

Organiser la valorisation de l'or noir du XXI^e siècle

Par **Éric BROUSSEAU**

Université Paris Dauphine-PSL ; Chaire « Gouvernance et Régulation »

La donnée n'est pas un bien collectif pur qui pourrait ressortir d'un régime de partage généralisé. Elle n'est pas non plus un bien se prêtant aisément à l'organisation d'un système efficace d'échange marchand. Sous ces deux régimes cardinaux d'organisation de leur circulation, la diversité et la granularité des données disponibles donneraient lieu à des gains en efficacité très inférieurs à leur potentiel. Telles sont les raisons pour lesquelles il convient de dépasser une approche purement « réglementaire », en prenant en considération les relations entre producteurs et utilisateurs de données, et l'éventuelle intervention d'intermédiaires qui peuvent, de manière neutre, fournir des services allant du traitement technique des données à la fourniture de services issus de leur analyse. Il convient cependant de veiller à ce qu'alliances ou intermédiaires n'abusent pas de la position centrale qu'ils peuvent occuper.

Depuis le milieu des années 2010, l'idée que les données seraient l'or noir du XXI^e siècle a fait florès dans le débat public. Cette comparaison avec la révolution énergétique, avec ses conséquences économiques et géopolitiques, qu'a constituée l'exploitation des hydrocarbures au tournant du XX^e siècle a permis de souligner à la fois l'immense gisement de valeur qu'elles constituent et les dangers de concentration du pouvoir économique entre les mains de quelques grandes entreprises prenant le contrôle de leur exploitation. Bien évidemment, de nombreuses contributions ont aussi souligné la différence entre les problématiques liées à une ressource naturelle finie et une ressource immatérielle infinie... et donc la limite de la métaphore.

Autorisons-nous cependant à la filer quelques instants. Pas plus que les molécules d'hydrocarbures emprisonnées dans une roche poreuse, la donnée n'est utilisable à « l'état de nature ». De la même façon qu'il faut concevoir et déployer un appareil industriel d'extraction, de transport et de raffinage pour obtenir des combustibles ou des composants chimiques, il convient de mettre en œuvre toute une organisation « industrielle » pour créer la donnée, la qualifier, l'organiser et la transformer en intrants potentiels dans des processus analytiques susceptibles de produire de la valeur. Mais là encore, de nombreux actifs complémentaires sont nécessaires. Les hydrocarbures ne produisent de la valeur que lorsqu'ils sont transformés en services (transport, éclairage, chauffage, etc.) ou en biens matériels. Les données, quant à elles, doivent être transformées en information et connaissance qui, réinjectées dans les appareils de production ou de prestation de service, rendent ces derniers plus efficaces ou adaptés aux besoins des usagers. *In fine*, la métaphore souligne que la donnée n'est pas un « don de la nature », mais le fruit et l'ingrédient d'un appareillage industriel et organisationnel créé par des entrepreneurs et des organisations espérant un retour sur investissement et des rentes à hauteur des risques encourus. Elle montre également que ces appareils industriels sont cruciaux pour maîtriser la manière dont la valeur est produite et répartie.

Cette capacité qu'auraient quelques grands acteurs industriels ou techniques de s'approprier les données et de les exploiter à leur seul bénéfice a induit un débat intense

sur l'obligation de partage. Au-delà des opérateurs de service public, les entreprises en position centrale dans une filière, voire toutes les entreprises, devraient partager leurs données, non seulement parce que cela aurait des vertus pro-concurrentielles, mais aussi parce que ce serait cohérent avec leur nature de bien non rival en consommation. Dans le même temps, les données personnelles pourraient faire l'objet de droits de propriété permettant à des individus de les céder à des tiers en échange de services gratuits, voire de rémunération.

Dans les lignes qui suivent, nous nous proposons de développer une réflexion sur l'économie de la donnée destinée à mieux comprendre les impasses auxquelles pourraient mener une approche trop simplificatrice des obligations de partage (ignorant notamment les incitations à produire la donnée utile à des tiers ou à l'intérêt collectif), ou au contraire une approche centrée sur l'organisation de la circulation des données sur la base du seul échange marchand (ignorant alors les nombreuses sources d'échec de marché, sans parler des échecs de concurrence, propres aux marchés informationnels). Il apparaît en effet utile de considérer le rôle de la coopération volontaire et de l'intermédiation pour envisager la manière dont les échanges de données peuvent être gouvernés ; ce qui n'est pas sans évoquer certains arrangements organisationnels articulant concurrence et coopération, mis en œuvre dans l'industrie pétrolière pour exploiter en commun sites d'extraction, systèmes de transport, et capacités de raffinage ou de distribution. Cela nous conduira à conclure sur la nécessité d'une supervision du fonctionnement d'alliances et d'intermédiaires, toujours sur le « fil du rasoir » entre la tentation d'abuser d'une position clé et la nécessité d'une coordination au bénéfice de l'intérêt général.

LE COÛT DE LA DONNÉE

Peut-on définir un régime d'accès, d'usage, de bénéfice des services dérivés ainsi que de transformation ou d'aliénation en fonction des caractéristiques de production et d'usage de la donnée ? Les réflexions engagées dans ce sens insistent le plus souvent sur leur caractère non rival (dans l'usage) et sur le coût marginal supposé quasi nul de leur reproduction et de leur conservation, ainsi que sur leur potentialité en termes de création de valeur, pour faire de la donnée une ressource qui ressortirait d'un régime d'accès largement ouvert et de propriété commune (cf. Duch-Brown *et al.*, 2016). Dans le même temps, la protection des droits individuels et des libertés conduirait à exclure les données à caractère personnel de ce régime, imposant le consentement des individus pour la collecte et les usages des informations les concernant. On exclurait aussi du régime général les données résultant d'une œuvre créatrice, par exemple parce qu'elles sont organisées en bases de données, ou qu'elles se réfèrent à des créations intellectuelles protégées. Évidemment, le secret industriel, ou celui des affaires, rendrait aussi nécessaires des exceptions au principe d'un accès totalement ouvert. Sans revenir sur les nombreux régimes d'exception qu'il faut imaginer, ni sur les difficultés de mise en œuvre pratique de cette multitude de régimes, on perçoit que la donnée n'est pas un objet si aisé à saisir à partir de ses supposées caractéristiques intrinsèques de production et d'usage découlant de son « immatérialité ».

L'idée, tout d'abord, que la donnée serait une ressource « naturellement » produite par les infrastructures numériques est en partie erronée. Certes, des traces numériques sont issues du fonctionnement des systèmes numériques, mais ces traces ne sont des données qu'une fois qu'elles sont articulées à une méta-information permettant d'associer du sens à une trace et précisant le contexte de sa production, la reliant à d'autres traces dans le temps, l'espace et les composantes du système concerné. Qu'il s'agisse d'un clic ou d'une information issue d'un capteur, il faut que l'information contextuelle – date, heure, information fournie sur l'écran où a été effectué le clic, localisation et caractéristiques du capteur, etc. – permette à la fois une interprétabilité de ladite trace et de la relier à d'autres. Il ne faut pas négliger non plus les éléments permettant d'authentifier la trace

et d'attester de son intégrité (ou des modifications qui lui ont été apportées). En bref, la donnée n'est pas un objet isolable, et ne trouve son sens qu'au sein d'un ensemble informationnel. Elle est une composante d'un système organisé de données.

Cette idée de système organisé de données renvoie d'emblée à une forte matérialité ainsi qu'à des coûts de production qui sont loin d'être négligeables. Concrètement, les données sont générées par une multitude de capteurs, transportées *via* des réseaux de communication vers des serveurs qui les traitent afin de les organiser dans des bases de données et les stocker. Les investissements sont considérables et le coût marginal est loin d'être nul, si l'on prend en considération les dépenses énergétiques des infrastructures physiques mobilisées. La volumétrie est telle que les opérateurs industriels décident la plupart du temps de ne rapatrier qu'une partie des traces générées, et de n'en traiter et stocker qu'une partie encore plus infime. Au-delà de ces coûts « physiques », il faut être conscient qu'une partie des données résultent d'expérimentations. Ainsi, un prestataire de service en ligne va tester une nouvelle interface ou une tarification et observe les réactions de ses usagers. Un opérateur industriel va expérimenter un aménagement de son système – par exemple un nouveau principe de gestion de la recharge d'une flotte de véhicules électriques, pour mesurer les conséquences sur la disponibilité des dits véhicules. De telles expériences ont des coûts, tant pour les mettre en œuvre que parce qu'elles peuvent avoir des effets de bord se traduisant par une dégradation de la qualité du service pour l'utilisateur et, *in fine*, des pertes de recettes ou une dégradation de la réputation de l'opérateur.

Au-delà du coût de la production de la donnée, il faut prendre en considération son coût de mise à disposition. Revenons sur la manière dont les données génèrent de la valeur. L'intuition générale est que la circulation en temps réel des données va permettre un pilotage/une coordination adaptative d'un système technique ou d'un ensemble d'opérateurs, notamment pour assurer une fiabilité systémique et une personnalisation des prestations. La collecte et le partage de données « historiques » doivent permettre quant à eux d'analyser le fonctionnement d'un système ou d'un processus afin d'en corriger les défaillances, d'en améliorer la conception et d'en automatiser le fonctionnement (grâce à des algorithmes produits de manière analytique ou par apprentissage automatisé). Cela étant, tous les utilisateurs potentiels des données n'ont pas tous les mêmes besoins. Un aménageur urbain doit disposer de données historiques pour comprendre comment organiser et dimensionner des infrastructures, quand un opérateur de micromobilité a besoin de données en temps réel sur les points de stationnement et de recharge disponibles pour gérer son parc, et que deux opérateurs de transport peuvent avoir intérêt à partager des données nominatives sur leurs passagers pour gérer les trajets de ces derniers de bout en bout. Chacun a besoin de données avec des granularités et temporalités différentes alors que, pour des raisons de volumétrie, de protection de la vie privée, ou encore de sécurité, toutes les données « brutes » ne peuvent pas être rendues accessibles à tous. Il faut donc façonner les jeux de données en fonction des besoins des utilisateurs, ce qui induit des coûts.

Il convient aussi de prendre en compte le contexte dans lequel les échanges de données ont lieu. Les données permettant l'identification et le traçage des personnes peuvent aussi bien être utiles pour personnaliser une prestation de service que pour manipuler les individus en les enfermant dans des « bulles informationnelles ». Celles qui rendent possible une coordination décentralisée par ajustement mutuel (travail collaboratif) entre participants d'une chaîne de valeur peuvent être utilisées pour mettre en place des mécanismes de surveillance et de gouvernance algorithmiques. Il en va de même des données non personnelles traçant le fonctionnement d'un système technique : elles peuvent être la source d'optimisation par identification des défaillances ou goulets d'étranglement, comme elles peuvent permettre à des concurrents ou des agresseurs d'identifier des vulnérabilités. La donnée peut donc facilement devenir « sensible », et il importe de contrôler les mains entre lesquelles elles se retrouvent. Pour minimiser les risques, il faut à la fois contrôler les

accès et mettre en forme les données (*i.e.* anonymiser, diminuer la granularité, retarder la mise à disposition, etc.). Ces efforts de “*packaging*” doivent être réalisés sous différents formats, en direction de différentes catégories d'utilisateurs, voire être réalisés sur mesure. Tout ceci représente des coûts de conception, mise en œuvre, et contrôle.

Si les coûts de mise à disposition des données ne sont pas nuls, il convient enfin de prendre en considération les conséquences des contraintes de circulation des données sur leur valeur d'usage pour les utilisateurs. Plus les jeux de données sont fournis de manière « générique » pour une large gamme d'utilisateurs, moins ils sont porteurs de sens, car la donnée est alors nécessairement très agrégée et conçue en vue d'une gamme limitée d'usages. Par ailleurs, une donnée initiale est *de facto* effacée si elle est agrégée dans une statistique ou un algorithme. Elle n'est alors plus disponible pour produire une information destinée à un autre usage.

LA CHAÎNE DE VALEUR DES DONNÉES

Au-delà de la donnée, il faut prendre en considération la chaîne de valeur de la donnée. La trace numérique n'est en fait transformée en donnée qu'au prix d'opérations de certification, codification et organisation qui lui donnent un signifiant. Une fois la donnée « produite », encore faut-il l'apparier à d'autres données afin d'analyser les bases de données qui en résultent à l'aide de techniques statistiques. On fait alors émerger des régularités ou l'on détecte des anomalies, qui constituent autant d'informations que l'on peut ensuite, dans un troisième temps, articuler les unes avec les autres pour réaliser des analyses et des apprentissages. Il faut donc distinguer trois fonctions, interdépendantes verticalement, mais pas nécessairement intégrées. Un même ensemble de traces numériques peut donner lieu à la production de différents jeux de données qui, eux-mêmes, peuvent être utilisés de différentes manières pour produire des ressources informationnelles utiles à la coordination ou comme sources de connaissance. La question de la circulation (des traces, des données et des informations) se pose donc à chaque étape de la chaîne de valeur.

L'approche de l'économie de la donnée par sa chaîne de valeur conduit à souligner que l'information et la connaissance produites grâce à des jeux de données ne génèrent finalement de la valeur que si l'on dispose d'organisations et d'opérateurs, capables de s'ajuster et de se transformer, de gérer le changement, d'apprendre et d'innover.

In fine, la réflexion sur la capacité de la donnée à générer de la valeur doit intégrer non seulement l'objet qu'elle concerne, mais aussi les usages qui en sont faits. Certains usages génèrent des gains nets en efficacité. D'autres ne provoquent en réalité qu'une redistribution entre parties prenantes. Cela induit des incitations très différentes à produire ces données et à les partager si la distribution entre coûts et bénéfices n'est pas alignée pour chacun des acteurs. Une des difficultés réside dans le fait que ces caractéristiques ne découlent ni de la nature des données, ni des catégories d'utilisateurs concernés, ni seulement de leurs intentions, mais de chaque situation particulière.

UNE TYPOLOGIE EST-ELLE POSSIBLE ?

On le comprend, la diversité des données, de leur contexte d'utilisation, de la nature des producteurs et des utilisateurs potentiels ainsi que de leurs motivations, conduit à s'interroger sur la possibilité de définir un régime « général » de droits d'accès, d'usage et d'aliénation des données en fonction de leur nature ; ou même sur la pertinence d'une

série de régimes *sui generis* en fonction des différentes catégories de données¹. Les débats juridico-économiques sur le sujet ne permettent pas de trancher. D'un côté, la non-rivalité des données en tant que biens immatériels plaide pour une absence de droits exclusifs. De l'autre, la nécessité d'inciter à créer et à mettre à disposition de tiers des systèmes organisés de données plaide en faveur de l'aménagement de régimes d'appropriation. Les réflexions sur le régime optimal dans différents contextes soulignent l'existence d'un *continuum* allant des données qui devraient ressortir d'une logique de bien commun à celles devant faire l'objet d'une privatisation, tout en ne négligeant pas les risques qui en découlent. En effet, il existe de fortes économies d'envergure dans l'usage des données. Plus large est le portefeuille de données auquel on a accès, plus la valeur d'usage de chaque jeu de données qui le compose augmente. Ceci est la conséquence du fait qu'un jeu de données n'est utile que s'il peut être croisé avec d'autres. C'est ce croisement qui permet de réaliser de l'inférence statistique, et donc à la fois de l'optimisation systémique et de l'analytique.

Dans un contexte marchand, cependant, l'accès à des jeux de données produits par des tiers peut engendrer des coûts de transaction supérieurs à la valeur marginale de chaque jeu de données, alors que leur combinaison permettrait de créer de la valeur. L'idée qu'une démultiplication des transactions peut bloquer la réalisation de transactions collectivement bénéfiques – situation qualifiée d'*anti-commons* (Buchanan et Yoon, 2000 ; Bertichini *et al.*, 2008) – n'est pas nouvelle, et, dans le domaine de la propriété industrielle, elle a donné naissance à la création de *patent pools* et de dérogation les concernant : les propriétaires de brevets complémentaires peuvent les abriter dans un véhicule commun qui permet à un utilisateur potentiel d'un ensemble de brevets, nécessaires à la mise en œuvre d'une technologie donnée, de ne négocier qu'une seule fois les conditions d'accès au portefeuille desdits brevets détenus par des acteurs différents.

L'autre effet sur les régimes d'accès de ces complémentarités entre jeux de données tient à leurs conséquences concurrentielles. En effet, les bénéfices qu'un opérateur peut tirer de l'agrégation de données hétérogènes peut le conduire à des stratégies d'intégration/diversification verticales et horizontales, afin d'accéder à des jeux de données complémentaires. Ce faisant, les optimisations et apprentissages qu'il réalise lui permettent de renforcer sa compétitivité par rapport à la concurrence dans chacune des activités qu'il intègre. Les grandes plateformes numériques représentent l'archétype de ce modèle et de ses conséquences concurrentielles, mais de nombreux autres opérateurs élaborent des stratégies similaires dans la mobilité, la santé, l'agriculture, par exemple. Quoi qu'il en soit, ce type de phénomène peut plaider pour un aménagement des régimes d'accès aux jeux de données de certains opérateurs (*e.g.* Mayer-Schönberger et Ramge, 2018)². Dans le même temps, la nécessité de cette ouverture des données des opérateurs dominants

¹ On peut multiplier la manière de catégoriser les données, par exemple en opposant données personnelles et données industrielles ou commerciales ; ou bien, données générées par les comportements humains et données générées par les systèmes techniques ; ou encore données liées au processus de production, aux transactions, ou aux utilisateurs et à leur retour d'expérience ; etc. L'intérêt de toutes ces classifications est de souligner l'extrême diversité et la très grande hétérogénéité des ensembles de données qu'il convient de prendre en considération quand il s'agit de penser l'économie de la donnée.

² Comme le soulignent Duch-Brown *et al.* (2016), les effets du partage de données sur la concurrence passent par une multitude de canaux qui font que les effets à attendre en pratique d'une obligation de partage ne peuvent être appréciés qu'au cas par cas.

soulève de nombreux problèmes, tels par exemple, la protection de la vie privée ou du secret des affaires, quand il s'agit de données sur les transactions³.

Les débats doctrinaux sur le régime optimal d'accès aux données sont bien évidemment nécessaires, mais leur portée pratique est limitée. En effet, ils aboutissent à l'idée qu'un régime *ad hoc* devrait être défini au cas par cas selon les coûts effectifs de production des données, les bénéfices collectifs espérés du partage, les situations concurrentielles, etc. Ils ne prennent en général pas en considération, au passage, les questions de contrôle de l'accès et de l'usage des jeux de données, de portabilité effective de ces derniers entre opérateurs, ni celles d'aliénation/modification des jeux de données. La donnée peut en effet facilement faire l'objet de modification ou de destruction ; soit par un utilisateur qui n'en n'a plus l'usage, ou a besoin de la modifier, soit de manière malveillante ou à la suite d'erreur de manipulation.

A contrario, l'ensemble de ces problèmes est bien pris en considération par les opérateurs de chaînes de valeur de données. Les *process* et les infrastructures qu'ils opèrent déterminent les qualités des jeux de données, leur transmissibilité entre chaînes de traitement, et, bien entendu, l'utilisation de l'information et de la connaissance générées comme *inputs* dans les *process* de production et de commercialisation. Par ailleurs, les modalités concrètes de circulation des données entre entités juridiques distinctes, ou entre personnes morales et personnes physiques, ressortent pour l'essentiel de la maîtrise d'opérations numériques (contrôle et restriction technique d'accès, cryptographie, etc.) combinées à des instruments juridiques contractuels (même s'ils sont influencés par des cadres de portée générale, tels le secret industriel et commercial ou la protection de la vie privée). *In fine*, c'est avant tout le contrat qui, en pratique, organise la circulation des données ; le producteur du système de données organisées occupant une place centrale dans le dispositif en posant des conditions générales d'utilisation ou en accordant des licences sur mesure⁴. Dans la suite, nous revenons donc sur les différents arrangements concrets qui pourraient être imaginés pour organiser la circulation et le partage des données.

PARTAGER

Revenons brièvement sur les questions de partage des données. Les incitations à ne pas partager les données sont nombreuses comme on l'a évoqué plus haut. Les jeux de données partagés sont coûteux à produire. Par ailleurs, les traitements qui sont réalisés pour anonymiser ou synthétiser les données en diminuent la qualité informationnelle et l'universalité. En outre, les données peuvent être sources d'avantage concurrentiel. Elles sont également des moyens d'extraction de rentes. Enfin, elles sont sources de risques. Il en résulte une logique de décroissance des incitations à partager à mesure que le cercle de bénéficiaires s'élargit. D'un autre côté, plus large est le cercle des bénéficiaires du partage, moins les coûts de protection sont élevés et plus les bénéfices espérés de l'accès à

³ Il existe aussi des arguments relativisant l'idée qu'un opérateur installé bénéficie d'un avantage informationnel et cognitif insurmontable pour les nouveaux entrants, insistant en particulier sur le fait que des apprentissages sur des questions similaires – par exemple la propension de différents types d'utilisateurs à payer pour tel ou tel service, où le prix auquel un prestataire est prêt à céder une capacité non utilisée – peuvent être réalisés à partir de jeux de données différents. C'est effectivement tout l'objet de la science des données que de savoir dériver des informations et des connaissances d'un jeu de données qui n'a pas nécessairement été conçu pour cela.

⁴ Comme dans l'Ouest américain au XIX^e siècle, les capacités de contrôle établissent les capacités à contractualiser, et donc, *de facto*, les droits de propriété (Umbeck, 1981 ; Elickson, 1986 ; Skaperdas, 1992). Tout est dans la combinaison de ressources techniques, humaines et organisationnelles qui permettent concrètement de façonner les données et de maîtriser leur circulation, au prix d'efforts d'adaptation aux besoins des tiers et de sécurisation de ces échanges.

des données complémentaires augmentent. Il existe donc un point d'équilibre potentiel en matière de taille (et de diversité) de la coalition décidant de partager des données.

Notons cependant que les incitations à partager vont aussi dépendre de la complémentarité anticipée, non seulement des jeux de données, mais également des modèles d'affaires d'acteurs impliqués dans le partage. La concurrence frontale, mais aussi à la marge, peut s'opposer à la réalisation des gains du partage. Ainsi, dans le cas de la mobilité électrique, on pourrait voir l'émergence de coalitions concurrentes associant constructeurs, fournisseurs d'énergie et prestataires de services à valeur ajoutée de recharge et d'effacement, plutôt qu'un système intégré de recharge et d'effacement plus fiable et permettant, par exemple, une diminution de la taille des batteries et donc du coût économique et écologique des véhicules électriques, ainsi qu'un usage plus efficace et une conception optimisée du système électrique. En bref, le partage de données entre membres d'un club, d'un consortium, d'une alliance, s'ils permettent de réaliser des gains en efficacité, ne se traduisent pas *ipso facto* par un résultat conforme à l'intérêt général, ni en termes d'efficacité, ni en termes de répartition des bénéfices du partage.

ÉCHANGER

L'organisation de marchés de données se heurte, de son côté, à plusieurs types d'obstacles. Tout d'abord, l'acheteur d'un jeu de données aura une difficulté fondamentale à comprendre la valeur d'usage qu'il achète tant qu'il n'aura pu prendre connaissance de ces données et qu'il n'aura pu les utiliser pour comprendre ce qu'il peut en tirer comme information et connaissance. Ce problème, dit d'antisélection (Akerlof, 1970), est intrinsèque à tout marché informationnel. Il aboutit à une sous-valorisation des biens informationnels par la demande, qui limite sa propension à payer. Des échanges potentiellement bénéfiques n'ont alors pas lieu. Ensuite se pose la question des coûts de transaction déjà évoquée plus haut. Vendre des données nécessite de la part de l'offre des efforts importants de sécurisation de la transaction pour s'assurer que les flux informationnels ne seront pas détournés de leur objet ou revendus à des tiers. Là encore, des échecs d'échanges mutuellement bénéfiques en résultent. Ces échecs de marché sont renforcés par le phénomène des *anti-commons*, évoqué plus haut, qui résulte de l'absence de prise en considération des externalités entre jeux de données par chacun des vendeurs de données. L'ensemble de ces défaillances de l'échange marchand est renforcé par la forte différenciation de la valorisation d'une même donnée par diverses catégories d'utilisateurs ; l'utilité marginale de cette dernière dépendant du potentiel de création de valeur de chacun en fonction de ses usages et de l'accès à des données complémentaires. Il semble au total difficile de garantir un fonctionnement du marché permettant la réalisation de toutes les transactions mutuellement bénéfiques.

ORGANISER L'INTERMÉDIATION

En dépit des apparences, la donnée n'est donc pas un bien collectif pur qui pourrait ressortir d'un régime de partage généralisé. Pour autant, elle n'est pas non plus un bien se prêtant aisément à l'organisation d'un système efficace d'échange marchand. Sous ces deux régimes cardinaux d'organisation de leur circulation, la diversité et la granularité des données disponibles donneraient lieu à des gains en efficacité très inférieurs à leur potentiel. Telles sont les raisons pour lesquelles il convient de dépasser une approche purement « réglementaire » du sujet pour aborder celle de l'organisation de la gouvernance des données. Cette notion couvre bien entendu les règles d'accès, de gestion, d'usage et de partage des informations/de la connaissance issues des jeux de données, mais la dépasse dans la mesure où il s'agit aussi de prendre en considération les relations entre acteurs ainsi que la présence d'intermédiaires fournissant des services allant du traitement technique de la donnée (*e.g.* anonymisation) à la fourniture de services issus de la donnée

(e.g. publicité ciblée, optimisation du système électrique), en passant par le traitement analytique de jeux de données ou l'entraînement d'algorithmes. Il s'agit également de définir les frontières et les modalités de fonctionnement de la communauté concernée (*i.e.* délibération et décision collectives, négociations et règlement des litiges bilatéraux, etc.).

L'industrie est familière de l'organisation de clubs et des régimes que l'on pourrait qualifier de *semi-commons*. Amesse et Cohendet (2001), par exemple, décrivent une forme d'alliance organisée sur le principe de cercles concentriques. Au cœur de l'alliance, les opérateurs les plus engagés participent à des efforts conjoints de recherche et développement, et partagent coûts, bénéfices et droits de décision. À mesure que l'on s'éloigne du cœur, les partenaires sont moins engagés, en situation d'opérateurs plutôt que de codécideurs, et bénéficient avec retard et dans une moindre proportion de la connaissance produite. De telles formes hybrides permettent de réaliser les arbitrages spécifiques aux multiples situations mentionnées plus haut.

La seconde solution est celle du marché intermédié/organisé par un tiers qui fournit des services permettant de sécuriser les partages de données – notamment en n'autorisant pas un accès aux données, mais en opérant, pour des tiers, leur exploitation –, en fournissant des services associés de traitement des données⁵, et surtout, en favorisant la mise en relation entre jeux de données et utilisateurs en fonction des besoins de ces derniers et des potentialités des premiers. Comme dans toute situation d'intermédiation, mais dans celle-ci en particulier compte tenu des importantes économies d'envergure informationnelles évoquées plus haut, il convient de veiller à ce qu'une concurrence effective se maintienne entre ces intermédiaires de données, de manière à éviter qu'ils ne capturent l'ensemble de la valeur et ne soient tentés d'abuser de leur position dominante.

CONCLUSION : GOUVERNER LES DONNÉES

Afin de superviser alliances et intermédiaires, il apparaît essentiel de maintenir une capacité de préserver l'intérêt général en dotant la collectivité de moyens d'observer finement et de débattre de la manière dont les données sont utilisées et des effets produits. La puissance publique doit très certainement se doter de moyens d'intervention, mais il faut aussi cesser de croire que seules des autorités publiques sont à la fois légitimes et compétentes. Les citoyens et les réseaux de compétences indépendants qu'ils peuvent constituer sont aussi indispensables, de même que les capacités des opérateurs privés sont fondamentales pour réfléchir aux alternatives en matière d'accès et d'usage des données, de création de valeur et de répartition des bénéfices. Dans ce contexte, le rôle de la communauté scientifique est aussi crucial. Elle peut et doit fournir aux différentes parties prenantes les moyens d'analyser et mesurer ce qui se passe. Les créateurs et opérateurs de données doivent, quant à eux, être soumis à des obligations de transparence et de vérification des usages qui sont faits des données.

BIBLIOGRAPHIE

AKERLOF G. A. (1970), "The market for 'lemons': Quality uncertainty and the market mechanism", *The Quarterly Journal of Economics*, 84(3), pp. 488-500.

AMESSE F. & COHENDET P. (2001), "Technology transfer revisited from the perspective of the knowledge-based economy", *Research Policy*, 30(9), pp. 1459-1478.

⁵ D'ores et déjà, des opérateurs fournissent de tels services en matière d'analyse de la demande (*marketing analytics*), d'analyse des risques (*rating* ou authentification) ou encore d'identification de prospects ou concurrents), mais on peut imaginer de nombreux autres services à valeur ajoutée. C'est aussi ce type d'approche que Google ou Facebook pratiquent en matière de ciblage publicitaire.

- BERTACCHINI E. DE MOT J. P. B. & DEPOORTER B. (2009), “Never two without three: Commons, anticommons and semicommon”, *Review of Law & Economics*, 5(1), pp. 163-176.
- BUCHANAN J. & YOON Y. (2000), “Symmetric tragedies: Commons and anti-commons”, *Journal of Law and Economics*, 43(1), pp. 1-13.
- DUCH-BROWN N., MARTENS B. & MUELLER-LANGER F. (2017), “The economics of ownership, access and trade in digital data”, JRC Technical Reports, digital economy working paper.
- ELLICKSON R. C. (1991), *Order without Law: How Neighbors Settle Disputes*, Cambridge, Harvard University Press.
- FENNELL L. A. (2011), “Commons, anti-commons, semi-commons”, in AYOTTE K. & SMITH H. E. (éd.), *Research Handbook on the Economics of Property Law*, Cheltenham, Edward Elgar, pp. 35-56.
- MAYER-SCHÖNBERGER V. & RAMGE T. (2018), *Reinventing Capitalism in the Age of Big Data*, Basic Books.
- OECD (2015), “Data-driven innovation: Big data for growth and well-being”, Paris, OECD.
- OECD (2016), “Maximizing the economic value of data: Understanding the benefits and challenges of enhanced data access”, Paris, OECD.
- SKAPERDAS S. (1992), “Cooperation, conflict, and power in the absence of property rights”, *The American Economic Review*, 82(4), pp. 720-739.
- UMBECK J. (1981), “Might makes right: A theory of the formation and initial distribution of property rights”, *Economic Inquiry*, 19(1), pp. 38-59.