

# Le *Data* altruisme : comment les données peuvent-elles être mises à contribution pour servir l'intérêt général ?

Par **Éric SALOBIR**

Président du comité exécutif de la Human technology Foundation et fondateur d'OPTIC

Le volume des données ne cesse de croître, mais pourtant celles-ci sont largement sous-utilisées. Ce paradoxe obère fortement des initiatives œuvrant pourtant au profit de l'intérêt général. Le *Data* altruisme, une innovation en matière de partage des données théorisée par la Commission européenne, pourrait être une solution à ce problème en supprimant la méfiance qui entrave ledit partage. Dans le rapport « Le *Data* altruisme : une initiative européenne, les données au service de l'intérêt général », la Human technology Foundation et l'Exploratoire Sopra Steria Next se sont attachés à montrer comment le concept encore théorique de *Data* altruisme peut devenir une réalité. Dans cet article, nous détaillons les propositions majeures de ce rapport et montrons qu'il est possible de construire un système facilitant grandement la mise à disposition de données pour aider des initiatives œuvrant en faveur de l'intérêt général à se construire.

## Introduction

Le volume des données numériques explose. Ce dernier a été multiplié par 30 entre 2010 et 2020, et l'augmentation à venir est encore plus exponentielle. En effet, les prévisions détaillent une croissance fulgurante qui devrait générer 181 zettaoctets de données, rien qu'en 2025<sup>(1)</sup> (64 zettaoctets ont été créés en 2020). Rappelons qu'un zettaoctet correspond à un milliard de teraoctets ! Les capacités de traitement de ces données vont donc jouer un rôle fondamental pour permettre d'utiliser ce volume toujours plus grand de données.

Pourtant, l'utilisation des données a parfois mauvaise presse en raison de scandales comme les affaires Cambridge Analytica ou IQVIA, lesquels ont montré une exploitation abusive et malveillante de la donnée. Cela génère une méfiance envers le partage des données, ce qui induit une grande sous-utilisation de celles-ci. Ainsi, la présidente de la Commission européenne, Ursula Von der Leyen, parle de « 80 % des données industrielles qui seraient non exploitées ». Ces données sont pourtant précieuses et permettraient de faire de grandes avancées dans le développement d'initiatives

en faveur de l'intérêt général, telles que l'amélioration des services publics, la lutte contre le réchauffement climatique ou l'amélioration de la mobilité.

Cette nécessité a ainsi amené à la création du concept de *Data* altruisme par la Commission européenne dans son Data Governance Act (DGA), un règlement approuvé par le Parlement européen et le Conseil de L'Union européenne à la fin novembre 2021.

Le *Data* altruisme n'est encore qu'un concept théorique. Dans cet article, nous reprenons les suggestions que nous avons formulées dans notre rapport « Le *Data* altruisme : une initiative européenne, les données au service de l'intérêt général », coécrit par la Human technology Foundation et l'Exploratoire Sopra Steria Next<sup>(2)</sup>. Dans un premier temps, nous analyserons le périmètre de ce concept ainsi que ses opportunités et ses limites. Ensuite, nous comparerons plusieurs modes de partage des données existants pour déterminer les meilleures pratiques en vue de l'élaboration de modèles *Data* altruistes. Sur ce point, la Figure 2 de la page suivante présente un récapitulatif des principales mesures que nous recommandons pour rendre concret le *Data* altruisme.

<sup>(1)</sup> « Le Big Bang du Big Data » (2021), Statista, <https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde/>

<sup>(2)</sup> Le rapport « *Data* altruisme : une initiative européenne, les données au service de l'intérêt général », <https://www.human-technology-foundation.org/fr-news/rapport-data-altruisme>

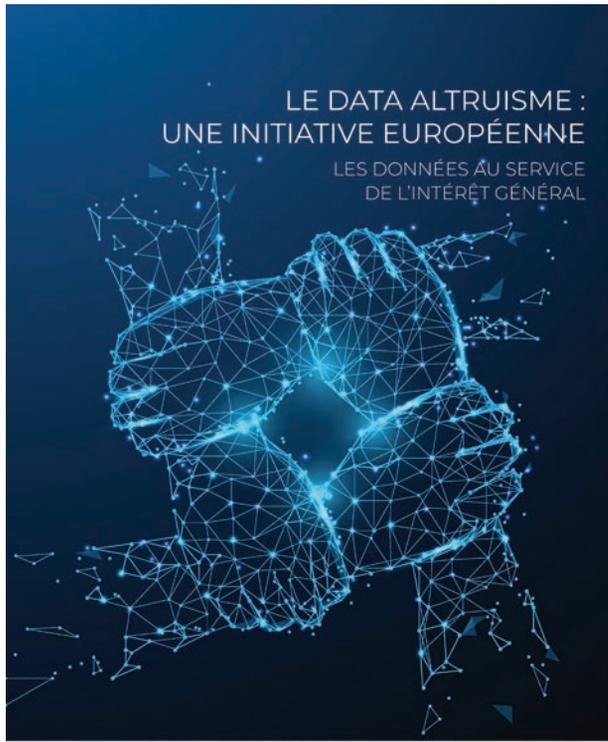


Figure 1 : Couverture du rapport « Le *Data* altruisme : une initiative européenne, les données au service de l'intérêt général ».

## Tout d'abord, qu'est-ce que le *Data* altruisme et comment intervient-il dans le partage des données ?

Le *Data* altruisme est défini par le DGA comme un « partage volontaire de données basé sur le consentement donné [...] ou les autorisations accordées [...] à des fins d'intérêt général, telles que la santé, la lutte contre le changement climatique [...] ou l'élaboration de politiques publiques. »

Il s'agit donc d'un cadre formel de partage des données que décrit le DGA. Pour inciter les acteurs à mettre en œuvre ce partage, le texte autorise des mécanismes de financement pour couvrir la mise en place et la maintenance des infrastructures de partage, ainsi que le recours à des incitations directes ou indirectes pour pousser les contributeurs à partager leurs données.

Il apparaît dès lors nécessaire de définir les différents acteurs impliqués dans un partage de données pour comprendre la nouveauté du *Data* altruisme.

De façon générale, les acteurs sont :

- les sujets (producteurs) des données : par exemple, un individu possédant une montre connectée et qui génère de la donnée *via* son application de sport ;
- le détenteur des jeux de données : par exemple, la société éditrice de l'application utilisée sur la montre connectée ; cette société rassemble les données de tous les utilisateurs de son application ;

### RECOMMANDATIONS PAR ÉTAPE

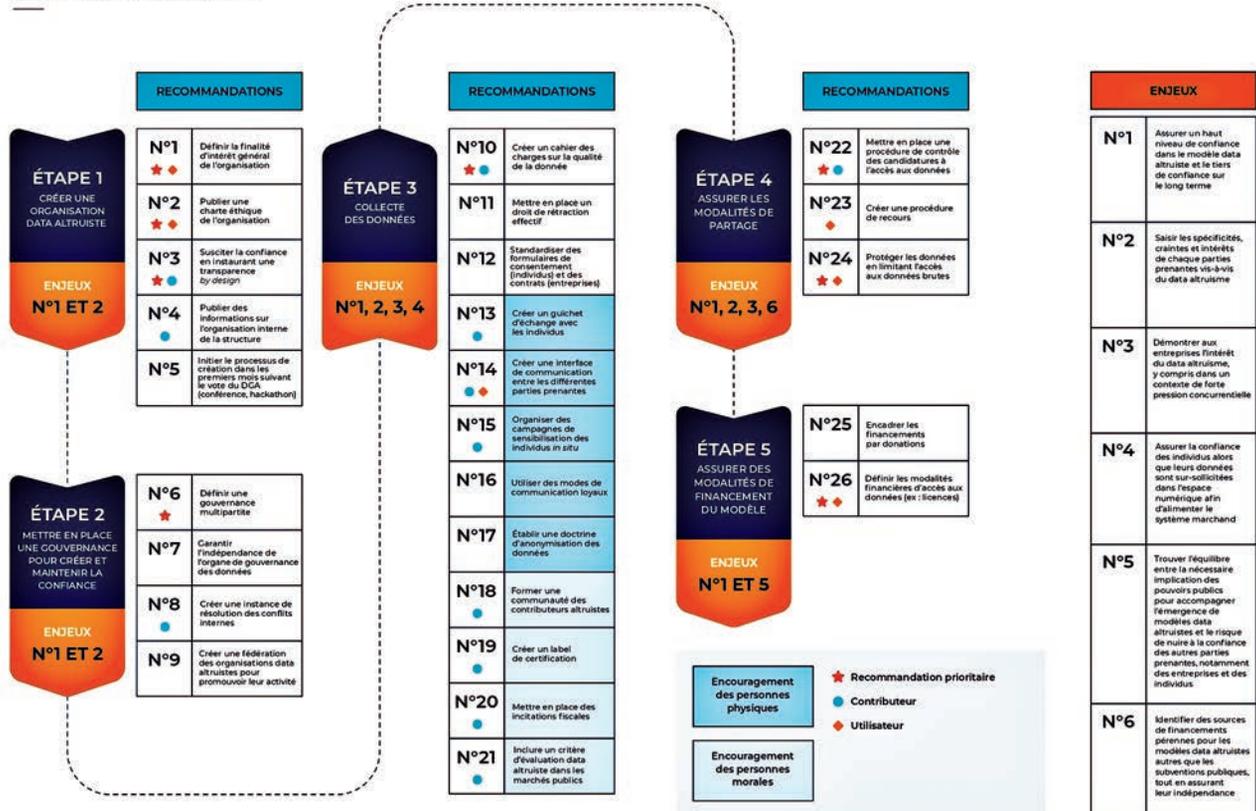


Figure 2 : Principales recommandations du rapport « Le *Data* altruisme : une initiative européenne, les données au service de l'intérêt général ».

- l'utilisateur des jeux de données : ce peut être le détenteur (voir ci-dessus), mais aussi un acteur tiers comme un équipementier sportif qui achèterait au détenteur ses jeux de données pour réaliser des études spécifiques dans le but, par exemple, de créer un nouveau produit ou lancer une campagne publicitaire ;
- l'intermédiaire de données : il s'agit d'un acteur neutre, dont la seule mission est d'opérer techniquement le transfert de données entre un sujet/détenteur et un autre acteur comme un utilisateur ou une organisation *Data* altruiste. Il se contente de faire circuler les données sans les exploiter ;
- Dans le *Data* altruisme introduit par le DGA, il est suggéré la création d'un cinquième acteur qui se placerait entre les détenteurs et les utilisateurs : il serait chargé d'assurer la collecte des données auprès des détenteurs ainsi que le partage de celles-ci avec les utilisateurs, avec le concours de l'intermédiaire de données qui garantit l'opérationnalité des infrastructures de partage.

Ce modèle est innovant, car il s'agit d'une approche spécifiquement européenne qui est radicalement différente de celles adoptées par d'autres grandes puissances comme la Chine ou les États-Unis. L'absence de reconnaissance d'un droit à la protection des données personnelles aux États-Unis ou en Chine comparable au modèle européen explique l'originalité du modèle *Data* altruiste. En effet, ce modèle découle des choix européens faits en matière de protection des données, il cultive donc une différence certaine par rapport aux autres systèmes dominants. La Chine a fait le choix d'un contrôle étroit permettant une supervision des transferts de données. Les données sont considérées comme un bien marchand et peuvent être échangées à ce titre sur le Shanghai Data Exchange ; cette approche est à l'opposé d'une vision altruiste et est en contradiction avec les règles de protection des données personnelles développées dans l'Union européenne. Les États-Unis, quant à eux, s'appuient sur le principe d'extraterritorialité qui s'applique à leurs réglementations et à leurs multinationales pour collecter auprès d'entreprises ou d'individus étrangers leurs données personnelles. Cette approche est impossible pour l'Europe qui ne dispose pas, contrairement aux États-Unis, de géants des données justifiant d'un quasi-monopole ; cela explique pourquoi la Commission européenne a cherché à promouvoir la stratégie d'un « pas de côté » pour rattraper le retard des pays européens en matière de captation et d'exploitation des données face aux deux modèles hégémoniques que sont la Chine et les États-Unis.

L'innovation vient également du but poursuivi par ce système altruiste, à savoir œuvrer en faveur de l'intérêt général. Il convient donc de définir au préalable cette notion « d'intérêt général » pour pouvoir pleinement appréhender le périmètre attribué au *Data* altruisme.

« L'intérêt général » n'a pas de définition légale officielle, mais certaines de ses caractéristiques sont détaillées dans différents textes. Ainsi, en droit fiscal français, l'intérêt général se conçoit comme une activité non lucrative à caractère philanthropique qui n'est pas mise en œuvre au profit d'un cercle restreint de personnes,

et dont la gestion est désintéressée. Sa finalité transcende les intérêts individuels et permet aux individus de participer à la société. Cette notion incite à créer un projet commun incarné par un intérêt supérieur qui permettrait à chaque individu d'apporter sa contribution à la collectivité.

Toutefois, l'approche du *Data* altruisme reste une approche nouvelle et doit donc résoudre un certain nombre de défis pour pouvoir être acceptée par le plus grand nombre.

Le premier défi est celui d'obtenir la confiance des différents acteurs. Plusieurs scandales ont fortement fragilisé la confiance des parties prenantes envers le partage des données. Un chantier sur la confiance est nécessaire ; cela implique de réussir à trouver les bons arguments pour convaincre, notamment en insistant sur les modalités de sécurisation des données, en rappelant la finalité du partage, son caractère altruiste, en décrivant les institutions garantes de la bonne exécution de ce partage, tout en prenant en compte les spécificités culturelles de chacun des pays de l'Union européenne (UE).

Le deuxième défi est l'indépendance des organisations *Data* altruistes, et donc la nécessité de s'assurer qu'elles disposent de sources de financement pérennes.

En effet, nombreux sont les coûts liés à l'activité *Data* altruiste :

- la collecte des données ;
- le stockage de celles-ci ;
- l'enrichissement des données : ajout d'informations supplémentaires ;
- la standardisation : ramener à un même format des données venant de différentes sources ;
- combler les manques d'informations (avec des données pivot, par exemple).

Un modèle de redevances proportionnées ou de frais administratifs pour pouvoir avoir accès aux données mises à disposition par l'organisation est une solution viable prévue par le DGA. En plus de ce modèle économique, il est tout à fait possible d'avoir recours à des financements extérieurs (mécénat, subventions, soutien technique) à condition de garantir que l'organisation maintiendra son indépendance et que les contributeurs ne seront pas exclus de la création de valeur.

## Quelles bonnes pratiques parmi les modes de partage existants pourraient servir au développement du *Data* altruisme ? Mais il faut être prudent au regard des dérives liées à ce partage

Dans notre rapport, nous étudions 56 initiatives qui sont divisées en quatre secteurs (santé, transports, environnement et gestion des informations personnelles) afin de comprendre l'écosystème du partage des données. Les modes de partage novateurs étudiés montrent que plusieurs formes de *Data* altruisme sont possibles, avec des orientations tout à fait différentes, mais aussi

qu'un certain nombre de risques existent et menacent la confiance sur laquelle repose tout le mécanisme. Les systèmes étudiés nous renseignent en outre sur les bonnes pratiques à transposer dans la démarche *Data* altruiste ; ces dernières sont détaillées pour chaque secteur.

### Enseignements à tirer des dynamiques observées dans le secteur de la santé

Le domaine de la santé est marqué par une quantité importante de données notamment du fait de l'existence de systèmes de sécurité sociale dans de nombreux pays européens. Toutefois, ce volume important ne signifie pas que les données soient facilement utilisables, car cette exploitation se heurte à deux enjeux majeurs : la qualité des données et la question de la sécurité de cette exploitation. Il faut en effet parvenir à avoir des données utilisables qui soient représentatives de tous les cas rencontrés (la qualité), tout en garantissant que ces données ne permettent pas d'identifier un individu (la sécurité).

Le problème crucial pour le *Data* altruisme dans le secteur de la santé serait donc la difficulté à décrire une finalité en amont du projet pour inciter au partage. En effet, la recherche médicale se caractérise par l'existence de besoins non identifiés. Comment alors arriver à convaincre les contributeurs si l'on ne peut définir en amont une finalité suffisamment précise et compréhensible ? Les questions de gouvernance évoquées précédemment sont également cruciales pour rassurer la grande variété des acteurs de la santé dont les intérêts sont souvent différents : l'Assurance maladie, les hôpitaux, les pharmacies, les laboratoires, etc.

### Enseignements à tirer des dynamiques observées dans le secteur des transports (la mobilité)

Le secteur de la mobilité est en avance dans le domaine du *Data* altruisme grâce à l'existence de diverses initiatives et propositions émanant des pouvoirs publics. Ainsi, la notion de donnée d'intérêt territorial évoquée dans un rapport parlementaire de 2017<sup>(3)</sup> aborde déjà l'idée de la transmission des données générées dans le cadre privé vers la sphère publique pour des finalités d'intérêt général. Parmi les initiatives privées, peuvent être cités Strava Metro ou Uber Movement, qui visent à aider les acteurs de la mobilité à améliorer les infrastructures de transport.

Toutefois, ce secteur rencontre des difficultés similaires à celui de la santé, notamment en ce qui concerne la qualité des données. En effet, les différents jeux de données disponibles sont complexes à compiler car ils sont très hétérogènes, notamment dans leur structuration.

Les dynamiques observées dans la mobilité sont en revanche particulièrement portées par les réglementations, comme la portabilité citoyenne qui oblige les acteurs du secteur à transmettre les données générées

par leurs utilisateurs à d'autres acteurs si l'un de ces utilisateurs le demande. Ce mécanisme permet de constituer de nouvelles bases de données et d'améliorer la circulation des données sous le contrôle des citoyens.

### Enseignements à tirer des dynamiques observées dans le secteur de l'environnement

Deux sous-cas ont été étudiés.

#### L'énergie

Le recours croissant à des moyens de production non pilotables (solaire, éolien...) a provoqué un plus grand besoin en données pour la production et la distribution de l'électricité. Il faut en la matière parvenir à concilier la demande et l'offre à une granularité de plus en plus fine en tenant compte des aléas propres à ces moyens de production. Le développement des *smart cities* requiert également des systèmes permettant l'envoi et le traitement de données en temps réel, à une maille fine.

#### La biodiversité

Le *Data* altruisme serait un précieux atout pour la collecte des données portant sur des surfaces et des temporalités longues. Il aiderait en particulier le tiers secteur de la recherche en permettant à un ensemble d'acteurs très divers (associations, collectivités, groupements professionnels...) de pouvoir plus facilement transmettre leurs données.

### Enseignements issus d'une perspective canadienne dans le secteur de la gestion des informations personnelles : la fiducie des données

Dans son rapport « Tables de stratégies économiques du Canada » publié en 2018, le gouvernement canadien définit la fiducie de données comme « un organisme formé pour gérer des données pour le compte de ses membres. Ces derniers mettent en commun leurs données et conviennent expressément des conditions afférentes à leur partage ». Il s'agit donc d'un intermédiaire fiable pour le traitement des données s'apparentant au modèle développé en Europe. Cette initiative des fiducies a été encouragée par les pouvoirs publics, par exemple, à travers la Charte numérique du Canada de 2019. Mais ce concept n'a pas encore été réellement adopté. En effet, il manque encore un support clairement défini dans les lois canadiennes pour que ce genre d'entité puisse être créé. Un des échecs connus dans la mise en place de ces fiducies de données est celui de l'Urban Data Trust développé par Sidewalk Labs. Une forte défiance citoyenne, des complications juridiques et une structure parfois ambiguë du fait de son appartenance à Alphabet Inc. ont fait que le projet n'a finalement jamais vu le jour.

La leçon de cet échec est qu'une fiducie de données ne soulève pas uniquement des enjeux technologiques et de sécurité, dès lors que ce modèle vise aussi à bâtir une confiance réciproque entre tous les participants. Cet exemple canadien illustre la nécessité d'avoir un cadre réglementaire et légal clair, notamment en ce qui concerne la protection de la vie privée, pour que les

<sup>(3)</sup> BELOT Luc (2017), « De la *smart city* au territoire d'intelligence(s) – L'avenir de la *smart city* ».

intermédiaires comme les fiduciaires de données puissent être considérés comme fiables dans le partage des renseignements d'ordre personnel.

Tous les systèmes de partage étudiés ont également permis de déceler les principaux risques que rencontrerait le *Data* altruisme s'il était mis en pratique :

- la réparation : des entreprises ayant mené des actions néfastes pourraient participer à une démarche *Data* altruiste pour camoufler les dommages qu'elles ont générés en profitant du statut altruiste de l'organisation créée en accord avec les règles du DGA ;

- le cheval de Troie : il s'agit ici d'utiliser les mécanismes du *Data* altruisme de façon malveillante pour capturer les données stratégiques détenues par un concurrent par exemple (cela peut être réalisé de façon directe avec le piratage des serveurs d'une organisation *Data* altruiste ou alors de façon indirecte sous couvert d'un projet *Data* altruiste pour appâter les potentiels contributeurs) ;

- la tromperie : la démarche *Data* altruiste est ici mobilisée pour collecter des données personnelles, notamment celles de consommateurs.