme d'échange décentralisé. En outre, il est difficile d'établir rigoureusement la véracité des données, ainsi que leur valeur, en raison de leur unicité (absence de point de comparaison) ou de leur complémentarité (il faut combiner plusieurs bases de données pour extraire de l'information pertinente ⁽⁹⁾).

Conclusion

En résumé, une quantité sans cesse croissante de données est produite, collectée et utilisée mais, en définitive, une fraction assez limitée de ces données est échangée. Nous avons identifié trois explications : le caractère stratégique des données pour les entreprises, la difficulté d'organiser des places de marché décentralisées et le manque de contrôle des individus sur les données qu'ils produisent. Sur ce dernier point, on peut s'attendre à des transformations importantes dans un futur assez proche. En effet, un nouveau texte européen, intitulé Règlement général sur la Protection des Données (RGPD), vient d'entrer en vigueur ; il impose aux entreprises de donner aux individus davantage de contrôle sur leurs données personnelles. Cela signifie que les entreprises vont devoir obtenir un consentement explicite et positif des individus pour pouvoir utiliser leurs données et, également, assurer la portabilité de ces données (c'est-à-dire permettre aux consommateurs d'emporter leurs données avec eux lorsqu'ils changent de fournisseur). Comme l'expliquent Peitz et Schweitzer (2017), la portabilité empêche le verrouillage et facilite ainsi la concurrence dans l'accès aux données personnelles (à défaut de mettre en place un véritable marché secondaire des données).

Une autre source de changement est le développement de nouveaux intermédiaires qui proposent aux consommateurs des solutions pour gérer activement leurs données personnelles et, potentiellement, les monétiser ⁽¹⁰⁾. Enfin, le scandale Facebook/Cambridge Analytica (qui ouvre cet article)

(8) On pense aux entreprises collaborant à la mise au point de voitures autonomes.

(9) Pour une analyse plus détaillée, voir KOUTROUMPIS *et al.* (2017).

(10) On les appelle PIMS (*Personal Information Management Systems*) ou systèmes de gestion des informations personnelles. Les plus connus sont Datacoup, Digi.me et Meeco.

a suscité de telles réactions des internautes et des pouvoirs publics qu'on est en droit de penser qu'une nouvelle ère commence, où les transactions sur les données personnelles seront plus encadrées, plus transparentes et plus respectueuses des individus.

Références

- ACQUISTI A., JOHN L.K. & LOEWENSTEIN G. (2013), "What is Privacy Worth?", *The Journal of Legal Studies*, 42, pp. 249-274.
- GRALLET G., PONCET G. & PONS H. (2018), « Données personnelles : comment reprendre le contrôle ? », *Le Point* (25 janvier).
- KOUTROUMPIS P., LEIPONEN A. & THOMAS L. (2017), "The (Unfulfilled) Potential of Data Marketplaces", *ETLA Working Papers*, n° 53. <http://pub.etla.fi/ETLA-Working-Papers-53.pdf>
- PEITZ M. & SCHWEITZER H. (2017), "Datenmärkte in der digitalisierten Wirtschaft: Funktionsdefizite und Regelungsbedarf", Discussion Paper No. 17-043, ZEW, Mannheim.
- SCHOLZ E.-M. (2014), "Putting a Price on Privacy – an Introduction to the Economics of Privacy", IPdigit.eu (1 June), <http://www.ipdigit.eu/2014/06/putting-a-price-on-privacy-an-introduction-to-the-economics-of-privacy/>
- THIERAUF R.J. (1999), *Knowledge Management Systems*, Quorum Books.
- WILLIAMSON O.E. (1991), "Comparative Economic Organization: the Analysis of Discrete Structural Alternatives", *Administrative Science Quarterly*, 36, 269-296.