



BANQUE des  
**TERRITOIRES**



# Données personnelles: le modèle d'après

État des lieux de  
l'écosystème du self data



Alors que l'exploitation des données est un enjeu primordial, tant pour les entreprises que pour les administrations, la question de la place de l'individu et de la protection des données personnelles est un élément central de ce débat. Le Règlement Général pour la Protection des Données (RGPD), entré en vigueur en mai 2018 a défini un cadre pour la collecte et l'exploitation des données personnelles, créant de nouveaux droits pour les citoyens européens.

La combinaison des obligations de recueil du consentement, de droit à la portabilité des données personnelles, et du principe de minimisation de

collecte des données personnelles sont même les piliers d'un modèle de circulation des données personnelles, alternatif au modèle actuel reposant largement sur la monétisation des données et la publicité ciblée. En France, ce modèle porte le nom de « self data » et se définit comme « la production, l'exploitation et le partage de données personnelles par les individus, sous leur contrôle et à leurs propres fins » (définition donnée par la Fondation Internet Nouvelle Génération).

Avant même l'entrée en vigueur du RGPD, des initiatives privées, publiques et citoyennes se sont créées pour proposer des outils et explorer des cas d'application du self data. Cet écosystème d'acteurs se compose principalement :

**Des détenteurs de données :** certains ont mis en place des canaux de transmission de données, permettant l'export des données personnelles vers les individus qu'elles concernent, lorsqu'ils le demandent. C'est la base technique de la portabilité des données personnelles.

Ces canaux de transmission ont été mis en place dans le secteur bancaire et ont permis le développement des services d'agrégateurs bancaires (comme Linxo, Banking,...). Enedis s'est également doté d'API pour permettre la portabilité des données Linky; à ce jour, une quarantaine d'entreprises s'y connectent pour fournir à leurs clients des conseils sur leur consommation énergétique. La portabilité est également utilisée pour faciliter l'authentification d'une personne, par exemple avec son numéro de téléphone (Mobile ID), de manière à lui éviter de ressaisir toutes ses informations.

Toutefois, la plupart des détenteurs de données restent réticents à mettre en place de tels canaux de transmission.

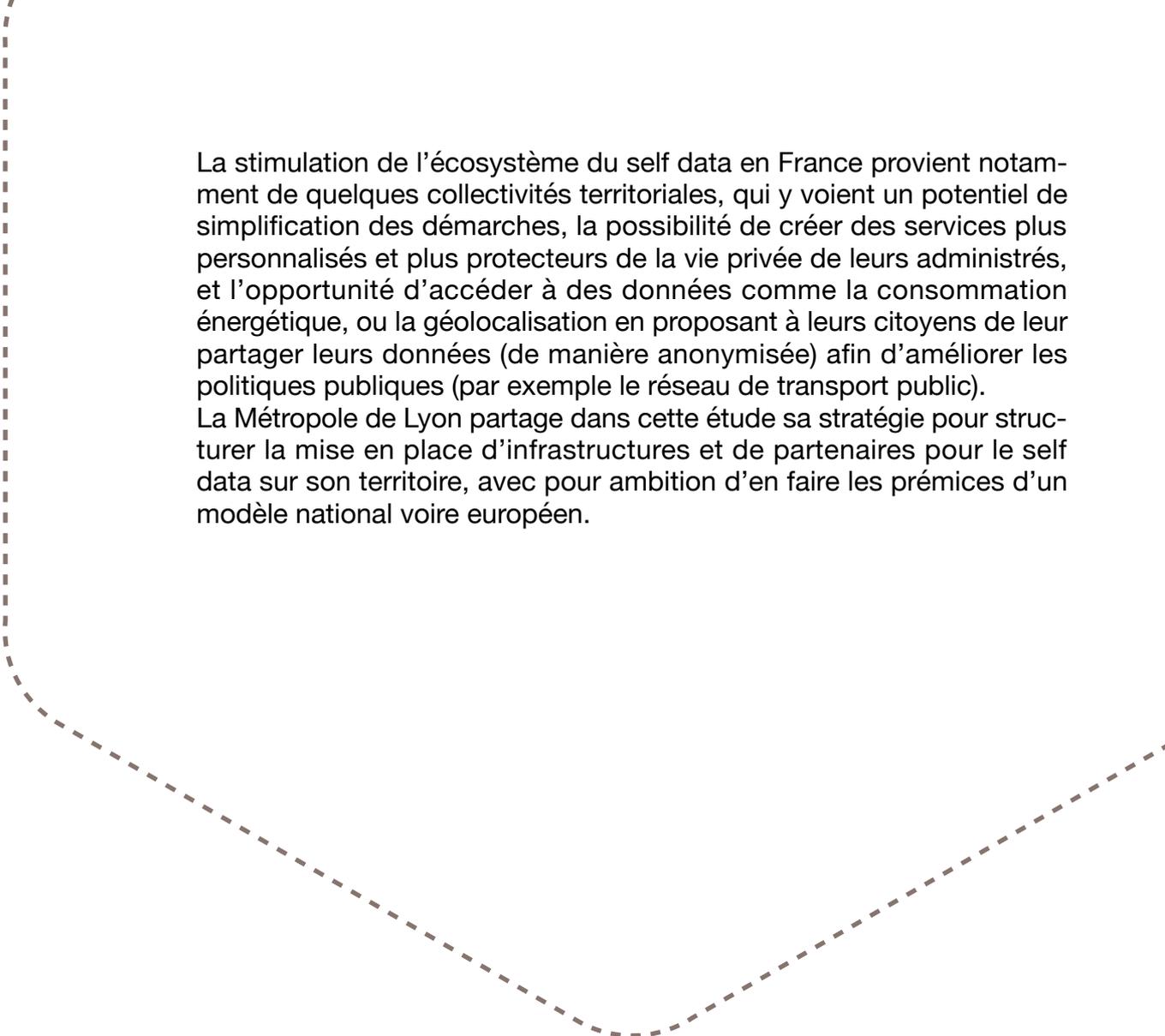
**Des fournisseurs d'entrepôts personnels de données (ou clouds personnels) :** proposés par des startups, ces espaces personnels permettent de stocker des données personnelles, notamment celles récupérées suite à une demande de portabilité. Pour le moment, ce sont davantage les coffre-forts numériques (stockage de documents personnels) qui se développent sur le marché B2B2C (stockage de fiches de paie mis à disposition du salarié par l'employeur).

**Des gestionnaires de consentements :** également développés par des start-ups, les gestionnaires de consentement permettent à l'individu de donner ou retirer leur consentement à partager telle ou telle donnée avec un tiers, depuis une seule interface et de manière simple.

**Des fournisseurs de services réutilisant les données personnelles tout en protégeant la vie privée :** plusieurs techniques éprouvées permettent aujourd'hui de réaliser des calculs sans jamais avoir accès aux données personnelles. D'autres services basés sur la portabilité des données personnelles existent tout en n'utilisant pas ces techniques.

Plusieurs enjeux expliquent le développement encore lent du self data :

- Le développement de cas d'usage attractifs pour les usagers, qui permette d'attirer des clients vers cette logique de réappropriation des données personnelles ;
- La définition de modèles économiques viables et adaptés à chacun des quatre types d'acteurs, avec une répartition de la valeur créée par le nouveau service ;
- La mise en place de capacités de transfert et d'exploitation des données entre les parties prenantes (standards techniques, compétences numériques, etc.) dans un cadre sécurisé.
- La finalisation un cadre de protection juridique et réglementaire adapté pour sécuriser les problématiques liées aux responsabilités dans la continuité du RGPD et en prenant appui sur les expériences sectorielles déjà en place.

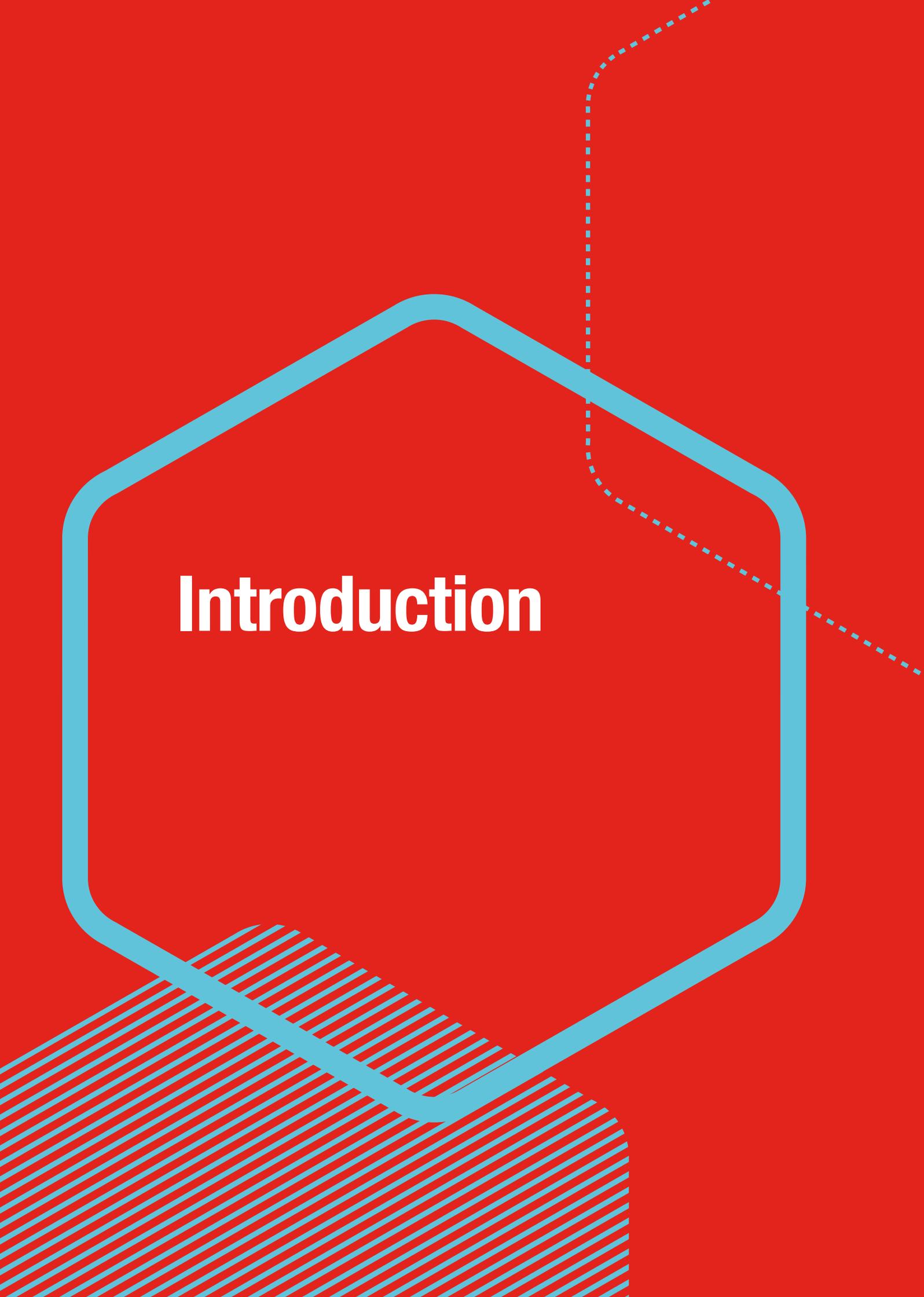


La stimulation de l'écosystème du self data en France provient notamment de quelques collectivités territoriales, qui y voient un potentiel de simplification des démarches, la possibilité de créer des services plus personnalisés et plus protecteurs de la vie privée de leurs administrés, et l'opportunité d'accéder à des données comme la consommation énergétique, ou la géolocalisation en proposant à leurs citoyens de leur partager leurs données (de manière anonymisée) afin d'améliorer les politiques publiques (par exemple le réseau de transport public). La Métropole de Lyon partage dans cette étude sa stratégie pour structurer la mise en place d'infrastructures et de partenaires pour le self data sur son territoire, avec pour ambition d'en faire les prémices d'un modèle national voire européen.

# SOMMAIRE

<b>I. Le contrôle des données personnelles : un enjeu économique et politique</b>	<b>13</b>
I.1 Un modèle actuel qui montre des signes d'essoufflement	14
I.2 Décider soi-même de l'usage de ses données personnelles, un enjeu collectif	16
I.3 Les collectivités également concernées par le self data	18
<b>II. Une nouvelle logique de partage de données personnelles émerge</b>	<b>19</b>
II.1 Les droits de l'individu, au cœur du modèle du self data	20
II.2 Les principes de la circulation des données décrits par les régulateurs	24
II.3 Les détenteurs de données plutôt réticents à la portabilité	27
<b>III. Des modalités techniques définies et déjà mises en œuvre dans quelques secteurs</b>	<b>29</b>
III.1 Plusieurs techniques sont mises en œuvre pour la transmission des données	30
III.2 Différentes techniques pour fournir des services numériques basés sur les données de manière plus efficace et moins intrusive	33
III.3 Deux axes pour analyser les services self data : richesse des données et degré de protection des données.	35
<b>IV. L'écosystème d'acteurs se développe lentement</b>	<b>37</b>
IV.1 Quelques services sur le marché basés sur des données accessibles à la portabilité	38
IV.2 Les Personal Information Management Systems, un marché d'infrastructure encore embryonnaire	43
IV.3 Des écosystèmes d'acteurs engagés pour lever les freins au développement du self data	47
<b>V. Des ambitions territoriales pour le self data : entre coûts et bénéfices pour les collectivités</b>	<b>51</b>
V.1 Des bénéfices certains mais difficilement quantifiables	53
V.2 La collectivité peut s'impliquer à plusieurs niveaux	55
V.3 Le passage à l'échelle : une démarche en plusieurs étapes	57

<b>Fiches techniques</b>	<b>60</b>
1.1 Construire une API pour repartager les données personnelles, un coût abordable mais qui nécessite la mise à niveau du système d'information	62
1.2 Méthodologie de l'évaluation du marché des intermédiaires	64
1.3 Les techniques pour développer des services plus respectueux de la vie privée : privacy by design, privacy by default, calcul à la source et preuve à divulgation de connaissance nulle	66
<b>Lexique</b>	<b>70</b>
<b>Bibliographie</b>	<b>73</b>



# Introduction

Dans un contexte de numérisation croissante des activités et des services, les individus se sentent de plus en plus menacés et insécurisés par l'utilisation de leurs données personnelles. Une majorité d'individus se sent en position de faiblesse vis-à-vis des acteurs du numérique et exprime une volonté de renverser le rapport de force en faveur de modèles fondés sur le respect de la vie privée. Ainsi, il semblerait que les modèles traditionnels exploitant massivement les données personnelles perdent de la vitesse. De nombreux signaux de changements sont perceptibles : usage accru des adblockers, fin programmée de l'utilisation des cookies tiers....

Le cadre réglementaire, par la voie du RGPD est venu appuyer ce mouvement en augmentant significativement les droits individuels et en renforçant le principe d'autodétermination informationnelle. Le droit à la portabilité des données personnelles fait partie de cet éventail de nouveaux droits. Il rencontre toutefois de nombreuses limites à sa mise en œuvre aussi bien par les entreprises (récupération partielle des données, difficultés techniques, manque de connaissance et peu d'incitation à l'action...) que par les individus pour qui ce droit est assez mal compris et difficile à revendiquer, en l'absence d'outil pour le faire.

Le self data est un concept, qui propose un modèle de circulation et d'exploitation des données personnelles replaçant l'individu au centre. Il peut se définir comme «**la production, l'exploitation et le partage de données personnelles par les individus, sous leur contrôle et à leurs propres fins**».

Les bases techniques et juridiques d'un tel modèle existent :

- ***la portabilité des données personnelles***, transcrite comme un droit par le RGPD ; ce droit est activable dès lors qu'un moyen (une API, un contact mail, un formulaire) est proposé pour demander et recevoir ses données dans un format lisible par machine. Certains acteurs se positionnent en tierce-partie pour gérer, à la demande des individus, leurs requêtes de portabilité. La portabilité se fait «**à la demande et sous le contrôle de la personne**».
- ***les données personnelles concernées par le droit à la portabilité sont celles qui sont collectées sur la base légale du consentement (défini par le RGPD) ou du contrat***
- ***la minimisation de la collecte de données personnelles*** (ou principe de privacy by default), qui repose sur des techniques de calcul permettant de minimiser voire d'éviter de traiter des données personnelles pour la fourniture d'un service personnalisé.

**La diffusion de ces outils, de ces pratiques et de ces méthodes reste marginale mais la réglementation est encore récente :**

——  *dans le secteur bancaire,* une législation avait déjà imposé la portabilité des données personnelles : la disponibilité de ces données a permis la création de nouveaux services, les agrégateurs bancaires qui compte plus de 4 millions d'utilisateurs en France.

——  *d'autres cas d'usages issus de la portabilité* des données personnelles se développent : les services de coaching énergétique sur la base de l'API d'Enedis, et les services d'authentification basées sur l'identité numérique (comme le numéro de téléphone).

——  *dans le secteur de la mobilité,* la centralisation des données par l'individu dans un Compte personnel de mobilité est particulièrement adaptée à la mise en œuvre du MaaS et fait l'objet de travaux expérimentaux.

**La création de canaux de transmission de données personnelles vers les individus est donc un levier important pour le développement de cas d'usage utiles mais elle fait encore l'objet de réticences de la part de la majorité des détenteurs de données.**

En l'absence de données disponibles et de cas d'usage suffisamment nombreux, les offres de clouds personnels peinent encore à s'imposer comme infrastructure de gestion des données. Toutefois, la stratégie européenne de la donnée publiée en février 2020 veut promouvoir le développement ces outils.

**Des acteurs réseaux d'acteurs soutiennent le développement de l'écosystème:** par l'expérimentation (collectivités locales, grandes entreprises), par le financement (fonds spécialisés ou non), par le lobbying et l'animation des réseaux d'acteurs (MyData, a New Governance, Data Transfer Project,...).

Finalement, le fonctionnement du self data repose sur l'entente entre quatre types d'acteurs pour lever les freins économiques, techniques et juridiques :

—— **Les individus** occupent une position centrale, leur consentement est indispensable à tout partage de données et ce sont eux qui sélectionnent les services qui ont accès à leurs données. C'est aussi leur usage des nouveaux services qui conditionne le succès de ces applications.

—— **Les détenteurs de données** sont les entreprises qui collectent les données personnelles, ils n'ont en théorie pas de pouvoir sur les données puisque la loi les contraint dans ce domaine. Mais en pratique, les évolutions efficaces ne peuvent pas se faire sans eux. Nombre de leurs inquiétudes sont légitimes et leurs réticences montrent qu'ils ne valorisent pas pleinement les données dont ils disposent. Certains ont néanmoins ouvert des canaux de partage plus ou moins automatisés de données.

—— **Les tiers ré-utilisateurs de données** sont des entreprises qui proposent de nouveaux services aux individus grâce à la portabilité de leurs données. Ils ne sont généralement pas concurrents des détenteurs de données mais plutôt complémentaires de leur offre. Ces services peuvent apporter de la valeur aux données des individus sous de nombreuses formes : optimisation de services existants, offre de nouveaux services personnalisés ou usages contributifs pour les besoins d'une communauté.

- **Les intermédiaires** sont des entreprises qui fonctionnent comme des sous-traitants des individus. Qu'ils soient des gestionnaires de consentement ou des entrepôts de données personnelles, leur fonction est de rendre techniquement opérationnelle la volonté du citoyen sur le partage de ses données personnelles. Ce marché est contraint par sa capacité à se développer massivement, selon la logique d'infrastructure et de plateforme : la communauté d'utilisateurs à laquelle ils donnent accès déterminera l'attractivité pour les tiers ré-utilisateurs de données.
- **Les collectivités territoriales** peuvent jouer un rôle dans la diffusion de ce modèle, tout en y trouvant des bénéfices : renforcer le contrôle des citoyens sur leurs droits, innover dans la gestion de la relation avec les administrés pour une meilleure gestion de la collectivité et des objectifs de politiques publiques. En France, les expérimentations de la Métropole de Lyon et de l'Agglomération de La Rochelle permettront de tester l'adoption des services self data, la soutenabilité du modèle, et les perspectives de valorisation pour les acteurs privés (détenteurs de données et tiers-ré-utilisateurs).

Cette étude repose sur les témoignages de plus de 25 professionnels du monde de la manipulation des données personnelles, parmi lesquels, des DPO, créateurs d'entreprises, acteurs du secteur public, en France mais aussi à l'international et plus de 450 acteurs technologiques identifiés à travers le monde. L'analyse étudie aussi des cas d'usage matures, et d'autres en cours d'expérimentation ou à venir : initiatives de la Métropole de Lyon (« mon parcours social », Energies multi fluides), de La Rochelle (Mon Coach CO2).



# **I. Le contrôle des données personnelles : un enjeu économique et politique**

# I.1.

## Un modèle actuel qui montre des signes d'essoufflement

Les données personnelles sont aujourd'hui devenues la ressource primaire d'un véritable secteur économique : elles sont collectées en masse, font souvent l'objet de traitement et de croisement pour être ensuite vendues à des plateformes dont la principale fonction est le ciblage publicitaire. De ce marché, les individus ne voient qu'une face : les publicités toujours plus personnalisées qui les incitent à acheter toujours plus de produits.

Mais cette personnalisation de masse et divers scandales, comme l'affaire Cambridge Analytica, ont intensifié la méfiance des individus à l'égard des géants du numérique : en 2018, seuls 3

français sur 10 considéraient que la confidentialité de leurs données personnelles sur internet était assurée <sup>(1)</sup>.

Des signaux faibles indiquent une attention de plus en plus importante à ces questions : 54% des internautes déclarent avoir déjà utilisé un bloqueur de publicités (ou adblocker). Pourtant, très peu ont changé leurs habitudes pour renforcer la protection qu'ils appellent de leurs vœux. Ils sont encore à 92% à utiliser Google comme moteur de recherche alors que 77% se déclarent inquiets concernant la manière dont les GAFAM utilisent leurs données personnelles. Cela s'explique en partie par le fait que les services numériques les plus performants sont ceux qui disposent du plus grand nombre d'utilisateurs.

<sup>(1)</sup> « Les français et les données personnelles »  
[www.bva-group.com/sondages/francais-donnees-personnelles/](http://www.bva-group.com/sondages/francais-donnees-personnelles/)

L'encadrement juridique de cette économie numérique a été adapté pour mieux protéger la vie privée et éviter les monopoles des grandes plateformes (Loi République Numérique, RGPD). La CNIL œuvre sans relâche pour le respect des droits des individus en ce qui concerne les données personnelles, avec des sanctions record : 50 millions d'euros pour Google en janvier 2019 pour « manque de transparence, information insatisfaisante et absence de consentement valable pour personnalisation de la publicité. »

Ceci pousse les géants du secteur à s'auto-réguler, après Safari (Apple) et Firefox (Mozilla), Chrome, le navigateur de Google, a annoncé la suppression des cookies tiers pour début 2022. Ces cookies permettent essentiellement le re-ciblage publicitaire, une pratique très mal perçue par les utilisateurs consistant à leur proposer des publicités ciblées sur un achat qu'ils ont commencé (mais sans effectuer l'achat) afin de les inciter à finaliser cet achat. Pourtant, la fin de ces cookies tiers pour Google signifie aussi l'éviction des « petites » entreprises de la publicité au profit des géants comme les GAFAM, qui concentrent le plus de données.



Aujourd'hui l'accès à la donnée est plus compliqué qu'avant le RGPD pour les départements marketing, les régies publicitaires et les Data brokers. La donnée n'a de la valeur que si on a le consentement explicite de la personne auprès de laquelle elle est collectée.

-

**Christian-François Viala,**  
*Expert indépendant RGPD et éco-système Data*

## I.2.

# Décider soi-même de l'usage de ses données personnelles, un enjeu collectif

C'est dans cet écosystème presque hostile à l'utilisateur que s'est développé le concept de Self data.

Les premières initiatives viennent des États-Unis avec le concept de Vendor Relationship Management (VRM), qui vient s'opposer au Customer Relationship Management (CRM).

Le principe est simple: au lieu de donner la responsabilité de la gestion des données aux entreprises qui

administrent une base de clients, le VRM propose d'inverser la logique. Ce sont les individus qui stockent et administrent leurs données et qui y donnent accès (ou non) aux entreprises. En adaptant cette logique au secteur de la santé, le gouvernement américain a développé un service, le Blue Button. Il permet à chacun d'accéder à ses données de santé et de les visualiser. Initialement réservé aux vétérans à son lancement en 2010, il a été ouvert à tous les américains et rencontre un certain succès <sup>(2)</sup>.

En France, c'est la Fondation Internet Nouvelle Génération (FING) qui a développé le concept de Self data autour d'une idée: «la production, l'exploitation et le partage de données personnelles par les individus, sous leur contrôle et à leurs propres fins » à partir de 2012.

À travers cette définition, le centre de gravité se situe au niveau de l'individu, à l'opposé des effets de réseau <sup>(3)</sup> qui ont fait la force des grandes plateformes. En Europe, le RGPD, dans son article 20, a créé un droit à la portabilité des données personnelles pour les individus, c'est-à-dire le droit de demander un accès aux données qui les concernent. Mais ce droit n'est aujourd'hui que rarement utilisé.

<sup>(2)</sup> Depuis 2010, selon un rapport de la Maison Blanche, 5,6 millions de vétérans auraient téléchargé leurs données

<sup>(3)</sup> L'effet de réseau est le phénomène par lequel l'utilité réelle d'une technique ou d'un produit dépend de la quantité de ses utilisateurs.

Pourtant, une transition de l'économie des données vers un modèle de self data porte plusieurs avantages :

## Au niveau individuel

- **Disposer pleinement des moyens pour activer ses droits numériques:** les droits que crée le RGPD sont encore trop peu activés par les individus car il n'existe pas encore assez d'offre adéquates.
- **Proposer des services réellement personnalisés aux consommateurs et aux citoyens:** le conseil sur les pratiques (un conseiller mobilité, bancaire, sportif, culinaire, etc. par exemple : des suggestions de recettes en fonction de ses achats); des comparateurs basés sur les habitudes réelles de consommation (le contrat de fourniture d'énergie, d'assurance, de téléphonie, ... le plus adapté à sa consommation et à son profil)
- **Simplifier la complétion des informations grâce aux données existantes,** dans une démarche maîtrisée par l'utilisateur (par exemple: compléter automatiquement un formulaire qui permet de connaître les aides énergétiques adaptées à mon profil et mon logement, connaître les aides auxquelles j'ai droit)

## Au niveau économique

- **Remettre en cause les monopoles sur lesquels reposent les grandes plateformes:** c'est l'argument privilégié par l'Union Européenne dans sa stratégie pour la donnée publiée en février 2020. Elle postule que le fait de casser le monopole de certains acteurs sur l'accès aux données des européens permettra de favoriser la concurrence et rééquilibrera le rapport de forces entre l'économie numérique européenne et américaine.
- **Un effet collatéral sur l'économie de la donnée:** la moindre dépendance aux plateformes pour accéder aux données permettrait aux acteurs publics et parapublics de retrouver une capacité de dialogue avec les citoyens, et aux acteurs privés avec leurs clients, pour accéder à des données (par exemple les données de géolocalisation, qui, partagées à la collectivité de manière anonyme, peuvent alimenter une meilleure connaissance des usages de mobilité). Cela impacterait donc à la baisse l'achat de données car ces données seraient mises à disposition par les individus.

## I.3.

# Les collectivités également concernées par le self data

Le self data permet de créer des canaux responsables et justes de transmission des données. De plus, les données transmises par ces canaux sont de meilleure qualité que celles que les différents acteurs des marchés des données peuvent s'échanger. C'est donc une alternative majeure qui pourrait permettre, si elle est suffisamment adoptée, la suppression des canaux de transmission actuellement aux dépens des utilisateurs.

Ceci permet l'émergence d'un véritable citoyen numérique, doté de droits numériques dont il peut véritablement faire usage puisqu'il en a les moyens à disposition. Mais il faudra alors passer de la capacité à l'action, d'un droit à un fait. Les individus pourront utiliser leurs données de manière simple et s'en serviront. Ce mouvement leur permettra

de prendre conscience de ce que sont ces données personnelles, de l'ampleur du champ qu'elles recouvrent. Il y a donc là un puissant levier de compréhension des enjeux numériques rapportés au citoyen, et une solution pour proposer un numérique plus transparent, un modèle vecteur de confiance. Les collectivités locales peuvent accompagner et accélérer ce mouvement et s'appuyer dessus pour l'inclusion numérique dont leurs administrés ont besoin.

Mais le self data est aussi, pour les collectivités, **une ouverture vers la simplification des démarches administratives, et vers une plus grande souveraineté sur les données personnelles en lien avec les politiques territoriales.** Le self data permet à chaque citoyen de choisir de faire circuler des données les concernant. Pour la collectivité, cela peut avoir plusieurs avantages :

<sup>(4)</sup> Le principe Dites-le-nous une fois (DLNUF) consiste à éviter aux usagers des services publics de fournir, lors de leurs démarches en ligne, des informations ou pièces justificatives déjà détenues par d'autres administrations, en s'appuyant sur le partage automatique de données via des API.

**simplifier la mise en œuvre du programme « Dites-le nous une fois »<sup>(4)</sup> et même étendre son champ d'application ;**

**solliciter un partage de données (anonymisées ou non) de la part des administrés pour participer à l'amélioration de leur cadre de vie en partageant leurs données (utilisées par la suite de manière anonymisée et agrégée) : par exemple pour repérer les quartiers où le commerce est en décroissance, les changements dans la mobilité, l'utilisation des équipements publics, et ce beaucoup plus rapidement et beaucoup plus précisément qu'en commandant des enquêtes ponctuelles.**



## **II. Une nouvelle logique de partage de données personnelles émerge**

# II.1.

## Les droits de l'individu au cœur du modèle du self data

Dans les modèles « classiques » de transfert de données, l'utilisateur est à la fois le fournisseur et le client de tout un modèle économique reposant sur la publicité ciblée :

- Il fournit des données (en renseignant des formulaires, en laissant des traces de navigation sur les sites),
- d'autres sont générées en observant son comportement (comme le temps passé sur un site, les liens sur lesquels l'utilisateur clique, ou même la liste des achats d'un porteur d'une carte de fidélité).

Ces données sont analysées, enrichies, croisées, anonymisées ou non. L'analyse de ces données est faite par des entreprises peu connues du grand public dont l'objectif est d'établir des profils d'individus toujours plus précis. Ces profils sont ensuite revendus aux annonceurs qui diffusent aux individus des publicités toujours plus ciblées. Il existe de nombreux angles morts dans les contrôles de ces pratiques lorsque les détenteurs ne sont pas sous le regard d'autorités adaptées (comme la CNIL en France).

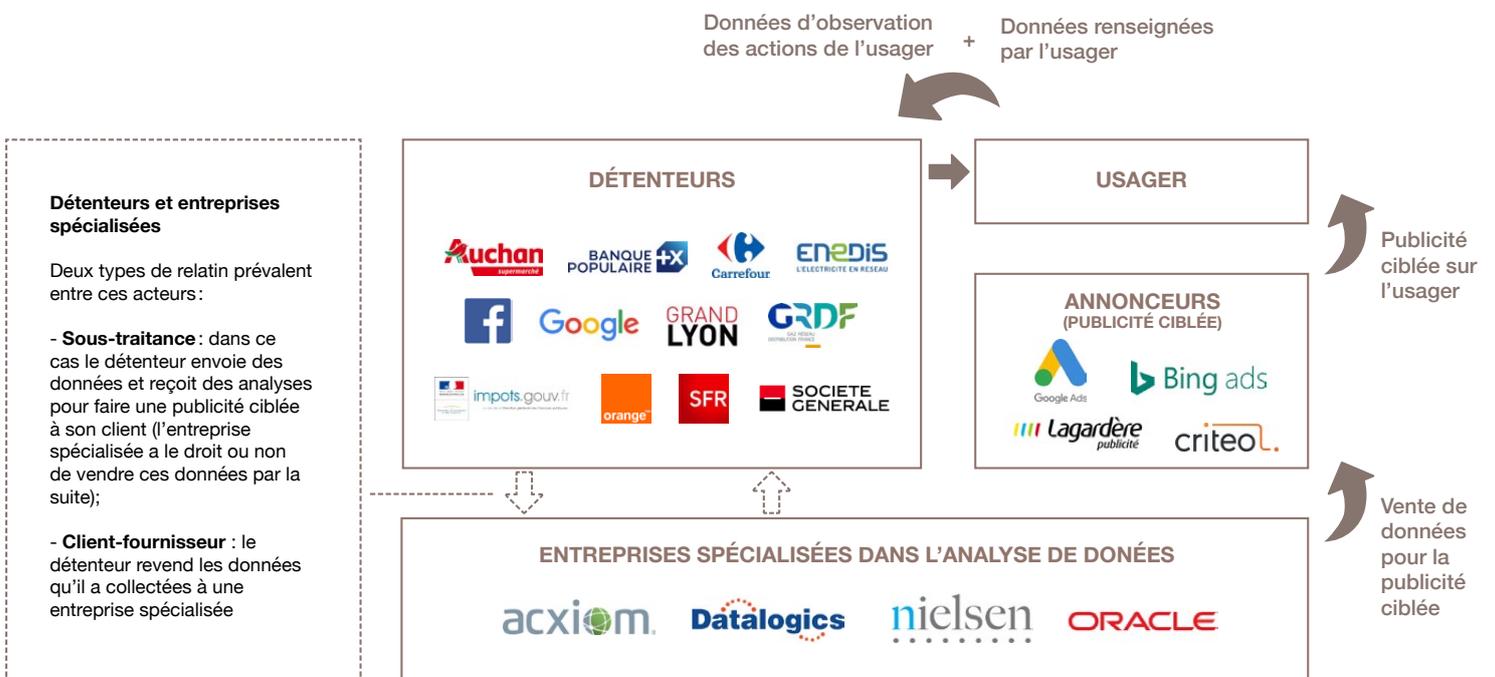


figure 1. Modèle de circulation des données classique avec 'third party data'

Le self data crée une alternative, complémentaire au modèle actuel, dans laquelle l'utilisateur est véritablement le point central de la circulation des données. Autrement dit, toute circulation des données personnelles est subordonnée à l'expression de la volonté de celui qu'elles concernent : fourniture des données sur la base du contrat ou du consentement, demande de portabilité, autorisation de transfert vers un tiers. Ceci permet à l'utilisateur de recevoir des services qu'il a choisis, sans diffuser de données superflues.

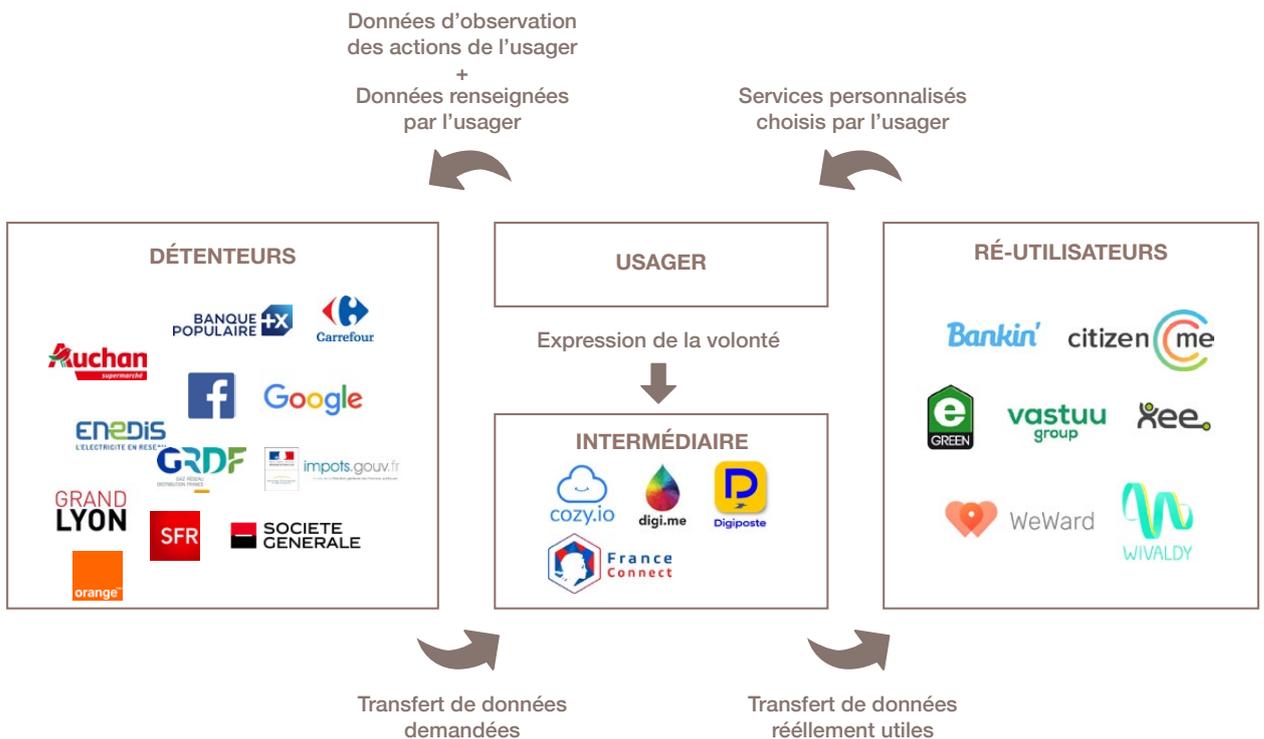


figure 2. **Modèle de circulation des données régi par le self data**

Les bénéfices peuvent être directs : la proposition d'un service ou d'un usage / ou une valorisation financière. Ils peuvent aussi être indirects : participation à l'amélioration du bien-être collectif, renforcement des pratiques contributives et collectives, etc.

Dans le modèle self data, les intermédiaires jouent un rôle capital. Ce sont eux qui permettent la transmission des données. Les intermédiaires appliquent et traduisent techniquement la volonté de l'utilisateur. Deux types d'intermédiaires se distinguent :

- **les gestionnaires de consentement** organisent les transferts mais ne stockent pas les données. Comme un notaire ou un huissier numérique, ils indiquent au détenteur quelles données il doit transférer à quel ré-utilisateur, et lui prouvent que l'utilisateur a consenti à cette transmission. C'est le cas de Fair & Smart et de OneCub par exemple.
  
- **les entrepôts de données personnelles ou clouds personnels** stockent les données que l'individu demande à récupérer. Ils peuvent aussi donner accès à ces données à d'autres services, toujours sur demande explicite de l'individu. Parmi les clouds personnels, certains proposent des services avec des fonctionnalités plus avancées en termes de protection de la vie privée. Par exemple, Cozy Cloud peut éviter la transmission de données à un nouveau service en proposant un espace de calcul : le service fournit son algorithme, qui est exécuté sur les données de l'utilisateur dans son cloud personnel, et l'utilisateur reçoit le résultat du calcul, sans qu'aucune donnée ne sorte de son cloud personnel. Concrètement, si une personne souhaite un conseil sur son contrat de fourniture d'énergie : au lieu d'envoyer toutes ses données de consommation à un comparateur d'offres, l'application (qui connaît les offres de tous les fournisseurs d'énergie) calcule directement dans le cloud personnel lequel serait le moins cher - de la même manière que l'on peut acheter un logiciel pour le faire fonctionner sur son ordinateur en coupant internet au lieu d'utiliser un service en ligne. **Il s'agit d'amener l'algorithme aux données et non l'inverse.**

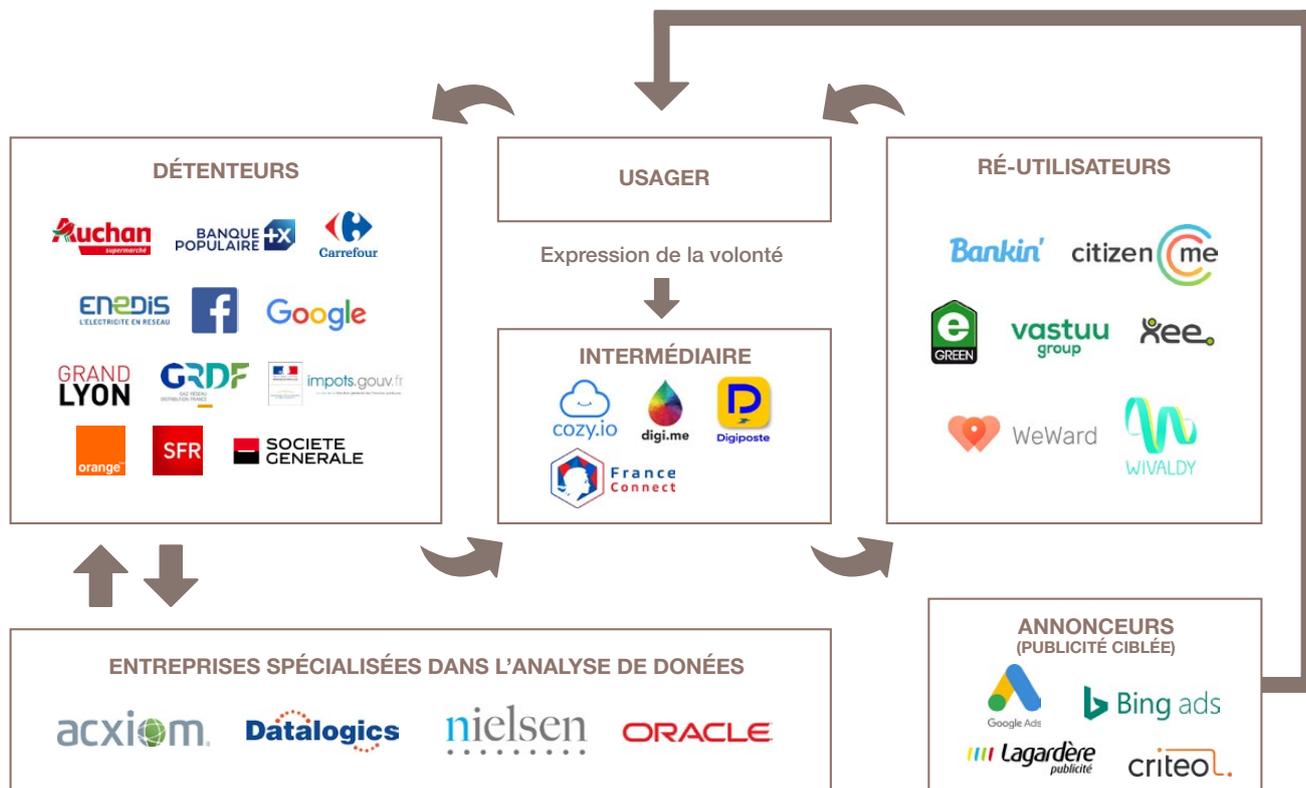


figure 3. Combinaison du modèle self data et du modèle classique

Le modèle de self data n'exclut pas le modèle classique de transfert de données personnelles. Les deux circuits peuvent sans problème cohabiter dans les faits, bien que cela pose une question de multiplicité du stockage de données et donc d'énergie consommée. En revanche, l'existence d'un modèle self data crée une alternative, la possibilité d'une décision qui pourrait mener au choix de l'arrêt des canaux classiques de transmission (ici en gris) via deux modes distincts :

- une somme de décisions individuelles: le libre choix des consommateurs pourrait les pousser à se tourner vers les entreprises qui s'engagent à ne pas valoriser leurs données personnelles. L'adoption de bloqueurs de publicité pourrait encore progresser et les marchés des annonceurs se tarirait par leur incapacité à produire de l'offre. Faute d'espace et de données, ces acteurs ne seraient plus dans une position économique viable et le canal ici en gris disparaîtrait.
- une décision collective, dont la forme la plus immédiate est la législation, pourrait acter la présence d'une voie plus respectueuse des libertés des citoyens, et donc interdire les procédés classiques de transfert de données, une décision marquerait un changement plus brutal que le premier cas.

## II.2.

# Les principes de la circulation des données décrits par les régulateurs

Le self data demande une révolution dans le mode de transfert des données personnelles. Ce changement est une source légitime de questionnement juridique. La France a une longue histoire de protection des données personnelles commencée en 1978 avec une exigence : « L'informatique doit être au service de chaque citoyen. (...) Elle ne doit porter atteinte (...) ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques. » <sup>(5)</sup>

Cette histoire a une forte influence sur le cadre actuel de protection des données personnelles, désormais européen : le Règlement général sur la protection des données (RGPD). Il s'applique depuis mai 2018 et consacre, en plus des droits désormais bien connus, un droit à la portabilité des données personnelles qui est un accélérateur majeur pour le self data.

Il s'applique depuis mai 2018 et consacre, en plus des droits désormais bien connus, un droit à la portabilité des données personnelles qui est un accélérateur majeur pour le self data.

L'article 20 du RGPD dispose que chacun est en droit de demander à récupérer les données personnelles qu'il a fournies, dans un format usuel structuré et lisible en machine, et d'en faire par la suite l'usage qu'il souhaite. Il est également possible que le citoyen ne demande pas à recevoir ses données directement, mais demande leur transfert vers une autre entreprise.

<sup>(5)</sup> Loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, article 1 : « L'informatique doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques. »

<sup>(6)</sup> Source : [CNIL](#)

### Données éligibles <sup>(6)</sup>

1.	2.	3.	4.	5.
Est-ce que les données concernent bien la personne qui fait la demande et sont des données personnelles ?	Est-ce que les données sont traitées par des moyens automatisés ?	Est-ce que le traitement de données se fonde sur le consentement ou du contrat ?	Est-ce que les données sont fournies par la personne concernée ?	Est-ce que ces données ne portent pas atteinte aux droits et libertés de tiers ?

Si la réponse est oui à toutes ces questions, les données sont portables.

Ce droit à la portabilité souffre tout de même plusieurs limites :

- \_\_\_\_\_ il est limité aux données fournies par l'utilisateur ou collectées pour l'exécution d'un contrat. Rien n'est donc précisément établi à propos des données d'observation du comportement de l'utilisateur ;
- \_\_\_\_\_ il est limité par les autres droits (droit à l'oubli des tiers par exemple) ;
- \_\_\_\_\_ il ne s'applique pas lorsque la collecte est faite à l'occasion d'une mission de service public ;
- \_\_\_\_\_ il est volontairement non prescriptif sur les modalités techniques.

C'est ce dernier point qui est le plus complexe pour encadrer la portabilité : afin de ne pas enfermer l'écosystème dans des normes d'échange complexes, le règlement ne précise pas les formats à utiliser pour le transfert. Ceci mène à ce que chaque entreprise puisse utiliser le format de son choix et n'incite pas à l'harmonisation.

## FOCUS

### La législation peut être plus précise sur les normes d'échange des données: l'exemple des données bancaires

*Directive sur les services de paiement 2 (DSP2)  
et Règlement délégué (UE) 2018/389*

La directive DSP2 et le règlement ont deux objectifs principaux pour le secteur bancaire :

- \_\_\_\_\_ permettre l'émergence de fournisseurs tiers de services bancaires (comme Bankin ou Linxo) grâce à des normes d'échange de données,
- \_\_\_\_\_ renforcer la sécurité des transferts d'informations bancaires.

Pour cela, ces textes établissent des normes très précises sur les protocoles d'échanges, et l'aboutissement principal est l'apparition, en 2018-2019 de services financiers qui ne sont pas gérés par des banques, et permettent plus de services: agréger les informations de ses comptes dans plusieurs banques dans une seule application, bénéficier d'un bilan et de conseils financiers, etc. Les banques se sont positionnées sur ce marché et ceci permet aussi de visualiser l'état de son compte Caisse d'Épargne grâce à son application Société générale par exemple.

Ces textes sont beaucoup plus précis que le RGPD et ont permis une transmission effective des données. Mais il faut garder à l'esprit qu'ils concernent un secteur bien précis, avec assez peu d'acteurs. Les retombées n'ont pas non plus été à la hauteur du projet envisagé, avec assez peu de nouveaux services développés en France, essentiellement car les grandes banques s'en sont tenues à l'application minimaliste de la loi. À l'inverse, au Royaume-Uni, certaines banques, comme Starling font du *DSP2 by design* en intégrant directement la portabilité de leurs données dans la conception.

## II.3.

# Les détenteurs de données plutôt réticents à la portabilité

Certains détenteurs, comme les banques et les opérateurs d'énergie, ont ouvert des canaux de transmission de données pour se mettre en conformité avec la réglementation. D'autres comme les GAFAM ou Le bon coin, ont créé des possibilités pour télécharger ses données personnelles (en fichier .csv).

En dehors de ces quelques exemples, les entreprises qui détiennent des données personnelles sont plutôt réticentes à l'ouverture de ces canaux de données pour plusieurs raisons :

- **la peur de perdre le contact privilégié avec ses clients :**  
les entreprises ont peur de voir s'immiscer un conseiller indépendant, ou des applications qui les couperaient du contact direct avec leur client ;
- **la peur de la concurrence d'être « disrupté » :**  
les entreprises historiques font face à la concurrence de nouveaux acteurs numériques, mieux équipés pour exploiter les données ; à ce jour, les données sont souvent vues comme un levier stratégique, et le meilleur atout de l'acteur historique sur ce marché ;
- **les difficultés soulevées en interne sur la gestion des données :**  
pour faire circuler des données efficacement, il faut avoir des données bien organisées, lisibles, fiables. La plupart des organisations ne sont pas matures sur ce sujet data, qui comporte un coût de transformation interne élevé et constitue un prérequis au self data ;

- **les risques juridiques et de sécurité informatique** sont aussi présents, et encore mal connus, ainsi la jurisprudence est presque inexistante sur la responsabilité du détenteur original en cas de mauvais usage des données après transfert demandé ;
- **la difficulté à converger vers des standards d'échange de données** est importante, des normalisations sectorielles pourraient être lancées, mais de nombreuses entreprises sont présentes dans deux, et souvent plus de secteurs, il serait difficile pour elles d'être conformes à plusieurs normes si elles parvenaient à être définies.

Ni les sanctions financières liées au non-respect du RGPD, ni la pression des consommateurs pour exercer le droit à la portabilité, qui permettraient d'inciter les entreprises à ouvrir les canaux de transmission de données ne sont effectifs. De ce fait, aucune entreprise ne s'est lancée sur la promesse d'une disponibilité effective des données personnelles comme un argument commercial pour se distinguer de ses concurrents, et les clients ne les sollicitent pas dans ce sens.

D'autre part, si la contrainte légale existe avec l'article 20 du RGPD, les autorités chargées de la faire respecter (i.e. CNIL) ne sont que très peu sollicitées sur ce cas. Toutefois, la CNIL anime régulièrement des ateliers « portabilité » à destination des porteurs de projet. Elle prévoit de publier des recommandations relatives au mandat pour l'exercice des droits fin 2020 et organise à cet occasion un événement pour la promotion du droit à la portabilité.

La situation est donc complexe pour l'ouverture des canaux de données. Les freins sont nombreux et les incitations sont rares pour les entreprises. Aujourd'hui, seules les entreprises qui sont contraintes par une réglementation sectorielle (secteur bancaire, énergie) et quelques entreprises de bonne volonté ont ouvert des canaux. Mais la réglementation est récente, il pourrait donc y avoir d'autres signes positifs dans un futur proche.



**III. Des modalités  
techniques définies  
et déjà mises  
en œuvre dans  
quelques secteurs**

## III.1.

# Plusieurs techniques sont mises en œuvre pour la transmission des données

La logique de self data repose sur trois types d'acteurs, tous liés à l'individu par une relation qui traduit un consentement libre et éclairé :

- **le détenteur est l'acteur habituel**, il collecte des données personnelles sur la base légale du contrat ou du consentement, de manière tout à fait classique ;
- **l'intermédiaire est soit un acteur** qui stocke les données dans un entrepôt de données personnelles, **soit un gestionnaire de consentement** qui indique aux détenteurs de donnée à qui envoyer quelles données ;
- **les ré-utilisateurs**, eux, sont multiples, ils peuvent être à la fois des acteurs très innovants qui offrent un nouveau service (comme Wivaldy), de nouveaux acteurs d'un service existant (comme Bankin'), et ils peuvent aussi être des acteurs historiques d'un secteur qui rendent leur démarches plus efficaces grâce au self data. Ils ont pourtant tous en commun le cœur de leur valeur : ils croisent des données qui concernent la même personne mais proviennent d'entreprises différentes afin de produire un service.

Les canaux de transmission des données sont l'enjeu majeur de l'écosystème. Pour extraire les données des bases des détenteurs, il existe plusieurs méthodes répondant aux exigences du RGPD, mais sont plus ou moins structurées, robustes, efficaces :

- Dans le cas où les données sont affichées sur internet (par exemple, les détails des comptes bancaires), il est possible de scraper ces données : si le *scraping* est une méthode peu appréciée des développeurs et peu robuste, elle est pourtant la plus rapide à mettre en place à petite échelle. Cette méthode consiste à utiliser un algorithme muni de l'identifiant et du mot de passe de l'utilisateur.

Ce robot entre sur le compte de l'utilisateur et lit toutes les informations afin de lui diffuser les données sous un format structuré. Cette méthode peut être employée à l'insu du détenteur des données, ce qui pose alors un problème moral. De plus, elle comporte des limites techniques: le robot a été programmé pour lire certaines informations à certains endroits de pages web précises. Si le détenteur change l'affichage des données (par exemple qu'il change l'endroit où s'affiche le nom ou l'adresse), le robot devient inopérant (car il a été conçu pour aller chercher les informations à un endroit donné), et il faut en créer un nouveau.

—— La communication de fichiers informatiques (tableaux, bases de données) contenant les données est mise en place par certains acteurs (comme Le Bon Coin). Ce sont des initiatives intéressantes, mais pas toujours techniquement robuste (comme pour le *scraping*, si le détenteur change l'ordonnement des données, cela peut perturber les programmes qui extraient ces données). Elles sont en revanche beaucoup plus morales, dans la mesure où le détenteur de données est informé et transmet les données de lui-même. Cette méthode de transmission des données a pour avantage d'être moins coûteuse et plus rapide à mettre en place que la création d'une API. Si la fréquence de transmission de ces données est très faible, cette méthode peut être suffisante. Toutefois, elle ne permet pas un passage à l'échelle industriel.

—— Les API sont la solution technique la plus pratique pour exploiter les données. Ce sont des formes très simples de communication entre logiciels. L'intermédiaire peut ainsi dialoguer efficacement avec le détenteur et obtenir les données de manière simple, fiable, et rapide (si l'API est bien faite). Certaines API intègrent des fonctions de sécurisation pour s'assurer que l'intermédiaire est bien autorisé à y accéder. Cependant, le déploiement d'API nécessite que les données soient déjà structurées et de qualité, ce qui est encore un défi à relever pour nombre de détenteurs de données. C'est donc une meilleure solution in fine, mais qui suppose un travail préalable relativement coûteux de mise à niveau des systèmes d'information (voir Fiche technique n°1 ) .



Voir fiche  
technique n°1

En fonction des types de détenteurs, les données sont plus ou moins exploitables. Les données qui ont généralement le plus de valeur se caractérisent par :

- **Une fréquence d'utilisation élevée**, avec un utilisateur qui utilise régulièrement le service, chaque jour voir plusieurs fois par jour
- **Un segment de marché avec assez peu d'acteurs**, car la concentration facilite l'accès aux données : le nombre d'acteurs à contacter et avec lesquels s'accorder est moins important
- **Une accessibilité et une interopérabilité effective**, avec des standards de partage existants ou facilement constructibles car les bases de données des acteurs se ressemblent, mais aussi avec des canaux techniques adaptés et efficaces pour la transmission des données
- **Des données les plus fiables possibles**, les données d'origine douteuse ou difficilement vérifiable ne sont pas réellement utiles, mais celles qui sont certifiées peuvent être utilisées pour fournir des preuves

En France, les principaux détenteurs de données peuvent être rassemblés dans ce tableau en considérant les caractéristiques ci-dessus.

Acteurs	Fréquence d'utilisation	Couverture du marché par un nb restreint d'acteurs	Accessibilité / Interopérabilité	Fiabilité des données
	● ● ● ● ●	● ● ● ● ●	● ● ● ● ●	● ○ ○ ○ ○
	● ● ● ● ●	● ● ● ● ●	● ○ ○ ○ ○	● ● ● ● ●
	● ○ ○ ○ ○	● ● ● ● ●	● ● ● ● ●	● ● ● ● ●
	● ● ● ● ●	● ● ● ● ○	● ○ ○ ○ ○	● ● ○ ○ ○
	● ● ● ● ○	● ○ ○ ○ ○	● ○ ○ ○ ○	● ● ● ● ●
	● ● ● ● ●	● ● ● ● ●	● ● ● ● ●	● ● ● ● ●
<b>Banques</b>	● ● ● ● ●	● ● ● ● ●	● ● ● ● ●	● ● ● ● ●

## III.2.

# Différentes techniques pour fournir des services numériques basés sur les données de manière plus efficace et moins intrusive

Aujourd'hui les services numériques basés sur l'utilisation de données personnelles sont majoritairement fournis en trois étapes :

- \_\_\_\_\_ le fournisseur du service achète ou obtient les données qui lui sont nécessaires en les demandant aux utilisateurs (par exemple en lui demandant de compléter un formulaire), ou grâce à des partenariats commerciaux avec d'autres acteurs ;
- \_\_\_\_\_ il utilise ses algorithmes, sur ses serveurs, pour produire le service ;
- \_\_\_\_\_ le service est « livré » à l'utilisateur (sous forme d'une information issue d'un calcul).

Mais il est également possible, grâce au self data, de procéder d'une manière beaucoup plus protectrice de la vie privée et qui favorise l'indépendance de l'utilisateur :

- \_\_\_\_\_ le ré-utilisateur fournit à l'utilisateur l'algorithme pour traiter ses données ;
- \_\_\_\_\_ l'utilisateur exécute cet algorithme dans son espace, sur son propre serveur ;
- \_\_\_\_\_ l'utilisateur obtient le service (le résultat du calcul fait à partir de ses données).



Voir fiche  
technique n°3

Il existe plusieurs techniques pour réaliser ces calculs (voir Fiche technique n°3 ). Très protectrices des données personnelles (seul l'utilisateur y a accès), et elles ont l'avantage de limiter les échanges de données en les traitant directement à la source. Ce principe visant à amener les algorithmes vers les données plutôt que de déplacer les données vers les algorithmes est une tendance de plus en plus courante en informatique. Ce procédé peut être utilisé pour un grand nombre de services, et il est la base d'une technique de plus en plus connue : la preuve à connaissance nulle, ou *Zero Knowledge Proof*.

Cette technique peut être illustrée par un exemple simple : les fiches de paie ou les avis d'imposition, documents nécessaires à la location d'un logement, contiennent énormément d'informations personnelles que le propriétaire n'a pas besoin de connaître alors qu'il cherche uniquement la réponse à la question «gagne-t-il plus de X € par mois ?».

Avec un cloud personnel et un algorithme adapté, il est extrêmement simple que le propriétaire puisse obtenir une réponse simple : «oui» ou «non». Il pourrait même poser des questions plus complexes et plus pertinentes comme connaître le revenu moyen sans jamais avoir accès au détail des revenus du foyer. Nombre de clouds personnels envisagent la fourniture de service, toujours avec le consentement de l'individu, et cela serait particulièrement adapté à ceux qui ont dans leur cœur de produit la gestion des fiches de paie, comme Digiposte qui explore ce nouveau type de services.

## III.3.

# Deux axes pour analyser les services self data : richesse des données et degré de protection des données.

Les services self data peuvent être très variés selon les cas d'usage qu'ils proposent mais également d'après la richesse des données qu'ils utilisent et le degré de protection de la vie privée qu'ils permettent (voir le schéma ci-dessous).

La richesse des données repose sur deux éléments :

- **la diversité des données**: certains services peuvent ne se baser que sur un type de données (uniquement la consommation d'énergie du foyer par exemple), ou bien plusieurs (produits achetés auprès d'enseignes sportives et activité sportive par exemple)
- **la fréquence d'actualisation des données** est également importante avec des données très statiques, comme l'identité ou la date de naissance, et des données plus « fraîches » comme l'usage de la carte bancaire.

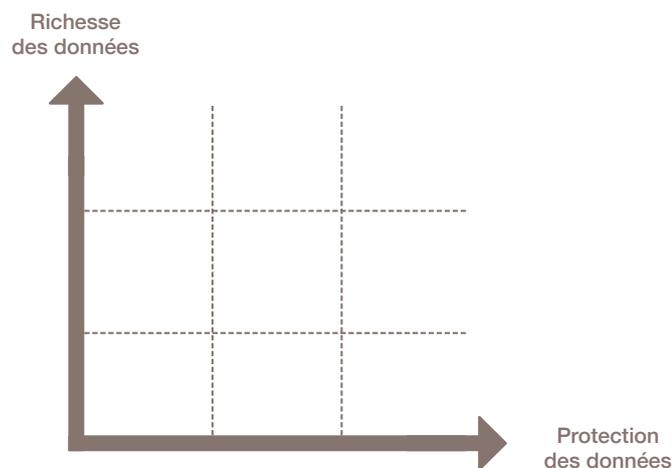


figure 5. Grille d'analyse des services vis-à-vis des principes du self data

Le niveau de protection des données personnelles peut être résumé par trois possibilités :

- **l'entreprise qui utilise les données en fait un usage peu clair**, qui, même s'il est légal, reste mal lisible par les utilisateurs et diffuse les données personnelles à différents acteurs sans consentement véritablement explicite ;
- **l'entreprise qui utilise les données explique clairement l'utilisation** qu'elle fait des données personnelles et recueille le consentement de manière explicite ;
- **l'entreprise qui propose le service n'a pas techniquement la possibilité de faire un usage discutable** des données personnelles car elle n'y a pas accès, ou car l'accès est restreint et suffisamment contrôlé. Cette catégorie de services utilise **les techniques décrites dans la partie III.2.**



Voir fiches  
techniques  
décrites dans  
la partie III.2



**IV. L'écosystème  
d'acteurs  
se développe  
lentement**

## IV.1.

# Quelques services sur le marché basés sur des données accessibles à la portabilité

La portabilité des données bancaires et des données de consommation d'énergie a prouvé que la portabilité des données personnelles permet l'émergence de nouveaux services et que ceux-ci sont adoptés par des utilisateurs :

— **La portabilité des données bancaires intégrée à la deuxième directive sur les services de paiement (DSP2) a encadré l'émergence d'agrégateurs et de comparateurs bancaires :**

La DSP2 (voir 1.2) est venue imposer une portabilité des données des comptes courants à compter du 14 septembre 2019. Ceci a permis le développement d'agrégateurs bancaires.

À ce jour, il existe une dizaine de services d'agrégateurs bancaires opérant en France : Linxo, Banking, Iswigo, Fiducéo (Boursorama), Cozy Banque, Budgea, Yolt (ING Direct), Max (Crédit Mutuel Arkéa). Ces services compteraient 4 à 4,5 millions d'utilisateurs en France. Ces agrégateurs ont pu être lancés ou rachetés par des banques elles-mêmes ou créés par de nouveaux acteurs. Les premiers améliorent le service fourni à leurs clients, et se construisent une nouvelle base de prospects. Les seconds proposent des modèles freemium tout en se rémunérant par de la publicité ciblée (à l'exception de Cozy Banque).

**En ce sens, ces services s'appuient sur la portabilité sans promouvoir les principes du self data sur le contrôle des données par l'utilisateur (notamment grâce à un consentement éclairé).** Toutefois, ils prouvent que des opportunités existent à la création de services basés sur la portabilité des données personnelles. Bankin' est une application bancaire qui permet de visualiser la situation et l'évolution de son ou ses comptes bancaires, mêmes s'ils sont ouverts dans différentes banques, dans une seule application. Bankin' permet aussi de bénéficier de conseils indépendants de sa banque. Ces services permettent à la fois de mieux comprendre son comportement et de bénéficier d'un assistant.

Bankin'

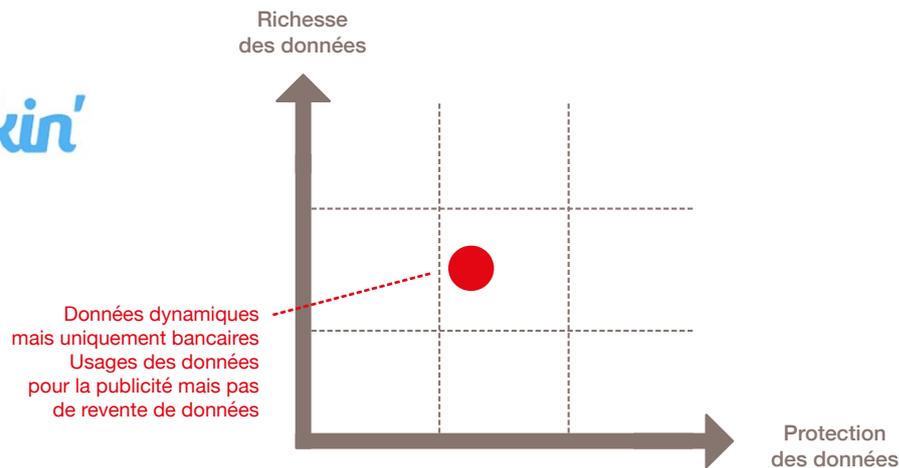


figure 6. Analyse de Banking vis-à-vis des principes du self data

Bankin' repose sur un double modèle économique: d'une part, l'application est un freemium, c'est-à-dire qu'elle est gratuite pour ses fonctionnalités de base, mais que des options plus avancées sont payantes. D'autre part, Bankin' est un apporteur d'affaire pour les services bancaires: l'application étudie le comportement des utilisateurs et leur suggère des services pour renégocier un crédit, ou une assurance par exemple, si l'utilisateur concrétise cette suggestion, Bankin' reçoit une commission. À la différence d'autres acteurs, Bankin' ne revend pas les données mais fait elle-même les suggestions personnalisées avec ses algorithmes. En 2019, Bankin' comptait 3,2 millions (en France, Angleterre, Espagne et Allemagne). Créée en 2011 elle a levé depuis 31,8 millions de dollars (US).

### La portabilité des données de consommation énergétique a permis la création de services de suivi de consommation énergétique.

Le cadre légal (Directives Electricité 2019/944) a accéléré la mise en place d'API par les acteurs de l'énergie. La fiabilité, l'exploitabilité et la disponibilité des données du secteur de l'énergie les ont rendus propices au développement de services. Le déploiement des compteurs Linky et Gazpar, initié en 2015 permet de générer des données de consommation d'électricité précis pour les foyers (35 millions de foyers équipés de compteurs Linky d'ici 2021, 11 millions de foyers équipés de compteurs Gazpar d'ici 2022).

Enedis s'est «APIsée» et début 2020, près de 40 entreprises se connectent à travers l'API. GRDF sera «APIsée» très prochainement.

Ces services de coaching énergétique pourraient être porteurs pour des enjeux sociétaux: 14% des ménages français souffrent de précarité énergétique, soit 3,8 millions de logement occupés par des ménages en précarité énergétique. Les Défis Familles à Energies Positives ont permis de montrer que l'adoption d'éco-gestes permet de réduire la facture énergétique d'un foyer de 200€ en moyenne sur la période de chauffe.

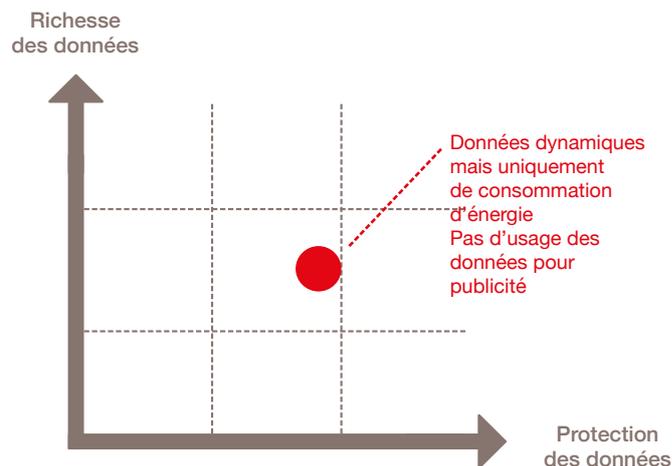


figure 7. Analyse de Wivaldy vis-à-vis des principes du self data

Wivaldy **recueille et analyse les données des compteurs Linky** (ou capteurs wifi du logement) et les **restitue à l'individu sous forme de bilan de consommation** électrique du logement.

Sur la base du bilan, Wivaldy propose une identification des surconsommations du logement pour **ajuster les comportements des habitants et leurs options tarifaires** et une analyse des **offres économiques plus adaptées** aux besoins et habitudes des individus. Si l'interface de suivi de la consommation et les bilans simples sont gratuits, les diagnostics plus approfondis ne sont accessibles qu'aux abonnés (49€/an).

Au printemps 2020, Wivaldy ne dispose pas encore d'une application mais d'une web-app (hybride entre une application et un site web, plus flexible). La startup n'a pas encore levé de fonds.

**La portabilité des données simples et stables est un atout pour faciliter le remplissage de formulaires.**

La question de la certification d'une identité numérique est clé pour divers acteurs. Si certains ont besoin d'obtenir des informations vérifiées, d'autres entreprises souhaitent simplement faciliter à leurs clients le remplissage des formalités d'inscription.

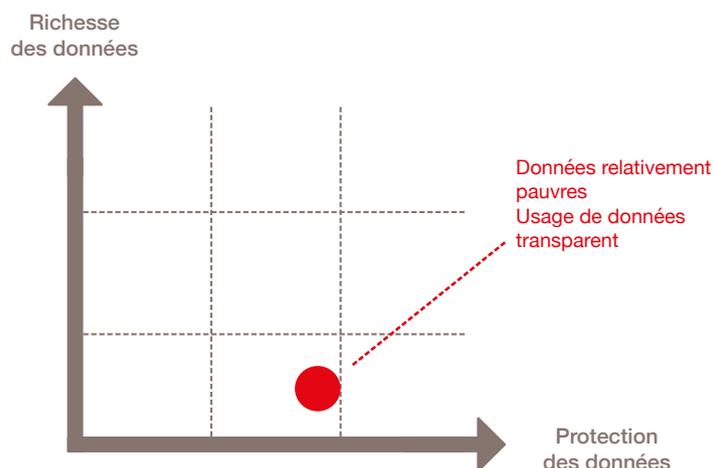


figure 8. Analyse de Mobile ID vis-à-vis des principes du self data

Mobile ID est une solution mise en place par les opérateurs de la téléphonie mobile en France. Ce service permet aux usagers de services en ligne, lorsqu'une entreprise partenaire de Mobile ID leur demande de remplir des informations personnelles, de donner uniquement leur numéro de téléphone et après une vérification, permet de remplir automatiquement le formulaire.

Le service réalise une portabilité et permet à l'utilisateur une vision très claire de l'utilisation de ses données lors de la demande de consentement.

Le service trouve son modèle économique dans une tarification auprès des entreprises partenaires qui y voient un moyen de concrétiser plus de ventes en évitant que les consommateurs soient découragés par le volume d'informations à saisir.

**La mobilité durable comprend aussi de nombreux défis pour les collectivités et les individus.**

La transition vers une mobilité durable intègre des objectifs complémentaires : favoriser l'intermodalité, les modes de transport collectifs, doux, électriques. Pour cela, les collectivités disposent de plusieurs leviers : adapter les réseaux de transport en commun (tarification, implantations des stations, fréquence/horaires de passage), créer des parking-relais, mettre en place des réseaux de vélos partagés, aménager la voirie, développer les stations de recharges électriques.

**Dans ce cas, l'application des principes de self data est vertueuse en permettant une migration plus simple vers un système de compte mobilité unique où tous les services de mobilité d'un territoire sont accessibles avec une facturation unique.** La Fabrique des Mobilités de l'ADEME a commencé à construire les briques d'un Compte personnel de Mobilité, basé sur la portabilité des données et un cloud personnel. Une documentation sur les demandes de portabilité auprès de la SNCF, Blablacar, Uber, etc. et l'exploitation de ces données est disponible sur leur blog.

Des services self data embryonnaires sont développés dans un domaine très proche. Le fournisseur de cloud personnel Cozy cloud a développé une application permettant de calculer le budget de mobilité d'une personne sur la base de ses données bancaires. Ce service a été développé sur la base du protocole DISPERS (voir Fiche technique n°3 ).



Voir fiche  
technique n°3

**En lien avec la Fabrique des Mobilités, l'Agglomération de la Rochelle va expérimenter avec plusieurs centaines de bêta-testeurs le développement d'un service self data calculant le budget mobilité et l'empreinte carbone, basé sur une architecture de cloud personnel à partir de 2021.** La collectivité envisage également de proposer aux bêta-testeurs de partager leurs données de mobilité avec la collectivité dans une approche contributive, les données permettant d'alimenter la décision publique grâce à une meilleure compréhension des usages.

## IV.2.

# Les Personal Information Management Systems, un marché d'infrastructure encore embryonnaire

Le self data repose sur des intermédiaires qui sont en quelque sorte des exécutants de la volonté de l'utilisateur : ils mettent en pratique la volonté du citoyen de récupérer certaines données et de les partager avec différents services.



L'expérience utilisateur est primordiale. L'existence d'un PIMS entre deux tiers est indispensable pour que les échanges soient « human centric » et offrent un management centralisé simple et efficace des consentements / flux de données.

Sans infrastructure type PIMS entre les tiers, le passage à l'échelle est empêché car la gestion des flux et consentements s'avère trop fastidieuse pour le citoyen lambda.

-

**Xavier Lefèvre,**  
*CEO de Fair & Smart*

Plusieurs entreprises proposent des services d'entrepôt personnel de données : Cozy Cloud, Digiposte, Digi.me, SOLID (Inrupt). D'autres proposent des services de gestion des consentements comme Fair & Smart. Afin de faciliter l'analyse, les principales caractéristiques sont résumées ci-dessous pour les deux types d'intermédiaires qui sont le plus développés dans l'étude.

## Comparaison des principales caractéristiques des gestionnaires de consentement et entrepôts de données personnelles

	Gestionnaire de consentements	Entrepôt personnel de données
Connexions techniques entre détenteurs et utilisateurs	X	X
Gestion des consentements	X	X
Tri / Minimisation des données partagées ou mises à disposition	X	X
Intéropérabilité des données	X	X
Stockage des données	-	X
Partage des données vers le ré-utilisateur	X	Optionnel
Contractualisation des services	Optionnel	Optionnel
Garant de l'identité numérique	-	Optionnel
Administration des flux de paiements	Optionnel	Optionnel

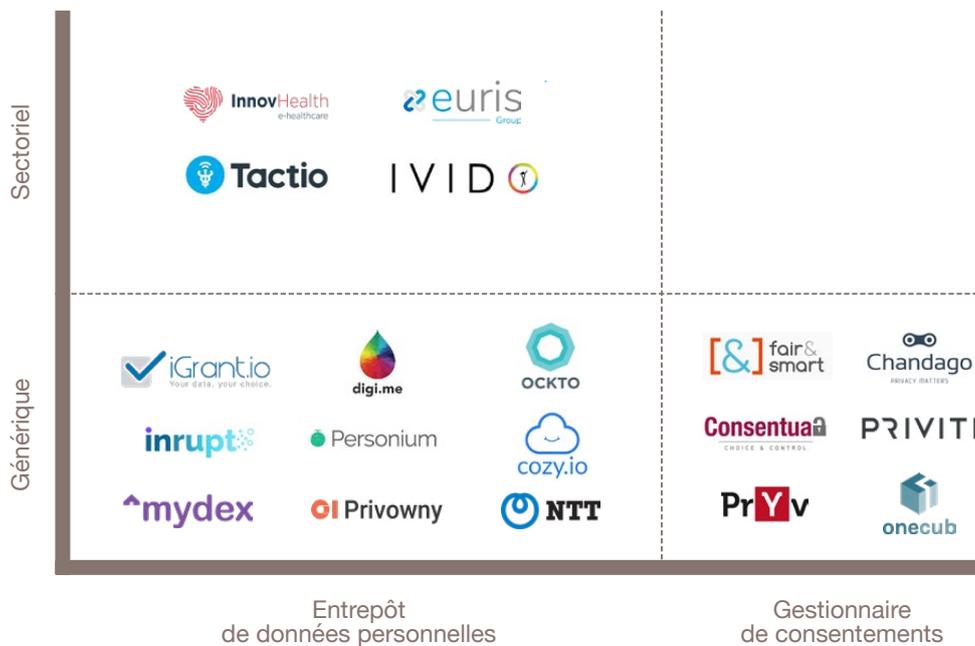


figure 9. Quelques exemples d'acteurs intermédiaires (fournisseurs d'entrepôt de données personnelles et gestionnaires) et leur positionnement

Les intermédiaires sont avant tout des plateformes et leur adoption est soumise à l'émergence d'un usage. Une plateforme ne peut passer à l'échelle qu'avec l'adoption massive du marché : par exemple, l'App store d'Apple et Android attirent les développeurs d'applications car ils donnent tous les deux accès à des millions d'utilisateurs potentiels. La diffusion des clouds personnels dépend de la même logique : équiper un grand nombre d'appareils, pour toucher une communauté d'utilisateurs suffisante pour attirer des développeurs d'applications.

Il s'agit d'un marché B2B2C, c'est-à-dire que les clouds personnels doivent être vendus aux entreprises afin qu'elles l'utilisent avec leurs clients. En effet, très peu de personnes acceptent aujourd'hui de payer un abonnement pour bénéficier de services leur permettant d'héberger leurs données, car elles disposent de clouds gratuits. Le modèle B2B2C semble plus adapté à ce marché, à condition d'identifier des services que des organisations sont prêtes à financer pour l'utilisateur final.

Par exemple, Digiposte gère et authentifie les fiches de paie, c'est l'entreprise qui rémunère Digiposte pour le service, mais c'est *in fine* l'employé qui en est le bénéficiaire (il peut continuer d'utiliser l'espace même après son départ de l'entreprise).

Dans une approche B2B, plusieurs sources de revenus potentielles sont identifiées à ce stade :

- La sous-traitance d'un point de vue « conformité » de la gestion des consentements auprès des détenteurs de données (Fair&Smart)
- La certification de l'identité de la personne auprès des services ré-utilisateurs (Mobile ID)
- La « certification » de la qualité des données auprès des services ré-utilisateurs (si l'intermédiaire peut apporter la preuve qu'elle provient directement du détenteur tiers de confiance par exemple un diplôme) (Digiposte)
- Une commission calculée sur la valeur que le ré-utilisateur génère grâce aux données ou bien calculée en fonction du coût de la portabilité (Hub of All Things)

## FOCUS

En 2020, Markets and Markets a estimé la taille du marché mondial des stockages sécurisés de données en ligne à 32,3 milliards d'euros. The Insight Partner estimait en 2019 la taille du marché mondial des gestionnaires de consentement pour les entreprises à 199 millions d'euros.

The Shift Project, a évalué le marché des intermédiaires à 14,6 milliards d'euros au Royaume-Uni. Si l'on transpose cette méthode au marché français, l'évaluation se porte à 10,3 milliards d'euros.

Selon autre méthode privilégiée OneCub et Visions, deux entreprises françaises du self data ont évalué **le marché des intermédiaires à 1,8 milliards d'euros et celui du self data à 61,8 milliards d'euros en France**. Ces estimations sont portées respectivement à 14,7 et 489,4 milliards d'euros dans l'Union européenne. Ces estimations sont faites considérant que 85% de la population adopte des services self data. Si le service reste marginal (3% de la population), le marché est évalué à 66 millions d'euros. (voir Fiche technique n°2 )



Voir fiche  
technique n°2

## IV.3.

# Des écosystèmes d'acteurs engagés pour lever les freins au développement du self data

Il existe quatre enjeux majeurs pour le développement des écosystèmes self data :

- \_\_\_\_\_ Le développement d'un cas d'usage attractif pour les usagers, qui permet d'attirer des clients vers cette logique de réappropriation des données personnelles
- \_\_\_\_\_ La définition de modèles économiques viables et adaptés à chacun des quatre types d'acteurs, avec une répartition de la valeur créée par le nouveau service
- \_\_\_\_\_ La mise en place de capacités de transfert et d'exploitation des données entre les parties prenantes (standards techniques, compétences numériques, etc.) dans un cadre sécurisé.
- \_\_\_\_\_ La finalisation un cadre de protection juridique et réglementaire adapté pour sécuriser les problématiques liées aux responsabilités dans la continuité du RGPD et en prenant appui sur les expériences sectorielles déjà en place.

Les initiatives prises par les différents acteurs impliqués dans la mise en œuvre du self data concourent à résoudre ces enjeux.

La Commission européenne a publié le 19 février 2020 sa stratégie pour les données. Cette stratégie est portée afin de résoudre de nombreux problèmes :

- \_\_\_\_\_ favoriser une meilleure disponibilité des données et leur partage ;
- \_\_\_\_\_ réduire les déséquilibres existants en termes de pouvoir de marché (entre les PME et les acteurs qui détiennent de grands volumes de données) en favorisant leur circulation ;
- \_\_\_\_\_ encourager l'application de formats et protocoles compatibles / interopérables ;
- \_\_\_\_\_ réduire ses dépendances technologiques dans les infrastructures stratégiques qui sont au cœur de l'économie des données ;
- \_\_\_\_\_ permettre aux individus de mieux exercer leurs droits.

La stratégie met en avant le besoin d'un meilleur équilibre des pouvoirs et de coordination entre tous les acteurs de l'écosystème. Elle aborde également le besoin de création d'infrastructures intermédiaires (par ses références à la fiducie de données personnelles, un dispositif juridique qui permet de transférer à un tiers la propriété de tout ou partie de son patrimoine, pour l'administrer au profit d'un ou de plusieurs bénéficiaires, et aux « Personal Information Management Systems ») afin de renforcer le droit à la portabilité.

La Commission ne formule pas de cadre légal contraignant à l'heure actuel mais aborde la problématique sous l'angle de terrains d'expérimentation. Ces « bac à sable législatifs » sont des espaces définis dans le temps et sur un territoire donné au sein desquels le cadre légal est volontairement flexible pour permettre des expérimentations. Ceci permet d'observer les usages et les possibilités afin de créer par la suite une régulation plus adaptée.

Afin d'atteindre les objectifs cités ci-dessus, des solutions sont directement citées :

- des outils de gestion des consentements,
- des applications de gestion des informations à caractère personnel, y compris des solutions entièrement décentralisées qui s'appuient sur la Blockchain,
- des coopératives ou les fiducies de données à caractère personnel faisant fonction de nouveaux intermédiaires neutres dans l'économie, fondées sur les données à caractère personnel. Cette initiative est soutenue par le programme Europe numérique.

Apple, Google, Microsoft et Twitter ont également entrepris une initiative commune pour favoriser l'interopérabilité. Depuis 2018, ils se sont montrés proactifs pour travailler sur le **Data Transfer Project (DTP)** <sup>(7)</sup> qui promet aux utilisateurs une meilleure circulation de leurs données personnelles d'une plateforme ou d'une application à l'autre. Ce projet de collaboration, principalement « technique », n'a pour l'instant pas donné lieu à une réalisation concrète mais a permis de comprendre que ces groupes souhaitent apporter des solutions pour ce type de problématiques.

Dans la même direction, Microsoft a lancé le « Projet Bali » <sup>(8)</sup> pour créer sa propre offre d'infrastructure intermédiaire. Il s'agit d'une **banque de données personnelles** qui permettrait aux utilisateurs de contrôler toutes les données collectées (brutes ou déduites) à leur sujet, leur permettant de les stocker, visualiser, gérer, partager et monétiser. Son ambition à terme est de devenir un PIMS, créant ainsi un nouveau pôle d'attraction de la donnée dans ses serveurs.

D'autres acteurs de l'écosystème français identifient également le self data comme une opportunité : la Maif, Orange, Le Bon Coin, Enedis, le Groupe Vyv, ... Certains d'entre eux sont engagés dans la recherche de cas d'usages utiles pour leurs clients.

<sup>(7)</sup> Les géants du web travaillent à la portabilité des données, <https://www.usine-digitale.fr/article/les-geants-du-web-travaillent-a-la-portabilite-des-donnees.N726099>

<sup>(8)</sup> Bali, l'outil Microsoft qui veut redonner aux utilisateurs le contrôle de leurs données personnelles, <https://www.zdnet.fr/blogs/securite-it/bali-l-outil-microsoft-qui-veut-redonner-aux-utilisateurs-le-contrôle-de-leurs-donnees-personnelles-39878683.htm>

« Nous sommes convaincus que le self data sera à terme un avantage pour la MAIF. Les projets que nous menons actuellement s'inscrivent dans une réelle transformation où nous renforçons la position centrale du sociétaire sur l'usage de ses données

-

**Chloé Beaumont,**  
*responsable du Hub, de la veille  
et de la prospective à la MAIF*

Le self data constitue un domaine où l'action publique et les acteurs privés ont des rôles extrêmement imbriqués.

« Les perspectives de cas d'usage self data dans la santé sont nombreuses et prometteuses mais il ne faut pas sous-estimer le besoin d'embarquer et de pré-positionner les pouvoirs publics.

-

**Sylvie Kermabon,**  
*directrice de projet, Groupe Vyv*

Ceci se révèle notamment en ce qui concerne les normes d'échange des données. Si le RGPD ne prévoit pas les modalités techniques du droit à la portabilité, le besoin d'un standard d'échange n'en est pas moins prégnant entre les intermédiaires.

## Une initiative pour la création de standards européens – a New Governance

L'initiative a New Governance, issue du mouvement MyData, a pour objectif la création d'un Public Private Partnership pour la gouvernance de l'infrastructure et des standards pour la circulation des données personnelles.

La Commission Européenne dans sa Stratégie sur les données publiée le 19 février 2020 a souhaité réunir tous les experts européens et les acteurs industriels organisés en « data spaces » pour mettre en œuvre une infrastructure globale et pour faciliter la circulation des données (personnelles et non personnelles). a New Governance propose de rassembler les acteurs pour le volet données personnelles.

Cette standardisation et collaboration permet de rendre interopérable la couche des intermédiaires et ainsi de permettre aux organisations de facilement s'échanger et de mutualiser des données en passant par la permission des personnes concernées.

Chaque intermédiaire pourra permettre à un écosystème de s'échanger des données. Cet écosystème va avoir une gouvernance qui va décider des règles et des finalités de l'échange de données que l'intermédiaire va avoir à respecter et mettre en œuvre.

La standardisation et la modélisation de ces règles et finalités permet de plus facilement passer à l'échelle et d'accélérer l'adoption plutôt que d'avoir à s'adapter à chaque écosystème. aNG travaille notamment, dans un premier temps, sur cette standardisation dans le secteur des compétences et de la mobilité avec des écosystèmes en Europe et à l'international.

FOCUS

Le self data s'inscrit dans une dynamique d'innovation pour une gestion plus respectueuse des données. Le self data appartient au domaine de la deeptech, des innovations de rupture qui demandent de grands efforts de recherche et développement avant de donner lieu à des produits adoptés par beaucoup. Dès lors, les fonds d'investissement spécialisés se sont positionnés sur le sujet afin de favoriser l'émergence d'acteurs. En Finlande, Vake.fi et SITRA (comparables respectivement à la Banque publique d'investissements et à la Caisse des dépôts en France) ont créé un incubateur dédié aux technologies self data. Le MIT et l'institut pour la recherche en informatique du Qatar soutiennent le projet SOLID de Tim Berners Lee.

Tous ces éléments favorisent l'émergence d'un écosystème self data avec une logique d'ensemble grâce à une action conjointe d'acteurs publics et privés. C'est aussi le sens du MyData global qui permet aux acteurs de différents pays de coordonner leurs réflexions, leurs travaux et leur communication.



**V. Des ambitions  
territoriales  
pour le self data :  
entre coûts  
et bénéfices pour  
les collectivités**



Quelques collectivités françaises s'intéressent au concept de self data. Nantes Métropole a exploré des cas d'usage relatifs à l'alimentation, le projet RUDI de Rennes Métropole intègre une dimensions self data et d'autres collectivités ont commencé à chercher des opportunités, comme à Grenoble. Des projets pilotes et des expérimentations vont être lancées au second semestre 2020 par la Métropole de Lyon et l'Agglomération de La Rochelle. Concernant ces expérimentations, il s'agit autant d'éprouver la faisabilité technique, que les bénéfices générés par le self data et des cas d'usage utiles aux politiques publiques et aux habitants. L'équilibre économique pour la construction et le maintien des infrastructures, et la création d'un véritable écosystème de services ont fait l'objet d'une analyse réalisée par le cabinet Wavestone pour la Métropole de Lyon au premier semestre 2020. Les conclusions de leur étude sont restituées dans les pages suivantes.

# V.1.

## Des bénéfices certains mais difficilement quantifiables

Très peu d'infrastructures self data sont actuellement déployées en France ; dès lors, pour que les collectivités développent des services self data pour leurs administrés, elles doivent financer l'exploitation de l'architecture technique, ce qui représente un coût significatif au regard de leurs budgets habituels d'exploitation de services numériques. Toutefois, le self data semble être porteur de réels bénéfices pour les collectivités. **On compte à ce jour quatre perspectives de valorisation financière pour les collectivités :**

- 1.** En contribuant à l'efficacité de certaines politiques territoriales, le self data doit permettre aux collectivités d'optimiser leurs investissements. Par exemple, un service self data mobilité, comme à La Rochelle, pourrait inciter les individus à adopter des modes de transport alternatifs, ce qui permettrait aux collectivités de moins investir dans des actions de sensibilisation aux mobilités douces, voire de limiter certains investissements dans des équipements routiers destinés aux automobilistes.
- 2.** Le self data peut permettre aux collectivités d'optimiser le fonctionnement de certains services administratifs ; en effet, la Métropole de Lyon a lancé un service self data social qui a pour but de simplifier les démarches administratives et de rendre les interactions entre les individus et les travailleurs sociaux plus pertinentes en limitant le temps passé sur des actes administratifs simples pour effectuer des tâches à plus forte valeur pour les individus et de pouvoir ainsi assurer un suivi plus personnalisé.
- 3.** En encourageant des changements de comportement, les services self data peuvent générer un regain de pouvoir d'achat pour les ménages concernés. Par exemple, le service self data énergie, développé par la Métropole de Lyon devrait permettre de diminuer les factures énergétiques des ménages grâce à des tableaux d'analyse multi-fluides et à des défis de sensibilisation à l'efficacité énergétique. Dès lors, une partie des sommes économisées par les ménages pourrait être dépensée dans l'économie locale ce qui sera bénéfique à la collectivité.
- 4.** Enfin, si le self data prouve qu'il peut être créateur de valeur pour des acteurs privés, les collectivités pourront bénéficier de recettes fiscales liées à la création d'entreprises self data sur leur territoire.

Dans le cas de Lyon, le cabinet Wavestone a estimé que ces postes de recettes restent incertains tant qu'ils n'ont pas été testés; dès lors, la phase pilote lancée dans la métropole en juin 2020 permettra de préciser les bénéfices du self data pour les collectivités.

**Au-delà des bénéfices économiques promis par le self data, la Métropole de Lyon a identifié trois typologies de bénéfices extra-économiques qui peuvent concerner les collectivités décidant de se lancer dans un projet self data :**

- 1.** Le self data promet d'être un concept utile à l'intérêt général dans le sens où il peut contribuer aux politiques publiques (ex: aide à réduire la consommation énergétique, contribue à la promotion des mobilités douces, optimise le fonctionnement des services sociaux) et peut renforcer la confiance et le lien entre les collectivités et les administrés.
- 2.** Il a pour vocation de faciliter la création de services numériques apportant de la valeur aux usagers (ex : gain de pouvoir d'achat, bénéfices personnels) et peut être vu comme un accélérateur de la médiation numérique pour aider les individus à prendre conscience des enjeux autour de leurs données personnelles.
- 3.** Enfin, le self data incarne un mode de traitement des données personnelles plus sain, qui se démarque des pratiques actuelles des entreprises dans l'économie de la donnée.

## V.2.

# La collectivité peut s'impliquer à plusieurs niveaux



En complément de notre projet « self data », nous souhaitons aller plus loin que les prérogatives du RGPD en termes de privacy : construire avec les citoyens un modèle de gouvernance des données à implémenter pour chaque projet. C'est un véritable levier concret d'engagement des citoyens dans les politiques locales.

-

**Virginie Steiner,**  
*CDO – Agglomération de La Rochelle*

Les collectivités locales, surtout les communes et les EPCI, jouent un rôle majeur pour l'inclusion numérique de leurs administrés, et donc potentiellement dans l'activation de leurs droits numériques. Les collectivités peuvent apporter trois types de ressources pour promouvoir et développer le droit à la portabilité par l'intermédiaire du self data :

- 1.** D'abord, les collectivités sont elles aussi des détenteurs de données, et si l'article 20 du RGPD exclut les données relevant de l'intérêt public, elles peuvent détenir des données soumises au droit à la portabilité des données personnelles (dès lors qu'elles sont collectées sur la base du contrat ou du consentement). C'est le cas des données personnelles collectées dans le cadre de Délégations de Service Public par exemple.

2. D'autre part, les collectivités ont un atout majeur par leur taille et leur ancrage territorial. Elles sont assez accessibles pour être en contact avec de petites entreprises qui offrent des solutions techniques de self data, et assez importantes pour obtenir l'attention de grandes entreprises nationales. Elles ont donc un rôle à jouer dans la **coordination de cet écosystème**, pour amener les différents acteurs à échanger.

Les collectivités ont aussi accès à une ressource rare : la capacité à rassembler des testeurs pour mettre en place ces démarches, et à animer ce groupe de test. C'est ce que font Lyon ou La Rochelle. Les collectivités peuvent animer des groupes d'usagers des services publics, les former à ces nouvelles pratiques numériques, pour inciter les entreprises à développer des services. Cette connaissance se diffuse ensuite dans la collectivité et permettrait, à terme, des projets plus vastes.

3. Enfin, la collectivité peut occuper un rôle plus en retrait techniquement, mais tout aussi important avec une **posture de tiers de confiance**. Elle peut se porter garante, auditer les systèmes utilisés, ou même garantir un certain niveau de service, en somme, amener de la confiance dans un projet qui doit rassembler beaucoup d'acteurs avec différents niveaux d'expérience, et inclure pleinement le citoyen. Dans cette démarche, la collectivité peut aussi **prendre en charge tout ou partie du financement**. Les infrastructures de self data sont, comme toute structure numérique, très coûteuses pour les premiers utilisateurs, et beaucoup moins pour les suivants. Un EPCI peut prendre à sa charge l'apport initial pour fournir un service efficace à un prix raisonnable. Ceci peut également passer par le subventionnement de services pour les moins aisés, comme le conseil en consommation et rénovation énergétique.

La construction et la diffusion d'une infrastructure self data à l'échelle d'un territoire est un projet coûteux, qui ne peut se s'industrialiser qu'en s'appuyant sur une logique d'écosystème, de plateforme et de mutualisation. En ce sens, la Métropole de Lyon a conçu une approche progressive, dans laquelle son rôle sera de lancer et catalyser la mise en place des premières briques de ce nouveau modèle.

## V.3.

# Le passage à l'échelle : une démarche en plusieurs étapes

« Une solution serait de mettre en place des fédérations de villes pour capitaliser sur les travaux de chacun, notamment ceux de Lyon qui sont transposables à d'autres collectivités en France voire dans le Monde.

Les expérimentations de la Métropole de Lyon et les analyses du cabinet Wavestone s'accordent à dire que les collectivités peuvent appréhender un projet self data à travers trois phases clés ; le passage d'une phase à l'autre étant conditionné par des instances de validation permettant de limiter la prise de risques autour de ce concept encore émergent.

## Phase pilote

Avant de se lancer dans un projet de self data, les collectivités sont fortement encouragées à consulter les travaux déjà menés par d'autres collectivités ainsi que la documentation que celles-ci mettent à disposition en open source (ex : Métropole de Lyon, Agglomération de La Rochelle, Métropole de Nantes). Une fois familiarisées avec le concept, il est essentiel que les collectivités passent par une phase d'exploration pour identifier les cas d'usage à valeur et les données exploitables, choisir un modèle de gouvernance et les solutions techniques associées, tester l'appétence des individus et identifier des sources de revenus (ex : subventions). Cette phase peut durer entre six mois et un an en fonction des ressources mobilisées (chargé de projet, technicien à temps plein...).

Au terme de cette phase, les collectivités devront s'assurer que :

- **Les citoyens adoptent le self data**  
(ex : taux d'utilisation des services),
- **Les services lancés créent de la valeur à la fois pour les citoyens utilisant ces services et pour la collectivité**  
(ex : baisse de la consommation énergétique, diminution des rendez-vous infructueux pour les travailleurs sociaux...),
- **Le portage politique des élus locaux est suffisant pour passer à la phase suivante.**



- Kit Self Data Territorial de la FING : <https://fing.org/publications/kit-self-data-territorial.html>
- My Data France : <http://mesinfos.fing.org/mydata-france/>
- My Data Global : <https://mydata.org/slack><https://mydata.org/slack>
- Bibliothèque de cas d'usage My Data : <http://mesinfos.fing.org/cartographies/usecases/>
- Ressources de la Métropole de Lyon accessibles en open source :
  - Forge du Grand Lyon : <https://forge.grandlyon.com/explore/projects>
  - URL du projet Lyon Living Lab Energy : [vidéo de présentation du service Ecolyo»](#)

## Généralisation de la pratique du self data par les collectivités

Une fois l'infrastructure en place et les prototypes lancés, les collectivités pourront étendre leur périmètre self data en développant de plus en plus de services sur des champs d'action variés afin d'attirer plus de citoyens. Cette montée à l'échelle devra s'accompagner de la stabilisation d'un modèle économique territorial basé sur plusieurs leviers de revenus préalablement mentionnés dans le paragraphe sur les bénéfices financiers attendus du self data.

Cette phase peut également être l'occasion de se rapprocher de partenaires publics, parapublics ou privés démontrant de l'intérêt pour le self data, pour mettre en place un modèle économique hybride qui soit viable dans le long terme. Le choix des partenaires ainsi que leur nombre dépendront de l'approche et des ambitions de chaque collectivité. Chaque cas d'usage pourra attirer des partenaires différents mais il faudra veiller à ce que l'équilibre général du projet self data soit garanti.

Avant de passer à l'étape suivante, les collectivités devront veiller à ce que :

- \_\_\_\_\_ De plus en plus de citoyens adoptent les services self data développés,
- \_\_\_\_\_ Le modèle territorial soit financièrement soutenable,
- \_\_\_\_\_ Les acteurs privés commencent à s'intéresser au concept et qu'il existe des perspectives de valorisation pour eux.

## Adoption du concept par le secteur privé

Une fois le self data généralisé au sein d'une collectivité, celui-ci pourra véritablement être appréhendé comme un service d'intérêt général. Néanmoins, de nombreuses incertitudes demeurent encore à ce stade quant à l'équilibre du modèle économique territorial ; d'où la nécessité pour les collectivités d'avoir recours à la sphère privée afin que celle-ci s'empare du concept et contribue au modèle économique. Pour sensibiliser les acteurs privés au self data et les inciter à s'approprier le sujet, les collectivités peuvent mettre en place des laboratoires d'innovations (data lab) ce qui permettra aux acteurs privés d'identifier comment le self data peut leur apporter de la valeur. L'objectif étant de fédérer les intérêts des acteurs économiques au sein d'une stratégie et gouvernance communes.

Il est toutefois essentiel que les collectivités s'assurent que les acteurs entrant dans l'écosystème soient dans une démarche vertueuse et alignée avec la philosophie du self data ; pour cela, elles pourront mettre en place un cahier des charges et une gouvernance instaurant un protocole pour protéger les données et les individus et garantir les principes du self data (ex : création d'une Charte self data et/ou participation à l'émergence et à la gouvernance d'un label self data pour distinguer les acteurs du territoire qui adopteront ce nouveau mode de traitement des données personnelles)

À terme, ces efforts seront valorisés à travers l'infrastructure self data créée par les collectivités et le secteur privé pourra s'emparer du concept et le développer sur une échelle plus large.

La principale leçon à tirer des expérimentations de la Métropole de Lyon est la nécessité d'aborder un projet self data de façon graduelle ; en effet la bonne pratique consiste à se lancer progressivement dans le sujet en adoptant un développement par paliers. Les collectivités devront mobiliser progressivement des ressources, l'essentiel étant de s'assurer que les porteurs des projets seront engagés et convaincus par les promesses du self data afin qu'ils puissent démocratiser ce concept et garantir un portage politique progressif permettant au self data de se solidifier et de faire ses preuves dans le temps.



**FICHES  
TECHNIQUES**

Ces trois fiches techniques proposent au lecteur d'approfondir certains points critiques de l'étude :

### **Fiche n°1**

La construction des API comme canal de transmission privilégié des données personnelles, du détenteur vers l'individu, et le coût que représente ce chantier.

### **Fiche n°2**

La méthodologie utilisée pour évaluer le potentiel du marché des intermédiaires (clouds personnels et gestionnaires de consentement)

### **Fiche n°3**

Un état des lieux des techniques permettant de minimiser la circulation des données personnelles pour la fourniture de services numériques.

## 01

**Construire une API pour repartager les données personnelles, un coût abordable mais qui nécessite la mise à niveau du système d'information**

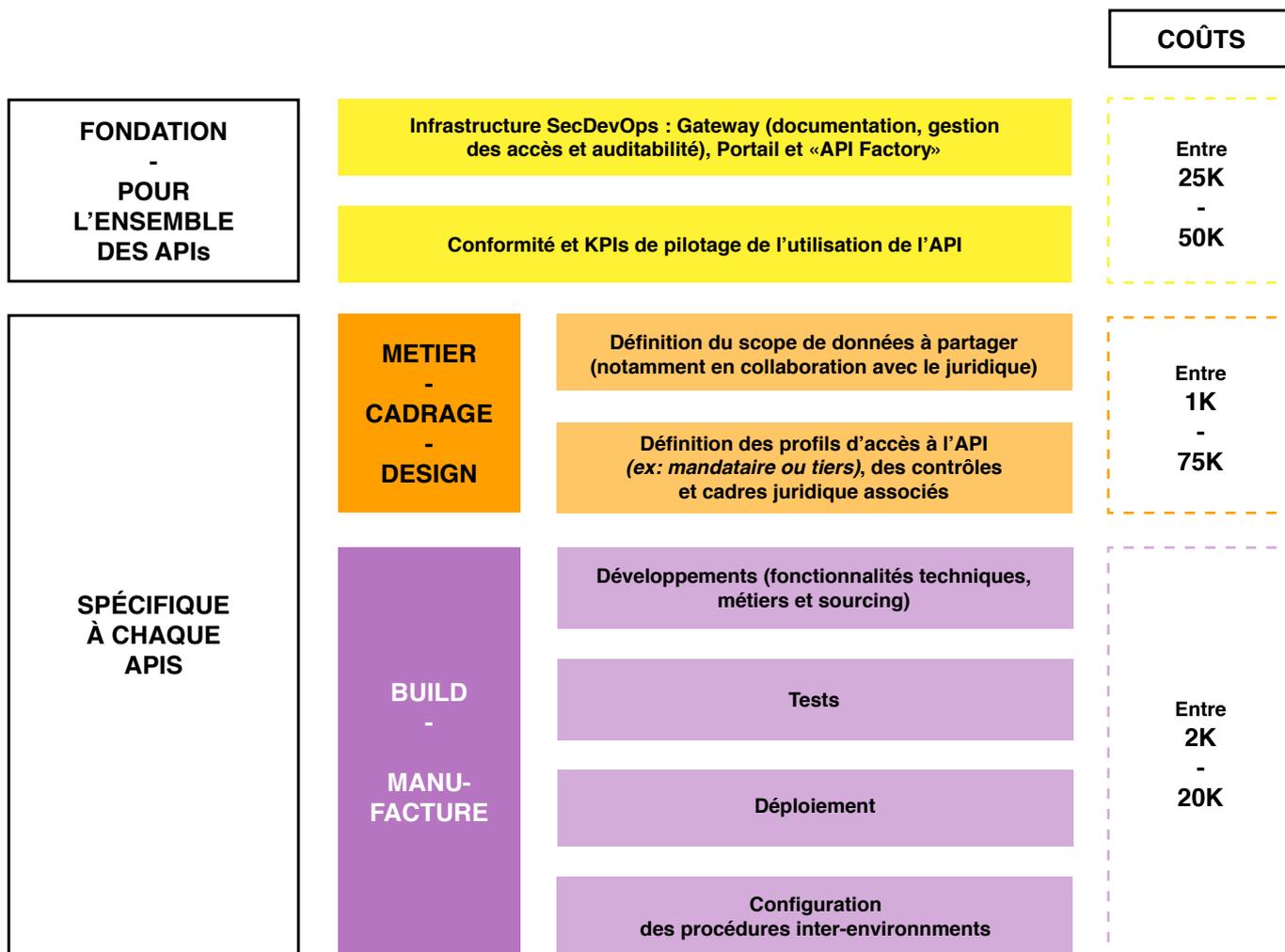
*Les API (Application Programming Interface) sont des outils de communication entre des logiciels, ou entre des ordinateurs et serveurs. Elles permettent des communications automatisées et simples. Elles sont particulièrement pertinentes dans le contexte du self data car elles permettent de demander certaines informations et de les recevoir dans un format établi d'un commun accord entre celui qui demande et celui qui fournit.*

Les API présentent tout de même des limites : elles ne sont que des transmetteurs, et elles ne sont utiles que lorsque les données détenues sont bien structurées ; d'autre part, certains évoquent une possible orientation de l'usage des données par la structure que l'API utilise, qui est généralement établie par le détenteur.

Le coût de mise en place d'une API est assez réduit, restant dans la plupart des cas en dessous de la barre des 100 000 € pour une création complète et personnalisée.

Si elles venaient à se généraliser, les entreprises qui opèrent des bases de données sur des outils standards pourraient acquérir des API déjà paramétrées beaucoup moins onéreuses. Ceci reste donc très abordable. En revanche, la structuration des bases de données nécessaires au bon fonctionnement des API s'évalue plutôt en millions, surtout pour les grandes entreprises. Mais il serait injuste d'affecter ce montant au self data puisqu'il s'agit d'un travail de fond qui permet aussi aux entreprises de pouvoir exploiter elles-mêmes leurs données, ce qui les intéresse d'ores et déjà.

Les API sont essentiellement le dernier maillon d'une chaîne de classification, de stockage, de gestion et de transmission efficace des données.



-----

*Critères différenciants*

-----



- On premise ou pas
- Niveau de complexité en termes d'authentification et d'auditabilité



- Taille de l'organisation
- Capacité d'alignement des décisionnaires
- Disponible du juridique



- Modernité / Legacy des systèmes d'information, notamment pour la partie sourcing
- Nombre des fonctionnalités techniques & métiers et niveau de complexité associé

# 02

## Méthodologie de l'évaluation du marché des intermédiaires

### Marchés similaires

Deux méthodes sont utilisées pour déterminer un ordre de grandeur de la valeur du marché.

D'abord, pour en fixer le contexte, une comparaison est possible avec trois types de services : les clouds personnels, les gestionnaires de consentements pour les entreprises, et les gestionnaires d'identité. Les estimations sont faites au niveau mondial

ESTIMATION	MARCHÉ	ÉTUDE
32,3Md€	Stockage sécurisé personnel	MarketsandMarkets (2020)
199M€	Gestionnaire de consentement pour entreprise	The Insight Partner (2019)
13Md€	Gestionnaires d'identité	MarketsandMarkets (2020)

Stockage sécurisé personnel	Proximité du marché	Gestionnaires de consentements des entreprises	Proximité du marché
Stockage local sur un disque dur (ordinateur...)	●	Gestionnaires de consentements	● ● ●
Solution de « cloud personnel »	● ●	Logiciels de gestion de la vie privée	●
Solution open-source de son propre cloud	● ●	Identity & Access Management (IAM)	●

## Méthode utilisée par Ctrl Shift en 2014

Le principe de la méthode est de distinguer des relation client-fournisseur par personne et par foyer, et d'évaluer la rémunération qu'un PIMS peut retirer de la création du lien correspondant. L'estimation se base sur la formule :

**Population**<sub>individuelle</sub> : le nombre de personnes étant concernées par un tel service (excluant les jeunes enfants)

**Relations**<sub>individuelles</sub> : le nombre de connexions qui pourraient être réalisés par an dans le cadre du self data entre une personne et un service (évaluation faite par Ctrl Shift en 2014)

**Population**<sub>ménages</sub> : le nombre de ménages (mesure statistique)

**Relations**<sub>ménages</sub> : le nombre de connexions qui pourraient être réalisées par an entre un ménage et un service (évaluation faite par Ctrl Shift en 2014)

**Valeur**<sub>connexion</sub> : valeur monétaire que l'intermédiaire pourrait facturer pour la création et l'entretien de la connexion (évaluation faite par Ctrl Shift en 2014 ajustée pour la France)

	ROYAUME-UNI	FRANCE
<b>Population</b> <sub>individuelle</sub>	52,1M	52,4M
<b>Relations</b> <sub>individuelles</sub>	45	45
<b>Population</b> <sub>ménages</sub>	24,5M	29M
<b>Relations</b> <sub>ménages</sub>	20	20
<b>Valeur</b> <sub>connexion</sub>	£4	3,5€
<hr style="border-top: 1px dashed black;"/>		
<b>Estimation du marché</b>	£11,5Md (14,6Md€)	10,3Md€

## Méthode utilisée par OneCub et Visions

Le principe de la méthode est d'évaluer la valeur globale du transfert des données et de considérer que les PIMS peuvent obtenir une partie de cette valeur comme rémunération. Elle se base sur la formule :

- : La population totale de l'aire géographique concernée
- : part des individus dans la population qui auraient recours à un service self data
- : le nombre de services digitaux avec lesquels un individu est en interaction chaque année
- : le nombre de transferts par an et par service nécessaires à un bon fonctionnement (en moyenne)
- : valeur d'un transfert de données
- : part de la valeur que le PIMS pourrait obtenir

	UNION-EUROPÉENNE	FRANCE
<i>Population</i> <sub>totale</sub>	512M	65M
<i>Taux</i> <sub>pénétration</sub>	85%	
<i>Nombre</i> <sub>services</sub>	45	
<i>Interractions</i> <sub>annuelles</sub>	50	
<i>Valeur</i> <sub>transfert</sub>	0,50€	
<i>Commission</i> <sub>intermédiaire</sub>	3%	
<hr/>		
<i>Estimation du marché</i>	14,7Md€	1,86Md€

## Comparaison des deux méthodes

Deux points expliquent la différence de résultat entre les méthodes (10,3Md€ avec la méthode de Ctrl Shift et 1,85Md€ avec la méthode proposée par OneCub et Visions).

Les revenus du PIMS par service sont très différents : Ctrl Shift propose une valeur nominale annuelle de £4, revue aujourd'hui à 3,5€ à l'occasion de l'étude là où l'autre méthode propose une valorisation à . Dans la première méthode, un PIMS pourrait facturer 3 ,5€ par personne, par entreprise par an ce qui est évalué dans la seconde à 0,75€

Le dénombrement des interactions : la méthode utilisée par Ctrl Shift peut être discutée sur la pertinence du double comptage avec d'une part les ménages et d'autre part les individus. Il reste complexe d'identifier ce qui correspond au ménage et ce qui correspond à l'individus dans des cas simples comme l'achat en hypermarché ou bien l'achat de mobilier.

Enfin, la méthode proposée par les acteurs français permet une adaptation en fonction de la part de la population qui adopterait des usages self data :

<b>TAUX DE PÉNÉTRATION</b>	<b>85%</b>	<b>30%</b>	<b>3%</b>
<b>ESTIMATION DU MARCHÉ</b>	<b>1,86Md€</b>	<b>658M€</b>	<b>66M€</b>

## 03

**Les techniques pour développer des services plus respectueux de la vie privée : privacy by design, privacy by default, calcul à la source et preuve à divulgation de connaissance nulle***Privacy by design et privacy by default*

Depuis plusieurs années, deux mouvements de conception de programmes, et d'interfaces se sont développées afin de mieux respecter et protéger la vie privée des individus.

Le privacy by design est essentiellement un principe de minimisation des données collectées, le principe est de ne recueillir sur une personne que les informations strictement nécessaires. Ce mouvement s'est en partie développé en opposition à une collecte massive des données personnelles dans une dynamique de précaution qui ne donnait lieu à aucune utilisation, créant un risque de fuite sans aucun intérêt.

Le privacy by default est un mouvement plus graduel : le principe consiste à ne recueillir en première intention que les données minimales nécessaires, mais de permettre à l'utilisateur de fournir plus de données pour obtenir plus de services. La fourniture de données prend donc la forme d'un opt-in : il faut faire une démarche active pour fournir plus de données, alors que dans la plupart des interfaces, l'utilisateur doit s'ingénier s'il veut fournir le moins de données possible.

Ces mouvements s'inscrivent aussi dans un courant visant à ne demander de s'identifier que si nécessaire là où beaucoup de sites internet par exemple demandent de se connecter pour accéder à un contenu pourtant public. En minimisant les données, ils simplifient leur récupération et préservent la vie privée de chacun, deux objectifs majeurs du self data.

*Des méthodes de calcul décentralisées « à la source »*

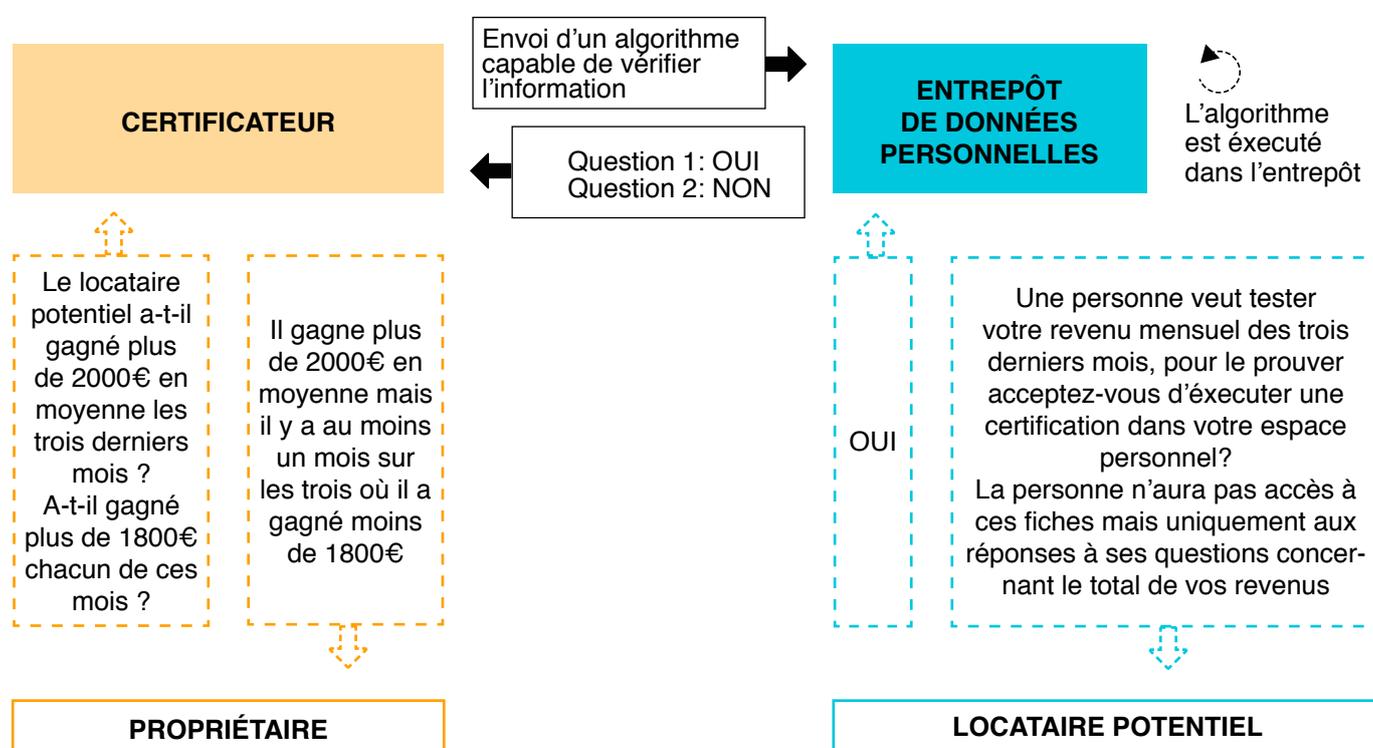
Il existe des protocoles permettant d'entraîner l'intelligence artificielle sans avoir à traiter de la donnée personnelle. C'est le cas, par exemple, du protocole DISPERS. Cozy cloud présente le fonctionnement de DISPERS sur son blog : « DISPERS propose ainsi une manière de distribuer les tâches et informations entre plusieurs domiciles numériques Cozy qui seront coréalisateurs du traitement. Le protocole incorpore des procédés pour cacher les détenteurs de l'information ou rendre les données incompréhensibles par l'acteur qui sera responsable d'une sous-tâche. Ainsi aucun acteur n'est en mesure de compromettre le calcul et, dans une certaine limite, ce qu'il manipule ! » (*extraits, Blog cozy cloud*). Ce type de solution permet de concevoir des services qui utiliseraient les données personnelles sans avoir besoin de les voir ou de les copier, préservant ainsi l'exposition de la vie privée des personnes.

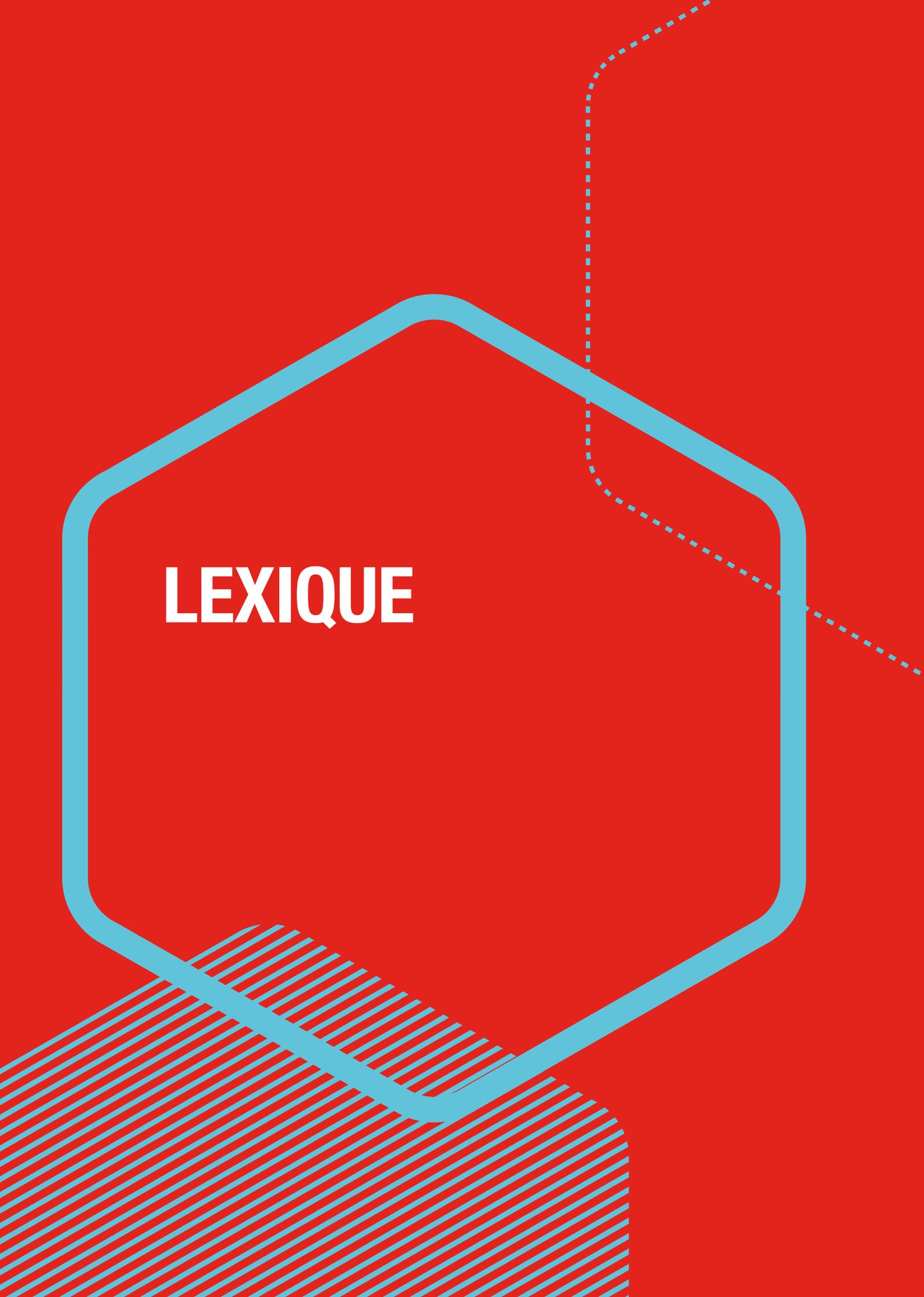
## Les protocoles ZKP : un atout pour la protection de la vie privée

Les procédures de preuve à connaissance nulle Zero knowledge proof (ZKP) ou « preuve à divulgation de connaissance nulle » sont apparues pour la première fois en 1985 au Massachusetts Institute of Technology (MIT). Elles permettent de donner l'information demandée sans avoir à divulguer plus d'informations détenues sur l'individu, ce qui rejoint les principes de minimisation de la donnée (ne donner que ce qui est strictement nécessaire au traitement) et de sécurité des données (ce qui transite est seulement le résultat).

Cette technique peut être illustrée par un exemple simple : lorsque l'on souhaite louer un logement, le propriétaire demande des garanties pour vérifier les revenus suffisants du locataire, pour se faire, essentiellement deux moyens, les fiches de paie ou les avis d'imposition. Mais ces documents contiennent énormément d'informations personnelles que le propriétaire n'a pas besoin de connaître alors qu'il cherche uniquement la réponse à la question « gagne-t-il plus de X € par mois ? ». Un schéma de preuve à connaissance nulle repose sur une réponse sur un certificateur, une entité qui possède la confiance des deux individus. Le locataire potentiel doit reconnaître que le certificateur n'est pas malveillant afin qu'il s'exécute dans son espace. Le propriétaire doit faire confiance au certificateur pour avoir réellement accédé à une preuve fiable.

Dans le cas de la recherche de logement, la démarche est assez simple, mais pour l'obtention d'un crédit par exemple, elle peut permettre d'accéder à plus de vérifications (comme la moyenne, le minimum, ou l'écart type des revenus sur une période) en maintenant une protection très forte du demandeur.





**LEXIQUE**

## **DONNÉE PERSONNELLE :**

est toute information se rapportant à une personne physique identifiée ou identifiable directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale.

## **AUTODÉTERMINATION**

### **INFORMATIONNELLE :**

capacité pour une personne d'être maître des informations qui la concernent, c'est-à-dire de pouvoir déterminer qui utilise des informations la concernant, qui les propage et à quelles fins.

## **DÉTENTEUR DE DONNÉES :**

personne morale qui détient les données personnelles des individus/utilisateurs, qu'elles aient été captées (ex. traces), recueillies (ex. formulaires) ou coproduites (ex. transaction).

# Les acteurs du self data

## **RÉ-UTILISATEUR DE DONNÉES :**

entreprise qui propose un service nécessitant différentes données sur le client qui ne sont pas obtenues en lui posant des questions, mais en lui demandant d'effectuer une portabilité de ces données déjà enregistrées par d'autres détenteurs

## **INTERMÉDIAIRE (AUSSI APPELÉ PERSONAL INFORMATION MANAGEMENT SYSTEM) :**

opérateur d'infrastructure essentielle du self data qui permet aux individus de récupérer, stocker et réutiliser leurs données sous leur contrôle et à leurs propres fins, il fait le lien entre l'individu, le détenteur de données personnelles et le service ré-utilisateur de ces données. Il se positionne comme un tiers de confiance qui remet l'individu au centre en lui redonnant le contrôle sur les données que les tiers détiennent sur lui. Il en existe différents types, principalement les gestionnaires de consentement et les entrepôts de données personnelles.

## **GESTIONNAIRE DE CONSENTEMENT :**

opérateur qui permet d'exécuter la volonté de l'utilisateur. L'utilisateur déclare au gestionnaire quelles données il veut transmettre à quel ré-utilisateur, et le gestionnaire de consentement se charge de créer le dialogue technique entre le détenteur des données et le ré-utilisateur qui en a besoin. Il crée le canal de transmission mais ne stocke pas les données.

## **ENTREPÔT DE DONNÉES PERSONNELLES**

### **(OU ENTREPÔT PERSONNELLE DE DONNÉES) :**

opérateur qui permet aux usagers de récupérer leurs données personnelles auprès des détenteurs et de les stocker sur un espace qui leur est propre. Les usagers peuvent ensuite autoriser différents ré-utilisateurs à accéder à certaines de leurs données ou à exécuter des algorithmes au sein de leur espace personnel.

# Types de marchés

## **B2B, « BUSINESS TO BUSINESS » :**

désigne un marché où les entreprises vendent aux entreprises.

## **B2C, « BUSINESS TO CUSTOMER » :**

désigne un marché où les entreprises vendent à des clients individuels.

## **B2B2C :**

désigne un marché où une entreprise A vend un service à une entreprise B dans le seul but de le mettre à disposition des clients ou employés de l'entreprise B pour leur usage. Par exemple, les tickets restaurants sont vendus à l'entreprise mais utilisés par l'employé.

**FORMAT DE DONNÉES :**

convention (éventuellement normalisée) utilisée pour représenter des données (des informations représentant un texte, une page, une image, un son, etc...). Lorsque ces données sont stockées dans un fichier, on parle de format de fichiers. Une telle convention permet d'échanger des données entre divers programmes informatiques ou logiciels, soit par une connexion directe, soit par l'intermédiaire d'un fichier. On appelle interopérabilité cette possibilité d'échanger des données entre différents logiciels.

**FIRST PARTY DATA :**

données utilisées par l'entreprise qui les a collectées.

**SECOND PARTY DATA :**

données utilisées par une entreprise qui les a acquises auprès d'une autre entreprise qui les a directement collectées auprès de l'individu. La transmission peut se faire dans le cadre d'un partenariat (les deux entreprises partagent leurs informations) ou bien contre rémunération pour l'entreprise qui a collecté les données

**THIRD PARTY DATA :**

données qui résultent du croisement de données collectées par différentes entreprises qui les transmettent ensuite à des agrégateurs. Ce traitement permet d'accumuler un grand nombre de données sur la même personne et de reconstituer son « profil ». Ces données sont principalement vendues dans des buts de publicité ciblée.

**COOKIES :**

fichier texte déposé sur le disque dur de l'individu par le serveur du site visité ou par un serveur tiers (régie publicitaire, service de web analytique, etc.). Le navigateur web le conservera pendant une certaine durée, et le renverra au serveur web chaque fois que l'individu se reconnecte. Les cookies tiers permettent d'éviter de devoir se reconnecter à chaque fois que la fenêtre du navigateur est fermée, mais aussi un suivi très fin de la navigation web qui peut être utilisé à des fins publicitaires.

**PUBLICITÉ CIBLÉE :**

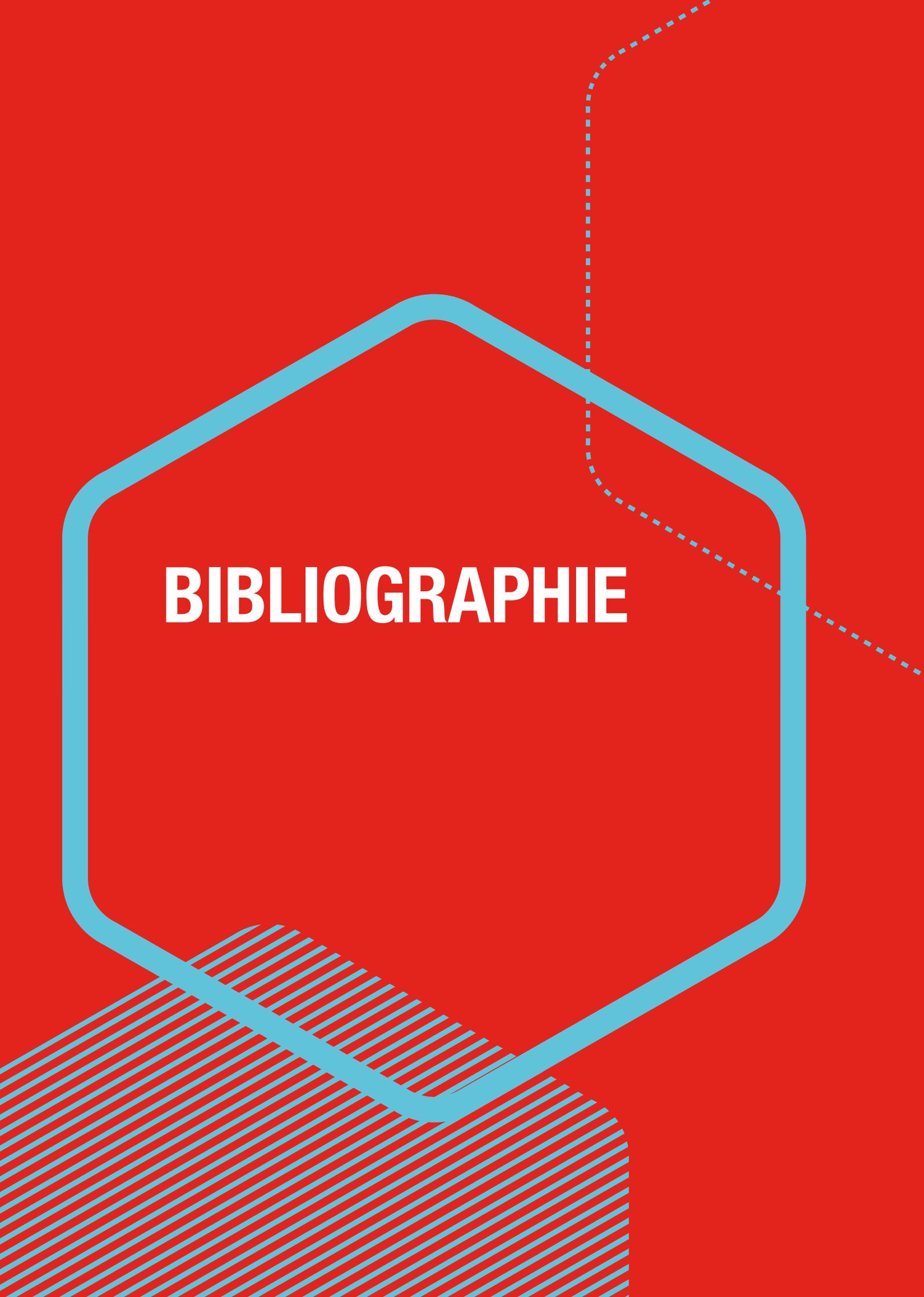
pratique publicitaire qui adapte le contenu d'une fenêtre publicitaire au potentiel client grâce aux différentes données acquises sur ce potentiel client.

**RE-CIBLAGE PUBLICITAIRE :**

pratique consistant à placer dans les fenêtres de publicité d'un utilisateur des annonces pour un produit dont il a commencé mais pas achevé l'achat (temps passé sur la page, ajout au panier, etc.) afin de le persuader de concrétiser son achat.

**BLOQUEUR DE PUBLICITÉ (OU ADBLOCKER) :**

outil intégré au navigateur internet ou qui s'y ajoute afin de bloquer un maximum de publicité (fenêtres qui apparaissent, ouverture non-souhaitée de nouveaux onglets, publicités pendant le visionnage de vidéos, etc.)



# **BIBLIOGRAPHIE**

---

## Études et publications

« **De la DSP2 à l'Open Banking, les impacts sur le modèle traditionnel bancaire** », VIATYS, A. DHELLEMMES, T. BERARD, P.KOLLEN, M.TASTES

« **MesInfos Santé, vers un Blue Button à la Française** », FING, 2015

« **MesInfos energies, Approfondir la richesse et la variété des usages possibles autour des données de d'énergies 5 concepts de services** », FING, Décembre 2015

« **Les consommateurs d'énergie et la transformation numérique** », Comité de prospective de la CRE, rapport du groupe de travail n°3, GT3, juillet 2018

« **Précarité énergétique face au défi des données** », T. ERARD, L. CHANCEL, M. SAUJOT, IDDRI Sciences Po, n°15.15 avril 2015

« **Personal Information Management Services : An analysis of an emerging market** », CtrlShift, juin 2014

« **The business opportunity of Personal Information Management Services** », Ctrl Shift, 26 juin 2014

« **Data mobility : the personal data portability growth opportunity for the UK economy** », Ctrl Shift, 2018

« **A Roadmap for a Fair Data Economy** », SITRA

« **Self Data : être acteur de ses données personnelles. Une réflexion portée par un Think Tank, mobilisant des techniques de conception collaborative, et explorant le sujet des données personnelles à l'usage des individus qui les produisent** » M. MOLINS, 2015

« **Cahier n°1 : identités numériques** », C.LEVALLOIS-BARTH, Chaire Valeurs et Politiques des Informations Personnelles, Institut Mines-Telecom.

« **Choose the Right API Monetization and Pricing Model** », Gartner, 4 mars 2019,

« **Data People Cities** » Eurocities citizen data principles in action, 2019

---

## Les indispensables pour comprendre le self data : les grands textes clés du self data

« **Kit Self, Data Territorial** » - FING, 2019

« **Self Data , cahier d'exploration MesInfos** » - 2ème édition, FING, mai 2019

« **Pilote MesInfos** » - FING, 2016-2018

« **Rapport de synthèse de l'expérimentation Mes données, Ma santé** » - FING, juillet 2019

« **Data portability, policy paper** » - Stiftung Datenschutz, 2019

« **Data Portability : What is at stake** » Datum future, Juillet 2019

« **MyData : a Nordic Model for human-centered personal data management and processing** » - Lois, réglementations, CNIL, avis et lignes directrices

---

Lois, réglementations,  
CNIL, avis  
et lignes directrices

**Directive 95/46/CE relative à la protection des données personnelles**

**Règlement n° 2016/679, dit règlement général sur la protection des données**

**Lignes directrices sur le consentement**, WP259 rev.01, 10 avril 2018

**Lignes directrices relatives au droit à la portabilité des données**, WP242, rev.01, le 5 avril 2017

**Avis du CEPD sur les systèmes de gestion des informations personnelles**, Vers une plus grande autonomie des utilisateurs dans la gestion et le traitement des données à caractère personnel, 20 octobre 2016

**« La plateforme d'une ville, les données personnelles au cœur de la fabrique de la smart city »**, Cahier IP innovation & prospective n°5, CNIL, LINC, septembre 2017

**« La forme des choix, données personnelles, design et frictions désirables »**, Cahier IP innovation & prospective, n°6,

CNIL, LINC, janvier 2019  
**« Le droit à la portabilité en question »**, CNIL, 22 mai 2017

**« Premiers éléments d'analyse de la CNIL sur la Blockchain »**, septembre 2018, CNIL

**« Communication de la commission au parlement européen, au conseil, au comité économique et social européen et au comité des régions - Une stratégie européenne pour les données »**, Commission européenne, 19 février 2020,

**« Etude sur les nouveaux enjeux liés aux questions de propriété accès, interopérabilité, réutilisation des données et questions de responsabilité associée à leur utilisation »**, Commission européenne, 25 avril 2018

**« Study on data sharing between companies in Europe »**, Commission européenne, 24 avril 2018

---

## Articles de presse généraliste et spécialisée

### Web, techno & GAFAM

« **L'affaire Facebook-Cambridge Analytica montre que nos données personnelles ont un impact collectif** », L'Usine Nouvelle, J.F. Marchandise de la FING, 26 mars 2018

« **Facebook permet de transférer des photos et des vidéos de sa plateforme vers Google** », (Data Transfer Project), Le Figaro Premium, I. VERGARA, 2 décembre 2019

« **La portabilité des données est la meilleure arme antitrust dans le numérique** », Le Monde Eco et entreprise, 13 décembre 2018

« **Les ZKPs, « Zero Knowledge Proof », une nouvelle ère pour le partage des données** », Medium, R. ACHACH, 27 septembre 2019

« **Personal Information management systems : a new era for individual privacy** », IAPP, A. LOAYZA, 21 mars 201

« **Parts de marché des moteurs de recherche (France, USA, Monde)** », Web Rank Info, dernière mise à jour du 2 mars 2020

« **Entraîner une IA sans posséder la donnée est possible** », M. MASSON, Cozy cloud (blog), 10 septmebre 2019

### Villes et smart city

« **A Lyon, deux mille personnes testent déjà la « portabilité » de leurs données personnelles** », Le Monde, C. LEGROS, 25 mai 2018

« **Rassembler les données pour gérer la ville** », L'Usine Nouvelle, 3 décembre 2018

« **Le Grand Lyon rend ses données à ses citoyens** », La Gazette des communes, G. ZIGNANI, 30 juillet 2018

« **Smart City : rassembler les données en vue de piloter la ville** », La Gazette des communes, B. CESSIEUX, 6 février 2018

« **Self data : un guide pour mettre les données personnes à disposition des habitants** », La Gazette des communes, G. ZIGNANI, 29 octobre 2019

« **Le self data met l'habitant au centre des décisions** », La Gazette des communes, G. ZIGNANI, 22 octobre 2018

« **DOSSIER / Les données réinventent les politiques publiques.**

**Le nouveau couteau suisse de l'action publique** », La Gazette des communes, 27 mai 2019

« **Le Partage des données, entre fluidité et surveillance** », Le Monde, C. LEGROS, 28 mars 2019

« **Self data territorial, explorer le potentiel du self data pour un territoire et ses citoyens** », M. MOLINS, La Fabrique des mobilités, 3 décembre 2018

### Données personnelles

« **Vendre ses données personnelles, un marché d'avenir ?** », Capital, cybersécurité, 16 avril 2019

« **La CNIL dézingue l'idée de revendre nos propres données personnelles** », Le Point, entretien de Isabelle Falque-Pierrotin », 30 janvier 2018

« **Les défis éthiques de l'identité numérique** », The Conversation, 3 mars 2019,

« **Identité numérique : la stratégie officielle se dessine** », Le Figaro, G. GUICHARD, 10 mars 2020

---

## Sondages

« **Baromètre générique sur les pratiques numériques et la maîtrise des données personnelles** », Médiamétrie, CNIL, 14 octobre 2019

« **Regard des Français sur la protection des données personnelles** », Ifop pour Reconnaissance Numérique, Avril 2018

« **Baromètre 2019 de la confiance des Français dans le numérique** », ASCEL, octobre 2019

« **Penser les territoires autrement, observatoire des usages et représentations des territoires** », L'OBSOCO, vague 2 janvier 2020

« **Cyber Safety : Insights Report Global Result** », Norton LifeLock, 2019

« **Données personnelles : les français se disent préoccupés par leur utilisation mais ne se prémunissent pas toujours** », Odoxa, 2019

« **Les français et les scandales liés aux données personnelles** », Opinion Way, juin 2019,

« **Les français et les fonctionnaires territoriaux** », ADRHGCT, octobre 2019

« **Six in Ten Consumers Willing to Share Significant Personal Data with Banks and Insurers in Exchange for Lower Pricing** », Accenture, 14 mars 2019

« **Est-ce que j'accepte de partager mes données de santé ? Baromètre ACCORDS, acceptabilité du partage de données de santé** ». Accords, baromètre données de santé de l'AFCROs, communiqué de presse, 18 juin 2019.

« **Why data trust is a priority in 2019** », PWC, 2019

---

## Remerciements

Nous remercions les entités et les personnes qui ont participé à cette étude en répondant à nos sollicitations pour des entretiens :

**Benjamin André** : Cozy Cloud, CEO, Co-founder - **Chloé Beaumont** : MAIF, responsable du Hub, de la Veille et de la Prospective - **Liz Brandt** : Ctrl Shift, CEO - **Régis Chatellier** : Laboratoire LINC - **Fabien Coutant** : Enedis, Chef de projet Gouvernance des Données - **Stéphane Cren** : GS1, Cas d'usage portabilité & Retail - **Caroline Dias** : Pôle emploi, Directrice de l'innovation et de la portabilité - **Bertrand Dolbeau** : La Poste, Directeur Marketing et Commercial Confiance Numérique - **Romain Gautier** : Didomi, CEO de Didomi - **Perrine Hagen** : GRDF, Responsable équipe data - **Sylvie Kermabon** : Groupe Vyv, Directrice de projet - **Xavier Lefevre** : CEO, Fair&Smart - **Maria Ines LEAL** : Chef de projet Développement des usages et services numériques, Métropole de Lyon - **Sarah Medjek** : FING, Présidente du Conseil d'administration de MyData - **Rusama Mikko** : Helsinki, CDO Ville d'Helsinki - **Géraldine Salord** : Avocate associée du cabinet Metalaw - **Jaana Sinipuro** : Directrice de projets, SITRA, Fonds public d'investissement pour l'innovation de la Banque de Finlande - **Patrice Slupowski** : Orange, EXPERT BUSINESS - **Virginie Steiner** : Administratrice générale des données, Ville et Communauté d'agglomération de La Rochelle - **Petri Tuomela** : Business owner, Vastuu - **Christian-François Viala** : Expert (indépendant), interrogé Yes profile

Cette étude a été réalisée par la Banque des Territoires avec Capgemini Invent (l'entité de conseil, transformation et innovation digitale du Groupe Capgemini), OneCub et Vision et avec la contribution de la Métropole de Lyon et le cabinet Wavestone.



BANQUE des  
**TERRITOIRES**



[banquedesterritoires.fr](https://banquedesterritoires.fr)



@BanqueDesTerr