

01 LES CAHIERS DE L'OBSERVAT OIRE

DONNÉES, INTÉRÊT GÉNÉRAL
& TERRITOIRES :
LA CONSTRUCTION
D'UN CADRE DE CONFIANCE





Ce cahier a été préparé par l'Observatoire Data Publica et a été publié en février 2022. Ses auteurs principaux sont Aurélie Legrand & Jacques Priol, avec les contributions de Anne-Sophie Hulin, Schéhérazade Abboub, Mathieu Caps & Joël Combin.

La conception & la réalisation graphique ont été réalisées par Olivier Legrand.

La publication des cahiers de l'Observatoire Data Publica bénéficie du soutien de la Banque des territoires.



BANQUE des
TERRITOIRES



DONNÉE, INTÉRÊT GÉNÉRAL & TERRITOIRES : LA CONSTRUCTION D'UN CADRE DE CONFIANCE

INTRODUCTION	07
LA DONNÉE AU SERVICE DE L'INTÉRÊT GÉNÉRAL	09
La donnée, les politiques publiques et le territoire	09
De quelles données parle-t-on ?	13
De l' <i>open data</i> aux <i>closed data</i>	16
Acteurs publics et privés : de nouvelles coopérations à inventer	18
Données personnelles et intérêt général : ambivalence et paradoxes	21
CONSTRUIRE UN CADRE DE CONFIANCE AU SERVICE DE L'INTÉRÊT GÉNÉRAL	27
Des règles et des objectifs	28
Des méthodes et outils	35
Le rôle des citoyens	41
Entre construction rationnelle et expérience subjective : prendre du temps	44
LES TIERS DE CONFIANCE	45
« Tiers de confiance » ou « acteur de confiance » ?	46
Les data trusts	48
Construire un tiers de confiance	53
CONCLUSION	58
REMERCIEMENTS	59
L'OBSERVATOIRE DATA PUBLICA	60

INTRODUCTION

EN janvier 2020, il y a exactement 2 ans, l'Observatoire Data Publica organisait à Paris dans les locaux du laboratoire d'innovation la 27^{ème} Région (Superpublic), une journée de réflexion autour d'une question : « Data trust : quelle définition en France ? ». Le sujet avait mobilisé une trentaine d'acteurs publics locaux à un moment où quelques rares projets étaient en gestation. Au niveau national, le Health data hub venait tout juste d'être créé pour rassembler et gérer les données du système de santé français, et personne n'imaginait l'ampleur des enjeux (et des polémiques) auxquels il devrait faire face quelques mois plus tard dans un contexte exacerbé par la pandémie. Au niveau local, la Région Occitanie avait créé l'association « Occitanie Data » et engageait une réflexion sur différents modèles possibles pour orchestrer un partage éthique de la donnée à l'échelon régional. La Région Centre Val-de-Loire envisageait de faire de même pour piloter ses politiques de transition à travers la construction d'un *climate data hub*. Rennes Métropole tirait les premiers enseignements de sa démarche de Service public métropolitain de la donnée et lançait son projet de plateforme RUDI. Nantes avait promulgué la première Charte territoriale de la donnée française.

Depuis, certaines de ces initiatives ont prospéré. D'autres sont nées, y compris dans des territoires de taille plus modeste. Avec à chaque fois une idée directrice : face à l'accroissement du volume des données produites par l'action publique et des données disponibles pour l'orienter, comment organiser en confiance et dans le respect de règles éthiques, l'accès, le partage et l'échange des données ? À quelles conditions ? Avec qui ? Ces réflexions ne sont pas propres à la France, loin s'en faut. Des travaux importants sont menés sur le sujet dans plusieurs pays depuis quelques années. Citons notamment ceux de l'Open Data Institute (ODI) au Royaume-Uni, de l'AAPTI Institute en Inde ou ceux de nos partenaires Nord Ouvert au Canada. Ce n'est d'ailleurs pas un hasard si en janvier 2020 le Professeur Christophe Abrassart de l'Université de Montréal participait à nos travaux.

La notion anglo-saxonne de *data trust* (ou de fiducie de donnée) peut nous inspirer ici en France. Mais elle repose sur des fondements juridiques qui sont ceux de la *common law* et ils trouvent difficilement à se transposer dans notre droit continental, *a fortiori* dans le droit public français. C'est d'ailleurs l'une des difficultés à laquelle sont confrontés les acteurs du Partenariat mondial sur l'intelligence artificielle, plus connu sous le nom de GPAI¹. Au sein de cette initiative internationale lancée sous l'impulsion de la France et du Canada (et qui regroupe aujourd'hui 25 membres²) un groupe de travail cherche à construire un cadre de référence qui permettra demain à des acteurs publics et privés

de mettre en commun des données au service du bien public. Dans la perspective d'un développement généralisé du recours à l'intelligence artificielle, les enjeux sont majeurs. Il s'agit d'inventer les modalités méthodologiques, techniques, économiques ou juridiques qui garantiront aux acteurs qui mettent des données à disposition à des fins d'intérêt général que les usages qui en seront faits ne s'avéreront pas contraires à cet intérêt collectif.

Au niveau européen, dans le cadre de la préparation du futur *Data Governance Act*³, un autre concept fait son apparition : celui d'altruisme des données. Cet « altruisme » pourra concerner des entreprises privées qui mettront leurs données à disposition au service de l'intérêt général ; des acteurs publics bien sûr, au-delà de leurs obligations réaffirmées d'*open data* ; ou encore des individus qui donneront ainsi leur consentement « RGPD compatible » à l'exploitation de leurs données personnelles, par exemple au service de la recherche scientifique ou pour l'amélioration des politiques publiques.

L'objet de ce cahier n'est pas de rendre compte de ces multiples travaux, ni d'anticiper ces évolutions à venir. Mais il était utile de les évoquer.

L'Observatoire Data Publica propose pour sa part de faire un état des lieux et de décrire la manière dont des collectivités territoriales françaises se saisissent de la question de la confiance dans l'usage qu'elles font des données. Nous l'avons fait à travers divers prismes méthodologiques et avec une préoccupation particulière : analyser la place accordée aux citoyens dans la définition d'objectifs, la gouvernance et le contrôle des actions menées au service de l'intérêt général, donc en leur nom.

Nous espérons que ce 1^{er} cahier de l'Observatoire Data Publica vous sera utile.



1. Global Partnership on Artificial Intelligence.

2. Les membres du GPAI sont l'Allemagne, l'Australie, la Belgique, le Brésil, le Canada, la République de Corée, le Danemark, l'Espagne, les États-Unis, la France, l'Inde, l'Irlande, Israël, l'Italie, le Japon, le Mexique, la Nouvelle-Zélande, les Pays-Bas, la Pologne, la République tchèque, le Royaume-Uni, Singapour, la Slovaquie, la Suède et l'Union européenne.

3. Le Data Governance act est un projet de règlement européen proposé fin 2020 par la Commission européenne, en cours de discussion après une concertation publique, dont l'adoption par le Parlement.

LA DONNÉE AU SERVICE DE L'INTÉRÊT GÉNÉRAL

LA DONNÉE, LES POLITIQUES PUBLIQUES ET LE TERRITOIRE

LE volume des données produites, transférées, stockées ou traitées dans le monde emprunte une courbe exponentielle. Il est communément admis que ce volume double tous les deux ans. Sous l'influence d'une digitalisation accrue, d'une dématérialisation systématisée et dorénavant de stratégies de « territoire intelligent⁴ », les données produites et utilisées par les collectivités territoriales croissent à leur tour. Cette évolution est peu ou mal quantifiée, notamment du fait de la multiplicité des flux de données et de la diversité des acteurs impliqués. Les données produites pour mettre en œuvre les politiques publiques d'un territoire peuvent être gérées en régie ou non, hébergées par la collectivité ou des opérateurs, recourir plus ou moins massivement au *cloud* et au mode *SaaS*... Faute d'étude quantifiée fiable, il est relativement intuitif d'imaginer que pour certains métiers le volume des données produites croît de façon très rapide. Ainsi par exemple, le suivi des transports en commun en temps réel (pour une meilleure information des voyageurs et une optimisation des dessertes), génère des volumes de données incomparables à ceux des systèmes anciens.

À ces données produites par les services publics locaux s'ajoutent des données produites par de multiples acteurs publics ou privés : celles du service public national de la donnée (l'État, l'INSEE, l'IGN...), celles d'entreprises délégataires de missions de service public, celles d'entreprises privées dont l'action impacte la vie de la cité (Waze, AirBnb et bien d'autres), sans oublier toutes celles produites par les habitants, certes protégées par le RGPD, mais massivement disponibles à l'échelle des territoires.

Cette explosion participe d'une transition numérique qui impacte l'ensemble de notre société, les services publics y compris. Elle pose de nombreuses questions. Elles sont juridiques, éthiques, environnementales, démocratiques ou politiques.

Ces nouveaux usages de la donnée (ou les usages de nouvelles données) se justifient-ils ? La question mérite d'être posée. D'autant que les investissements pour les produire, les capter, les transporter, les stocker, les traiter sont significatifs. L'utilisation massive de données, le recours à des outils algorithmiques et demain à l'intelligence artificielle, sont le plus souvent motivés par une combinaison d'objectifs.

4. Voir notamment : *De la smart city à la réalité des territoires connectés*, rapport officiel, ministère de l'Économie et des finances, octobre 2021.

La donnée sert d'abord à améliorer la connaissance d'un territoire. Elle sert ensuite à concevoir ou adapter des politiques publiques. Elle sert aussi à optimiser la gestion opérationnelle des services, avec parfois des gains significatifs (des gains de productivité et financiers, mais aussi des économies d'eau ou d'énergie). Elle sert bien sûr à améliorer le service rendu à la population (des services plus ciblés et plus rapides voire la création de nouveaux services). Elle favorise le développement de la recherche. Elle contribue au développement économique. S'ajoutent à cela des objectifs démocratiques et politiques. La publication des données est d'abord un exercice de transparence. C'est en tout cas ce qu'avaient imaginés les premiers promoteurs de l'*open data* en Californie en 2007 et ce qu'a repris le législateur français en 2016 dans la loi pour une République numérique⁵. Si l'*open data* a souvent été le déclencheur d'une prise de conscience de l'importance des données dans les collectivités territoriales, beaucoup d'acteurs ont rapidement compris que l'*open data* seul ne suffit pas à atteindre leurs objectifs de transformation et de pilotage des politiques publiques. Certes la publication des données reste un objectif en soi. Et un gage de confiance démocratique essentiel, nous y reviendrons. Mais l'impact des données sur l'action publique et l'intérêt général n'est significatif que si l'on opère des croisement de données. Il peut s'agir de croiser des données publiques entre elles, ouvertes ou non. Il peut s'agir de croiser des données publiques et des données privées, accessibles sous certaines conditions. Il peut aussi s'agir de croiser des données collectives et anonymisées avec les données personnelles d'habitants et d'usagers du service public, sous la réserve expresse d'une gestion rigoureuse de leur consentement.

5. Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique.

bretagne et pays de la Loire

L'association Résovilles expérimente en Bretagne et Pays de Loire de nouveaux usages de la donnée au service des acteurs de la politique de la ville à travers le programme « Data & quartiers ». Elle a conduit entre 2019 et 2021 plusieurs projets concernant l'emploi, la santé et la mobilité. Pour chaque projet, l'enjeu central a été d'organiser l'accès à des données et de permettre leur partage entre acteurs de statuts différents mais tous impliqués dans la Politique de la Ville.

Ainsi par exemple durant la crise sanitaire, Résovilles a réalisé une cartographie de la fragilité des quartiers face à la Covid-19⁶ en croisant des données publiques disponibles en *open data*, avec des données de santé sensibles expliquant certains risques et certains facteurs de comorbidité. Ce projet a mobilisé l'Agence nationale de la cohésion des territoires, des Agences régionales de santé, des métropoles et une mutuelle privée (Harmonie Mutuelle). Il avait pour objectif de porter à la connaissance des acteurs de terrain des réalités utiles à la conduite de politiques de sensibilisation et de prévention. Le croisement de données jusqu'alors inaccessibles a produit de la connaissance inédite.



6. Jordan Esnault, Émilie Sarrazin, *Quand le big data se met au service des quartiers*, Horizons publics, 2020.

montréal

À Montréal, l'organisme à but non lucratif Synapse C⁷ propose aux acteurs de la culture de mutualiser, en confiance, leurs données. L'objectif est de créer une vision partagée du développement culturel en faisant émerger une compréhension des habitudes et des attentes des spectateurs. Les travaux de Synapse C contribuent depuis plusieurs années à l'émergence de pratiques culturelles nouvelles, à l'évolution de l'offre culturelle et à sa diffusion. Elle sert aussi à coordonner l'action d'acteurs culturels, publics et privés, soucieux de développer collectivement des pratiques nouvelles et d'éviter des processus de cannibalisation entre offres potentiellement concurrentes. La mutualisation des données se fait sur la base de règles définies en commun avec un haut niveau d'exigence en termes d'anonymisation et de garanties de non-concurrence. Le rôle de Synapse C est d'autant plus essentiel que les organisations culturelles ne disposent souvent pas de l'expertise ou des ressources nécessaires pour gérer et valoriser leurs propres données.

région centre val-de-loire

Le Climate data hub est une initiative de la Région Centre Val-de-Loire pour permettre le partage en confiance de données publiques ou privées utiles au pilotage de la lutte contre le réchauffement climatique. Ce projet s'inscrit dans le cadre d'une gouvernance partagée, la Région ayant installé une « COP régionale » sur le modèle des COP mondiales.

Le Climate data hub ambitionne de rassembler des données qui auront un impact sur les politiques publiques (locales, régionales ou portées par l'État en Centre Val-de-Loire) mais aussi sur les activités d'acteurs privés, qu'il s'agisse des citoyens, d'associations ou d'entreprises du territoire. De multiples domaines sont concernés : énergie, eau, déchets, mobilité, consommation, agriculture... La création du cadre de confiance entre les acteurs mobilisés est un processus long. Il a été engagé sur plusieurs mois avec des travaux dans diverses commissions pour écrire un cadre éthique et encadrer juridiquement les échanges de données, identifier des « cas d'usage », choisir une infrastructure souveraine, imaginer un modèle économique.



7. synapsec.ca

DE QUELLES DONNÉES PARLE-T-ON ?

Dans le cadre de leurs missions, les administrations territoriales produisent de nombreuses données. Ce sont des données publiques. Les collectivités en ont la maîtrise et en sont responsables. En application de la loi pour une République numérique, dès lors que ces données présentent un intérêt collectif (social, économique, environnemental, démocratique...) elles sont soumises « par défaut » à une obligation de publication. C'est l'*open data*.

La loi française permet aussi d'affirmer que les données produites par des entreprises qui exercent une mission de service public (notamment dans le cadre d'une délégation de service public) ont un statut de données publiques⁸. Ce point est important car il garantit l'accès aux données pour la collectivité, ce qui est indispensable pour exercer un contrôle des entreprises qui agissent pour son compte, *a fortiori* lorsque les entreprises optimisent leur gestion par des dispositifs de plus en plus sophistiqués pilotés par la donnée. Mais ce point est aussi important car il permet la publication de certaines de ces données en *open data*. Ces dispositions sont souvent, volontairement ou non, mal comprises par les entreprises délégataires bien que ces données soient essentielles au pilotage des politiques publiques.

Les données publiques constituent dans leur globalité un patrimoine et un capital dont la collectivité est garante, pour ses propres usages et plus globalement au service de l'intérêt général.

Mais d'autres données concourent à l'intérêt général. Elles peuvent être aussi publiques, mais produites à d'autres échelles et par d'autres acteurs : l'INSEE, l'État, d'autres collectivités, des entreprises publiques...

Elles peuvent aussi être privées. De nombreuses données sont produites par des acteurs privés qui sont présents sur un territoire et dont l'action a un impact sur les politiques publiques. Le cas des données produites par Waze ou Airbnb est bien connu. Ces acteurs interfèrent avec les politiques de mobilité ou l'accueil touristique et l'immobilier. Et les collectivités peuvent légitimement

8. Schéhérazade Abboub, *Le Cadre juridique de la donnée en cinq points*, La Gazette des communes, janvier 2022

vouloir accéder à ces « données privées d'intérêt général ». Cette notion apparaît pour la première fois en 2017 dans le rapport « De la smart city aux territoires d'intelligence(s)⁹ » du député Luc Belot. Il recommande la création d'un statut particulier pour ces données de sorte à créer les conditions de leur mise à disposition. D'autres initiatives ont suivi. En 2020, le Conseil national du numérique propose que toutes les données utiles au pilotage des transitions écologiques soient considérées d'intérêt général¹⁰. Et dans l'attente d'un cadre européen imaginé comme facilitateur à terme, le rapport du député Éric Bothorel¹¹ préconise une approche pragmatique permettant de négocier au cas par cas la transmission de données privées au motif d'un intérêt général avéré. Une circulaire a d'ailleurs annoncé la création future d'une fonction de médiateur de la donnée d'intérêt général¹².

La question de l'accès aux données privées jugées d'intérêt général, de leur mise en partage, de leur échange ou de leur croisement avec des données publiques, est au cœur des enjeux d'un cadre de confiance. Les acteurs publics redoutent que leurs données, bien que publiées en *open data*, ne servent à des usages contraires à l'intérêt général. Des acteurs privés, bien que désireux de concourir à l'intérêt général, redoutent que leurs données ne dévoilent et fragilisent leur modèle d'affaires ou portent préjudice à leur développement.

9. Luc Belot, *De la smart city aux territoires d'intelligence(s)*, rapport au Gouvernement, 2017.

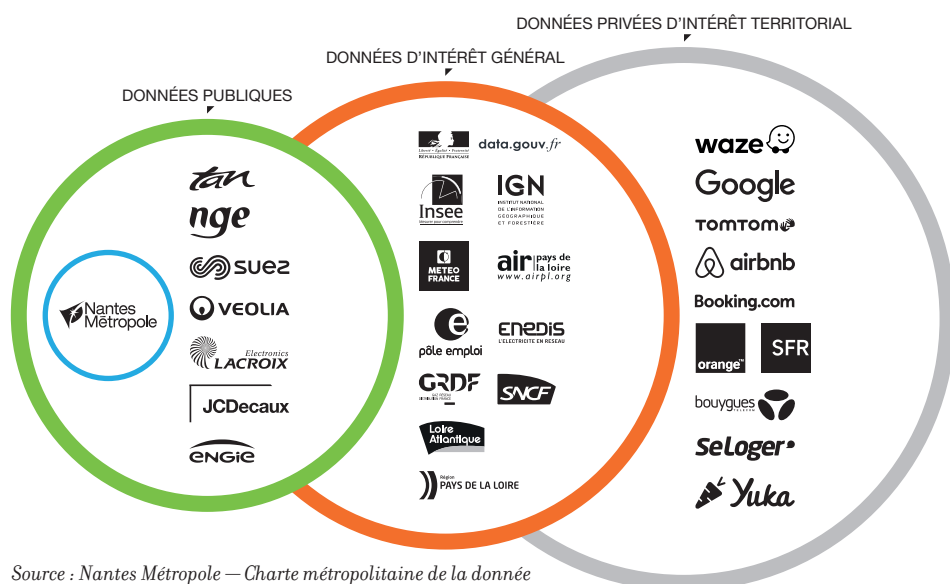
10. Avis du conseil national du numérique, *Faire des données environnementales des données d'intérêt général*, juillet 2020.

11. Éric Bothorel, *Pour une politique publique de la donnée*, rapport au Gouvernement, 2020.

12. Circulaire du 27 avril 2021, Politique publique de la donnée, des algorithmes et des codes sources.

nantes métropole

Nantes Métropole n'a pas attendu l'avènement d'un cadre juridique national ou européen stabilisé pour identifier les données privées qui présentent un intérêt pour les politiques métropolitaines. La charte territoriale de la donnée promulguée en 2019¹³ précise que « des acteurs divers interviennent dans la vie du territoire métropolitain et sont susceptibles de produire des données qui revêtent un caractère d'intérêt général. Certaines sont produites par des acteurs publics (services de l'État, collectivités territoriales, entreprises publiques ou Concessionnaires de l'État...). D'autres sont produites par des acteurs privés. Lorsqu'il est de l'intérêt de tous qu'elles soient partagées avec la puissance publique parce qu'elles peuvent contribuer à la mise en œuvre des politiques publiques du territoire, la collectivité propose un cadre de dialogue avec les acteurs concernés pour créer les conditions d'un accès à ces données respectueux des droits de tous. Ces données sont d'intérêt métropolitain. »



Source : Nantes Métropole — Charte métropolitaine de la donnée

¹³. <https://metropole.nantes.fr/charte-donnee>

DE L'OPEN DATA AUX CLOSED DATA

EN France, depuis octobre 2018, les données publiques sont soumises « par défaut » à une obligation de publication. Trois types de données sont néanmoins dispensés de publicité : les données personnelles protégées par le RGPD, les données couvertes par des droits d'auteur et les données couvertes par des règles de protection économique (secret des affaires, propriété intellectuelle...).

L'obligation s'impose aux collectivités de plus de 3 500 habitants et ayant au moins 50 agents en équivalent temps plein. Près de 5 000 collectivités sont ainsi concernées. Le bilan publié par l'association Open Data France¹⁴ en octobre 2021 montre que quasiment toutes les régions et plus de la moitié des départements et des communes de plus de 100 000 habitants ont engagé des démarches de publication. En revanche, moins de 10% des communes et des EPCI de moins de 100 000 habitants respectent la loi. Le choix des données publiées varie considérablement d'un territoire à l'autre, et la qualité des données publiées est très inégale. Certaines données sont même jugées inutilisables en raison de leur format ou de leur caractère trop local et/ou trop spécialisé¹⁵.

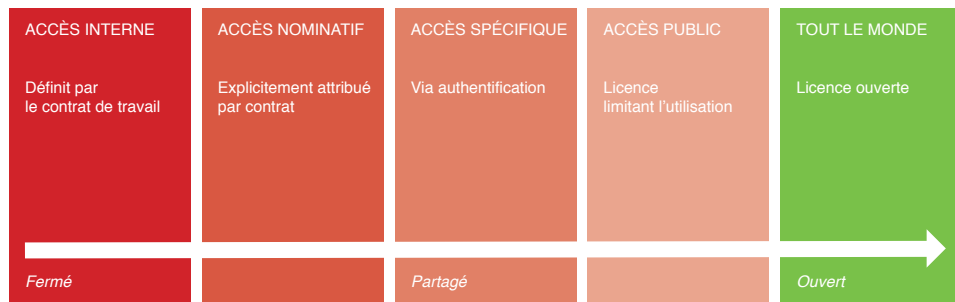
Mais cette situation n'empêche pas les territoires concernés d'utiliser des données au service de leurs politiques publiques, de les partager ou de les croiser. D'un strict point de vue juridique et démocratique, au regard notamment des principes de transparence voulus par le législateur, cette situation peut choquer. Mais dans bien des circonstances les acteurs territoriaux considèrent, à tort ou à raison, que des jeux de données utiles ne sont pas publiables. Ils peuvent comporter des données personnelles ou permettre des réidentifications. Ils peuvent présenter un risque de violation du secret des affaires, du secret fiscal ou du secret statistique. Le rapport Bothorel souligne le risque que ces données demeurent alors inexploitées bien qu'utiles à l'intérêt général et propose que l'on puisse « prévoir un cadre de partage des données intermédiaire entre l'ouverture au public et la fermeture des données¹⁶ ».

14. observatoire-opensdata.fr/resultats

15. FING, *Dix ans d'ouverture des données publiques, un bilan critique*, 2020.

16. Éric Bothorel, *Pour une politique publique de la donnée*, rapport au Gouvernement, 2020.

Pour comprendre ces processus, l'Open Data Institute propose une grille de lecture décrivant un continuum¹⁷ allant des données complètement accessibles aux données fermées.



Le spectre d'ouverture des données (ODI, traduit par Dataactivist)

17. theodi.org/data-spectrum

ACTEURS PUBLICS ET PRIVÉS : DE NOUVELLES COOPÉRATIONS À INVENTER

L'accès aux données, leur mise en commun pour procéder à des analyses puis leur exploitation à des fins de modélisation introduisent une nouvelle zone de dialogue et parfois de tensions entre acteurs publics et acteurs privés. Ces nouveaux procédés impactent d'abord les rôles traditionnels de client et de fournisseur, de délégataire et d'autorité concédante, d'entreprise et d'autorité de contrôle. Ils bouleversent aussi les règles de cohabitation d'entreprises concurrentes sur un même territoire, par exemple lorsqu'une commune impose à différents délégataires de déverser leurs données d'exploitation (éclairage public, gestion du stationnement public, distribution de l'eau...) dans une seule plateforme de données urbaines.

À chaque fois des questions se posent. Elles concernent l'intérêt légitime de la collectivité à accéder aux données, les conditions économiques, techniques ou juridiques de leur transfert, les limites éventuelles à leur exploitation, les limites certaines à leur publicité. En retour, les entreprises qui opèrent des missions de service public sont de plus en plus utilisatrices de données publiques et de données d'intérêt général qui permettent de mieux décrire et comprendre le territoire, et ce faisant de modéliser et d'optimiser leurs offres. L'encadrement juridique de la circulation de ces données constitue une préoccupation nouvelle pour les administrations locales et les entreprises qui les accompagnent. Plusieurs publications proposent un cadre applicable à la commande publique¹⁸ ou détaillent les enjeux juridiques qui naissent du développement d'usages de plus en plus massifs de la donnée au sein des villes intelligentes¹⁹.

Mais l'organisation de l'accès à ces données et plus largement la mise en commun de données publiques et de données privées ne reposent pas sur des considérants uniquement juridiques. La création d'un cadre de confiance qui garantisse et préserve en premier lieu l'intérêt général est une préoccupation légitime. Elle émane des acteurs publics, de certaines entreprises, mais aussi des citoyens.

18. Voir par exemple le *Guide des bonnes pratiques contractuelles et recommandations pour la mise en place d'une gouvernance territoriale de la donnée*, Banque des territoires, 2020.

19. *De la smart city à la réalité des territoires connectés*, rapport officiel, ministère de l'Économie et des finances, octobre 2021.

L'intérêt de mutualiser des données publiques et privées ne relève pas de l'évidence. L'association Occitanie Data a piloté en septembre 2019 une importante étude sur la perception des enjeux de la donnée par les Français mais aussi sur leurs attentes en matière de régulation et de contrôle des usages des données par les acteurs publics et privés²⁰. À la question « êtes-vous favorables à ce que des acteurs publics et privés partagent leurs données pour améliorer leurs services ? », 48 % des sondés expriment un avis assez ou tout à fait opposé. L'étude montre que l'acceptation de cet échange repose sur la formulation d'objectifs et de finalités concrètes (au bénéfice de l'intérêt général) et sur la mise en œuvre de principes qui, additionnés les uns aux autres, concourent à la création d'un cadre de confiance : transparence, garantie d'anonymisation des données, existence d'un cadre éthique, contrôles externes indépendants notamment.



²⁰. Enquête conduite par l'institut Harris Interactive, CIVITEO et Cap comportant plusieurs panels d'usagers ainsi qu'un sondage national « Les Français et le big data, le défi de la confiance » réalisé sur un échantillon de 1027 personnes en septembre 2019.

occitanie data

Occitanie Data est un tiers de confiance créé en 2019 sous forme associative regroupant des acteurs publics (Région Occitanie, Universités de Toulouse et Montpellier, Toulouse métropole, le CNRS, l'IGN, la Banque des territoires...) et des acteurs privés (Airbus, Orange, Sopra Steria...) qui coopèrent pour favoriser l'émergence d'une économie de la donnée en Occitanie. L'association a trois objectifs : développer une économie responsable de la donnée dans l'intérêt général et offrant des services innovants, définir un cadre de confiance éthique et souverain autour de la science des données et de l'IA, accompagner la transformation numérique de manière compatible avec la transition écologique.

Les membres d'Occitanie Data ont produit une charte éthique²¹ de la donnée et construit un label éthique qui garantit la conformité aux principes de la charte après contrôle par un certificateur indépendant²².

Occitanie Data propose à ses membres publics et privés un véritable accompagnement pour la mise en œuvre de leurs projets de partage et d'utilisation des données (aide technologique, éthique et juridique). Pour asseoir son développement et élargir son périmètre géographique d'intervention un processus de transformation en Groupement d'intérêt public (GIP) est engagé. En 2022, Occitanie Data se transformera en GIP EkitIA, premier tiers de confiance construit en France au service des politiques publiques locales et du développement des territoires.



21. <https://www.occitaniedata.fr/la-charte-ethique/>

22. <https://www.occitaniedata.fr/label-ekitia/>

DONNÉES PERSONNELLES ET INTÉRÊT GÉNÉRAL : AMBIVALENCE ET PARADOXES

LA question de l'utilisation des données personnelles des habitants et des usagers du service public est également au cœur des enjeux de la construction d'un cadre de confiance. En France et en Europe, des limites protectrices sont posées par le RGPD. Rappelons d'abord que le renforcement des règles de gestion du consentement au traitement des données personnelles n'obère pas la capacité des services publics à agir au quotidien. L'article 6 du règlement instaure comme base légale à la collecte des données un « intérêt légitime » sur lequel reposent nombre de traitements administratifs. La question qui est posée ensuite est celle de l'utilisation des données personnelles des citoyens pour d'autres usages ou des usages complémentaires à ceux prévus lors de la collecte initiale. L'utilisation de ces données peut présenter de multiples intérêts : personnalisation et adaptation des services aux besoins de chacun (à travers des applications notamment), actions ciblées de prévention (en matière de santé ou d'action sociale...), simplification des procédures (déclarations, paiements...), conseils et recommandations (à travers des applications pour la mobilité, la gestion des déchets...), modélisation et adaptation collective d'offres de service, gestion optimisée des ressources publiques, travaux de recherche, etc. Mais pour tous ces usages, un consentement complémentaire doit être formalisé.

La question de la confiance des citoyens dans l'utilisation faite de leurs données est essentielle. En France, la confiance à l'égard du service public est réelle. Le sondage réalisé pour Occitanie Data est à cet égard éclairant²³. À la question « à quels acteurs faites-vous confiance, concernant la gestion et l'utilisation des données ? », 62 % des répondants accordent naturellement leur confiance aux collectivités territoriales. *A contrario* ils ne sont que 42 % à faire confiance aux entreprises de vente en ligne, 38 % aux moteurs de recherche, 22 % aux réseaux sociaux... Mais au-delà de cette représentation sans doute sans surprise, il y a un paradoxe. Les Français confient toujours plus de données aux géants du numérique mais sont réticents à confier plus de données aux acteurs publics, fusse au bénéfice de services améliorés, fusse au service de l'intérêt général. Une autre étude réalisée en Ile-de-France sur les nouveaux services et la « ville de demain »

²³. *Les Français et le big data, le défi de la confiance*, Harris Interactive, CIVITEO, Cap pour Occitanie Data, septembre 2019.

montre que le sujet est clivant : 52% des personnes interrogées se déclarent réticentes ou opposées à partager plus de données personnelles pour améliorer la qualité des services rendus, contre 47% qui l'accepteraient volontiers²⁴.

Le chemin de la confiance, dès lors qu'il sera aussi fait usage de données personnelles des habitants n'est donc pas simple. On retiendra donc avec intérêt le fait que le secteur public français part avec un crédit de confiance positif, ce qui est une caractéristique hexagonale. Elle n'est pas nécessairement partagée ailleurs dans le monde. Ainsi par exemple 80% des Américains se déclarent préoccupés ou très préoccupés des atteintes à leur vie privée par leur propres administrations gouvernementales²⁵.



²⁴. *Les Franciliens et la ville de demain*, Harris Interactive pour Grand Paris aménagement, décembre 2018.

²⁵. Baromètre annuel sur les usages numériques et la confiance réalisé par IPSOS pour le CIGI (ONU) dans 25 pays (<https://www.cigionline.org/cigi-ipsos-global-survey-internet-security-and-trust/>)

toronto

Le projet de Sidewalk Toronto est emblématique de cet enjeu de la confiance des citoyens dans l'usage de leurs données au cœur de l'action publique locale²⁶. Porté par Sidewalk Labs, filiale sœur de Google au sein du groupe Alphabet, le projet de ville intelligente de Toronto au Canada a connu de nombreuses difficultés et a finalement été abandonné (officiellement au motif de la crise sanitaire). La gestion des données des habitants du futur quartier a été l'objet de vives polémiques. Le choix de collecter massivement ces données sans aucun procédé d'anonymisation à la source a entraîné une forte suspicion à l'égard de Google sur les usages qui en seraient faits *in fine*. L'annonce d'un hébergement des données aux États-Unis, à la législation moins protectrice que le Canada, a amplifié la méfiance. Face à l'émoi grandissant d'une partie de la population mais aussi des autorités locales ou nationales, plusieurs démarches ont été entreprises. Elles incluaient la création d'un *urban data trust* indépendant en charge de la gestion ultérieure des données ainsi qu'une démarche d'information des citoyens sur la présence de capteurs dans l'espace public, le type de données collectées et leurs usages possibles. Mais ces efforts, tardifs, n'ont jamais réussi à contrebalancer le fait que les citoyens n'avaient pas été associés dès l'origine à la conception d'un cadre général de gouvernance de la donnée. Le sentiment prédominant était que Google décidait seul, de manière illégitime et non démocratique, ce qui convenait au collectif, privant la puissance publique de ses prérogatives et plaçant les citoyens devant le fait accompli.

On notera avec intérêt que les outils de transparence sur la gestion des données dans l'espace public inventés sur place à Toronto ont été malgré tout très innovants. Ils sont disponibles sous licence Creative Commons. Leur déploiement est aujourd'hui expérimenté dans différentes villes par la société américaine Helpful Places²⁷.

²⁶. Jacques Priol, *Ne laissez pas Google gérer nos villes*, Éditions de l'Aube, 2020.

²⁷. La présentation du standard de communication et le programme expérimental déployé notamment à Boston sont consultables sur le site <https://dtp.helpplaces.com/>

stop covid

La gestion des données personnelles des habitants est devenue un sujet d'actualité et une préoccupation majeure durant les années 2020 et 2021 en raison de la pandémie du Covid-19. Le rapport du député Éric Bothorel souligne que « le manque d'appropriation de l'application StopCovid par les citoyens [en 2020] a mis en évidence la réticence de la population française à partager des données personnelles avec la puissance publique. Dans ce contexte, il apparaît nécessaire de créer une confiance minimale dans les services de l'État en associant davantage la société civile pour établir les prérequis indispensables permettant d'engager ce type de démarche ».

En mai 2020, l'Observatoire Data Publica avait fait réaliser un sondage sur la perception par les Français de l'application STOPCOVID et sur l'usage des données²⁸. Si une majorité de Français était favorable au déploiement de l'application pour combattre la pandémie (59% d'avis favorables), ils étaient 54% à s'inquiéter et à indiquer qu'ils n'avaient pas confiance dans le Gouvernement pour garantir les usages qui seraient faits des données collectées.

²⁸. Sondage portant sur les perceptions de l'application STOPCOVID et regards sur l'enjeu du partage des données personnelles, Harris Interactive pour l'Observatoire Data Publica en mai 2020 – 1051 répondants.

Face à ces enjeux spécifiques, plusieurs initiatives ont été prises pour permettre aux citoyens de conserver la maîtrise de leurs données et de gérer de façon autonome des consentements nouveaux (pour des données déjà collectées). C'est le sens du mouvement My Data²⁹ et des initiatives prises en France pour développer le *self data* territorial. En un mot, l'idée est d'outiller des citoyens volontaires pour qu'ils puissent accorder, ou retirer, leur consentement à l'usage de leurs données pour un projet d'intérêt général : la recherche médicale, des prototypes de réduction de la consommation énergétique, l'optimisation de la gestion des déchets, le changement d'habitudes alimentaires, etc. Ces dispositifs participent de la construction expérimentale d'un cadre de confiance pour la gestion des données locales. Nous renvoyons le lecteur à deux documents : le Kit du *self data* territorial produit par la FING à partir des expériences de Nantes, Lyon et La Rochelle³⁰ et l'étude « Données personnelles : le modèle d'après » publié par la Banque des Territoires³¹.

Ces dispositifs restent, à ce jour, limités dans leur déploiement. Mais ils permettent d'éprouver avec des citoyens volontaires le principe de portabilité des données personnelles prévu par le RGPD et ils anticipent sans doute le futur « altruisme des données » du *Data Governance Act*³².



29. <https://mydata.org/>

30. *Kit du Self data territorial*, FING, 2019.

31. *Données personnelles : le modèle d'après. État des lieux de l'écosystème du self data*, Banque des Territoires, décembre 2020.

32. Le *Data Governance Act* est un projet de règlement européen proposé fin 2020 par la Commission européenne, en cours de discussion après une concertation publique, dont l'adoption par le Parlement européen est attendue fin 2022.

métropole européenne de Lille

Tous les ans au mois de mai, les habitants de la Métropole européenne de Lille sont invités à prendre part au « Challenge métropolitain du vélo ». L'opération a pour but de les inciter à repenser leurs modes de déplacement quotidiens et à opter pour le vélo pour aller au travail, à l'école ou pour leurs différentes activités. Ils sont ainsi chaque année près de 5 000 à relever le défi.

Plus de la moitié d'entre eux ont également accepté de communiquer à la Métropole le détail de leurs trajets. Ils le font grâce à une application mobile qui enregistre les tracés via la puce GPS de leur téléphone mobile. Certains préfèrent en revanche renseigner manuellement leurs parcours sur Internet. Les données ainsi collectées s'ajoutent à celles de capteurs qui opèrent des comptages sur les voies cyclables. Consolidées, elles permettent de mettre à jour chaque année une carte de chaleur de la pratique cycliste dans la Métropole. La carte sert à comprendre les évolutions des trajets, à mieux connaître la fréquentation des aménagements réalisés, à identifier des axes dangereux.

La mise à disposition volontaire de données personnelles contribue à la priorisation de futurs travaux.



CONSTRUIRE UN CADRE DE CONFIANCE AU SERVICE DE L'INTÉRÊT GÉNÉRAL

LA construction d'un cadre de confiance est un processus qui prend du temps et qui repose sur la combinaison de plusieurs éléments. Il faut des règles, des principes, des valeurs et des objectifs partagés. Pour ce qui nous concerne ici, ces objectifs revêtent un caractère d'intérêt général. Il faut aussi des outils et des méthodes de gestion cohérents avec les finalités du projet, pilotés et gouvernés selon des modes de collaboration admis de tous. L'utilisation concrète des données et les « cas d'usage » sont aussi constitutifs du cadre de confiance en ce qu'ils donnent à voir la réalité des modes opératoires et des actions.

Parfois, mais pas toujours, la gestion « en confiance » des données passe par l'émergence d'un tiers c'est-à-dire d'un opérateur neutre, garant du respect des règles fixées par les parties prenantes. Nous y reviendrons dans le détail dans la dernière partie de ce cahier.



DES RÈGLES ET DES OBJECTIFS

LA définition des règles d'utilisation des données est un préalable indispensable à l'organisation du partage, des échanges et du croisement de données de statuts divers. Les craintes et les résistances sont en effet nombreuses. Elles sont souvent légitimes.

DES CRAINTES MULTIPLES ET LÉGITIMES

Les acteurs publics hésitent parfois à mettre entièrement leurs données à disposition, y compris lorsqu'il s'agit de données soumises à l'obligation d'*open data*. La crainte existe d'usages malveillants des données publiques, du fait d'opposants ou du fait d'observateurs peu scrupuleux. À l'usage cette crainte se dissipe souvent car il y a plus à perdre à refuser la transparence qu'à en gérer les conséquences. En revanche, l'exploitation des données publiques par des acteurs privés peut parfois gêner la mise en œuvre de politiques publiques ou concurrencer une offre de service voulue au bénéfice de tous et fragilisée par des usages rentables au service de quelques-uns. Le partage des données de mobilité en est un bon exemple. La SNCF a pendant longtemps refusé d'ouvrir les données de circulation des trains (malgré une obligation légale³³!) au motif que ces données seraient exploitées d'abord par BlaBlaCar principal opérateur de co-voiturage. BlaBlaCar ayant intégré le Groupe SNCF, c'est la concurrence potentielle de Google qui bloquait. Le même raisonnement a conduit plusieurs territoires, Toulouse Métropole et le Grand Lyon par exemple, à freiner ou à tenter de réguler les usages possibles des données rendues publiques. Le Grand Lyon a notamment expérimenté une licence restrictive dont l'objectif est de garantir que des données publiques ne seront pas exploitées à des fins contraires à l'intérêt général.

33. Loi n°2015-990 du 6 août 2015 pour la croissance, l'activité et l'égalité des chances économiques, dite « loi Macron »

grand Lyon

La « licence de réutilisation des données d'intérêt général » a été imaginée par le Grand Lyon pour contraindre les utilisateurs de certaines données mises à disposition gratuitement à déclarer en retour les utilisations qu'ils envisagent. La Métropole peut ainsi s'assurer que ces utilisations seront conformes aux politiques publiques et qu'elles ne contreviennent pas à l'intérêt général. Appliquée aux données de mobilité du réseau public, cette licence peut interdire le développement de services réorientant le trafic automobile sur des voiries à préserver (à proximité d'écoles ou d'hôpitaux); interdire la réutilisation de données du trafic routier en temps réel pour alimenter des applications encourageant exclusivement l'usage de l'automobile; ou encore éviter que ces données n'alimentent des services promouvant des pratiques non compatibles avec l'intérêt général, comme par exemple le signalement du positionnement des contrôleurs dans le réseau des transports en commun, incitant à des comportements inciviques.

On notera que cette licence lyonnaise n'a pas été homologuée au niveau national et reste donc à ce jour une licence expérimentale.



Les acteurs privés ont également des craintes et des réserves à faire valoir avant de partager leurs données, y compris au service de l'intérêt général. Il y a d'abord le souci de préserver la confidentialité d'informations qui peuvent intéresser la concurrence. Certaines informations peuvent sembler d'apparence anodine tandis que d'autres relèvent expressément du secret des affaires ou sont protégées par des clauses de propriété intellectuelle. Au cœur de la gestion publique, ces notions ne sont pas de manipulation facile. D'abord parce que le secret des affaires n'est pas opposable par les entreprises aux autorités concédantes. La loi pour une République numérique a précisé ce point rendant obligatoire et automatique la transmission des données liées à l'exploitation du service public³⁴. Mais la collectivité n'est pour autant pas autorisée à tout publier ou à tout partager. Détentrice des données, elle doit préserver l'entreprise de toute violation préjudiciable. Et ceci sera plus vrai encore si des données sont transmises en dehors de tout lien contractuel de droit public, par des entreprises volontaires (ou « altruistes ») au motif de l'intérêt général.

Quant aux citoyens, et parfois aux associations qui les représentent, ils expriment à la fois des craintes et font montre d'ambivalence. D'un côté ils acceptent que leurs données personnelles puissent parfois être exploitées à des fins d'utilité publique. L'étude réalisée pour Occitanie Data en 2019³⁵ montrait par exemple une acceptation forte d'un usage des données, y compris non anonymisées, au bénéfice de la recherche médicale (62%) ou pour l'amélioration de l'efficacité du service public (52%), et a contrario un refus net que s'accroisse l'exploitation des données pour optimiser des services marchands (refus à 65%). Mais les citoyens redoutent aussi que leurs données personnelles ne soient utilisées à des fins détournées par les acteurs publics. Ils craignent notamment des contrôles (de police, fiscaux...) ou des manipulations, notamment en période électorale. L'un des cas étudiés lors de l'amorçage du Climate data hub de la Région Centre Val-de-Loire est à cet égard intéressant. Les représentants des irrigants, intéressés à partager les données d'exploitations agricoles pour construire des modélisations utiles à l'élaboration de pratiques plus durables, se sont inquiétés de voir dans le tour de table des services de l'État par ailleurs en charge de la Police de l'eau. La question était simple :

34. Voir sur ce point Schéhérazade Abboub, *Le Cadre juridique de la donnée en cinq points*, La Gazette des communes, janvier 2022.

35. *Les Français et le big data, le défi de la confiance*, Harris Interactive, CIVITEO, Cap pour Occitanie Data, septembre 2019.

comment garantir que les données fournies au nom de travaux d'intérêt général ne seront pas utilisées à d'autres fins, notamment de contrôle ?

Chacun attend donc des garanties. La construction d'un cadre de confiance repose d'abord sur le partage des objectifs, l'acceptation des règles existantes et la construction, si besoin, de règles additionnelles.

LA DÉFINITION D'OBJECTIFS PARTAGÉS

La définition des objectifs n'est pas toujours un exercice simple, mais cette étape est essentielle : elle fixe la raison d'être du cadre de partage des données.

Il peut s'agir de finalités générales qui permettent à des acteurs d'engager la discussion. Ils partagent la conviction que leurs données mises en commun vont contribuer à l'intérêt général sans en avoir la certitude ou la preuve concrète. Synapse C à Montréal est né de l'idée que des institutions culturelles concurrentes, ou se considérant comme telles, avaient à gagner à disposer d'une connaissance meilleure de leur public commun. Occitanie Data est née de la volonté d'acteurs publics et privés de faire émerger une économie de la donnée souveraine et éthique (en opposition aux méthodes de quelques plateformes géantes). Le Climate data hub a engagé ses travaux parce que des acteurs publics et privés ont fait le constat, ensemble, du besoin d'outils de pilotage des transitions environnementales à l'échelle d'une région. À des échelles plus locales, des cadres nouveaux de mutualisation de données voient le jour autour d'enjeux multiples : fluidifier les déplacements et réduire les pollutions, réduire les consommations énergétiques, changer les habitudes alimentaires, réduire la production et améliorer la gestion des déchets, renforcer la sécurité...

La déclinaison de ces grandes finalités en objectifs précis permettra l'implication d'acteurs nouveaux. À la condition qu'ils soient associés à la définition de ces objectifs. Ceci est particulièrement vrai si l'on veut obtenir le consentement des citoyens à l'exploitation de données personnelles (y compris dans des processus anonymisés). L'expérience lilloise concernant les trajets des vélos dans la ville, ou les expériences conduites ailleurs avec des applications comme Géovélo³⁶, illustrent bien cet enjeu. Peu

36. <https://geovelo.fr/>

de cyclistes pourraient consentir à être « tracés » en continu par une balise GPS. Ils sont pourtant nombreux à souhaiter enregistrer leurs déplacements durant des périodes données pour cartographier des itinéraires et peser sur les acteurs publics en charge de l'aménagement des pistes cyclables.

UN CADRE LÉGAL À RESPECTER

Cela va de soit... mais ça va mieux en le disant. Un cadre de confiance pour la gestion des données au service de l'intérêt général est un cadre dans lequel la loi est appliquée. Et le premier point mis en avant par les parties prenantes est logiquement le respect du RGPD. Vient ensuite la loi pour une République numérique qui définit des exigences en matière de transparence (*open data*, transparence des algorithmes publics). Suivent également toutes les dispositions qui protègent les droits d'auteur, le secret médical, le secret fiscal, le secret des affaires, la propriété intellectuelle...

Échanger et partager des données en confiance impose de vérifier que tous agissent en conformité à cet environnement législatif. Des conventions d'échange de données peuvent en rappeler les termes. Elles peuvent être bilatérales ou multipartites. Le programme « Data & Quartiers » évoqué précédemment repose entièrement sur des conventions bilatérales qui concernent des acteurs multiples : une association de collectivités et des collectivités, un ministère, Pôle Emploi et un groupe de travail temporaire (Randstad), deux Agences régionales de santé, l'Assurance maladie, une Mutuelle (Harmonie Mutuelle), une Université... Au service d'une finalité partagée (améliorer la connaissance des quartiers prioritaires) chaque partenaire rassemble ou produit et met à disposition des données utiles sous la réserve expresse de la légalité de la transmission et de la légalité des traitements. Mais il est parfois utile de renforcer ce cadre par des règles additionnelles, convenues entre les participants. C'est notamment le rôle des chartes éthiques.

LES CHARTES

L'idée de chartes pour encadrer et favoriser l'échange de données au service de l'intérêt général avait été avancée en 2017 dans le rapport parlementaire sur les villes intelligentes du député Luc Belot³⁷. La charte est un objet juridique « mou » dont on peut parfois craindre qu'il reste une lettre d'intention. À bien y regarder pourtant, les travaux engagés par quelques territoires pionniers montrent que l'élaboration d'une charte est souvent un moment clef du processus de création d'un cadre de confiance. La charte rassemble des principes additionnels dont les partenaires conviennent. Chacun interroge ses propres pratiques au regard d'une certaine éthique de la donnée au service de l'intérêt général. Et chacun projette les principes convenus sur l'idée qu'il peut se faire des pratiques de ses partenaires. Dans plusieurs territoires, des chartes (ou des règlements de partenariat) régissent notamment la manière dont des entreprises partenaires peuvent, ou non, exploiter des données publiques à des fins commerciales. La primauté des objectifs d'intérêt général apparaît ainsi dans la Charte métropolitaine de Nantes³⁸, dans les travaux du Service public de la donnée de Brest Métropole (voir ci-après), dans la charte de DataGrandEst³⁹ et bien d'autres.

Certes, une charte peut demeurer en l'état d'engagements de principe. Elle joue alors le rôle d'un guide de bonnes pratiques. Mais une charte peut aussi servir de texte de référence au sein de contrats et de conventions. Visée de façon simple ou annexée, la charte devient opposable. Et dans certains cas, les partenaires s'accordent sur des dispositifs de contrôle et de sanction en cas de non-respect des engagements pris⁴⁰. Dès 2014, le parlement écossais avait anticipé la possibilité de sanctionner financièrement des entreprises qui feraient un usage détourné de données mises à leur disposition⁴¹. Les villes de Nantes et de Montréal coopèrent pour construire un dispositif d'évaluation de l'impact de leurs chartes de la donnée. Occitanie Data a lancé un label délivré par un auditeur externe.

37. Luc Belot, *De la smart city aux territoires d'intelligence(s)*, rapport au Gouvernement, 2017.

38. <https://metropole.nantes.fr/charte-donnee>

39. <https://www.datagrandest.fr/portail/fr/partenariat>

40. Voir à ce sujet *Big data dans les territoires, comment garder la main ?* Horizons publics n°18, 2020.

41. *A Data Vision for Scotland*, Data management board, The Scottish Government, 2014.

Le label ekitia

Occitanie Data a créé le label EKITIA pour inciter au respect des principes de sa charte éthique des données et pour valoriser les projets qui s'y inscrivent. Le label s'adresse à toute organisation, publique ou privée, ayant développé une solution numérique basée sur l'analyse ou le partage des données. Le label est un « repère de confiance » qui atteste des démarches éthiques entreprises au-delà du respect bien sûr de règles juridiques obligatoires. Il est aussi un gage de transparence pour rassurer les futurs usagers des solutions développées. Il répond également à une attente d'acteurs publics ou privés qui souhaitent valoriser leur démarche vis-à-vis de leurs partenaires (ou concurrents).

Le processus de labellisation repose sur un contrôle de conformité aux principes décrits dans la charte. Il se déroule en quatre temps. Une auto-évaluation en 10 questions détermine l'éligibilité du projet au processus de labellisation. Suit alors un travail déclaratif sous la forme d'un questionnaire en ligne, accompagné de l'envoi de pièces justificatives. Un organisme indépendant procède à des contrôles et à une évaluation complémentaire. Il soumet un avis contraignant qui induit, ou non, l'octroi du label par Occitanie Data.

La version bêta du label a été lancée en janvier 2022. Les premiers projets sont en cours d'évaluation.



DES MÉTHODES ET OUTILS

LA confiance naît aussi de l'élaboration de méthodes de travail communes. En cela, le choix des premiers cas d'usage sera déterminant car il permet de tester une communauté de vues et d'affiner des procédés. C'est la démarche qui a été retenue en 2019 par le Health data hub. Officiellement créé le 1^{er} décembre 2019 pour gérer et réguler les usages des données de santé au service de la recherche, le GIP encore en construction avait lancé un premier appel à projets dès le printemps 2019. Le Climate data hub en région Centre Val-de-Loire a, pour sa part, identifié près de 60 usages possibles de données mutualisées avant d'en faire le tri mais aussi d'alimenter à partir de cas concrets la réflexion sur les outils techniques à déployer pour collecter, héberger ou traiter ces données. Brest Métropole, pour engager une réflexion sur une charte éthique avec des citoyens (voir ci-après) a fait témoigner des experts et des praticiens exposant des projets et des réalisations concrètes avant d'identifier les options méthodologiques. La construction de la plateforme RUDI de Rennes qui permettra l'accès et le partage de données d'intérêt métropolitain (y compris les données personnelles de citoyens volontaires), a été lancée après de nombreux travaux sur des cas pratiques explorés en détail dans le cadre de la démarche de service public de la donnée.

QUI FAIT QUOI

Avant de détailler des outils et des méthodes, il est nécessaire d'identifier les parties prenantes et de convenir du rôle de chacune. Dans une récente publication, l'équipe de Nord Ouvert, distingue cinq rôles possibles entre partenaires⁴². Il y a les détenteurs de données, les utilisateurs de données, les services support qui amènent les moyens et une assistance fonctionnelle, juridique et technique, les bénéficiaires, ciblés par les finalités du dispositif, et enfin les autorités et les organes de contrôle qui garantissent la conformité ou le respect des règles. Par ailleurs, selon la forme choisie pour gérer la coopération, des organes directeurs peuvent être habilités à exercer une fonction de direction, de coordination, de supervision et de pilotage de l'initiative ou du cadre de gouvernance des données mis en place.

42. Sarah Gagnon-Turcotte, Miranda Sculthorp, Steven Coutts, *Les Partenariats de données numériques. Mettre les bases d'une gouvernance de données collaborative dans l'intérêt du public*, Nord Ouvert, 2021.

Ces rôles peuvent se cumuler. Ils peuvent aussi changer au fil de projets concrets. Voir permuer : un acteur sera fournisseur de données pour un usage précis et réutilisateur pour un autre.

DES OUTILS TECHNIQUES

Le choix des outils est une préoccupation importante. Les quelques projets français en cours accordent une importance majeure au choix de solutions souveraines notamment pour l'hébergement des données. L'objectif peut être à la fois stratégique, politique et juridique. La commission technique du Climate data hub a ainsi travaillé durant plusieurs semaines à définir les caractéristiques des outils numériques à mobiliser pour faire vivre les cas d'usage du projet. Des principes généraux ont été posés. Ils ont trait à la sécurité et à la protection des données, à la gestion des accès, à l'interopérabilité ou encore à la transparence (des données et des codes sources). En cohérence avec les finalités mêmes du projet, des exigences en termes de sobriété numérique ont été formulées. Au final, les partenaires ont décidé que l'hébergement et les traitements de données seraient orchestrés avec les infrastructures du Data Centre mutualisé piloté par le BRGM⁴³, l'INSA Centre Val-de-Loire et les universités de Tours et Orléans⁴⁴.

Occitanie Data a exploré d'autres aspects méthodologiques, engageant l'ensemble de ses membres à travailler de concert sur la robustesse des systèmes algorithmiques (notamment ceux reposant sur de l'apprentissage automatique), sur l'explicabilité de ces algorithmes ou sur la description et la traçabilité des données. L'équipe du futur GIP EKITIA explore également les diverses solutions technologiques qui existent pour orchestrer la gestion de données personnelles mises à disposition par des habitants à des fins d'intérêt général : coffres-forts numériques, *blockchain*, outils distribués gérés de façon coopérative, etc.

⁴³. Bureau de recherches géologiques et minières.

⁴⁴. Le Data Centre régional Centre Val-de-Loire est une infrastructure de haut niveau qui dispose notamment de moyens de calcul intensif. Opéré par le BRGM, il est l'un des deux premiers *data centers* labellisés par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.

dawex

Dawex est une société française leader des technologies de plateformes d'échange de données. Elle propose des outils d'orchestration du partage et de la circulation de la donnée reposant notamment sur des technologies de traçage et d'horodatage.

Soutenue par la Banque des Territoires qui est entrée à son capital, Dawex est l'un des acteurs importants d'une nouvelle économie du partage de la donnée appelée à se développer notamment pour la mise en œuvre des politiques publiques des villes intelligentes⁴⁵.

Commercialisée en France comme à l'étranger, la technologie de Dawex va notamment être utilisée pour la gestion du *Space data marketplace*, projet partenarial public / privé visant à faciliter l'accès aux données spatiales (projet soutenu par le CNES⁴⁶ et dont Occitanie Data est également partenaire).

Un outil ou un élément de méthode ne garantit pas seul et en soi le respect du cadre commun. Mais si l'un d'entre eux vient à défaillir c'est l'ensemble du cadre de confiance qui s'effrite. Les polémiques auxquelles est confronté le Health data hub suite au choix d'un hébergeur américain l'illustrent brutalement. Retenue sur des critères techniques et bien que reconnue comme très performante, l'offre de Microsoft est soumise aux lois extraterritoriales américaines faisant peser un risque sur la souveraineté des données de santé de 67 millions de Français. Quelle que soit la sophistication des procédures mises en place (pseudonymisation ou avatarisation des données, cryptage des données...), cette situation a été jugée inacceptable par plusieurs juridictions conduisant le Conseil d'État et la CNIL à imposer le « rapatriement » des données chez un hébergeur français ou européen. Ce récit simplifie et caricature sans doute un peu une situation très complexe, mais il montre comment des choix technologiques peuvent impacter la confiance collective dans un dispositif de gestion des données.

45. De la smart city à la réalité des territoires connectés, rapport officiel, ministère de l'Économie et des finances, octobre 2021.

46. Centre national d'études spatiales.

MAÎTRISER LES EXPÉRIMENTATIONS

Les experts de la donnée en sont friands : pour convaincre les décideurs de l'impact d'usages innovants de la donnée, il faut des POC⁴⁷ et des prototypes. Les expérimentations autour de la donnée ont un rôle important dans la construction d'un cadre de confiance, important mais délicat.

Elles sont souvent l'occasion d'associer les partenaires et d'engager un premier travail en commun autour de jeux de données simplifiés. Des données vont être mises en partage mais de façon limitée. On utilisera par exemple des données déjà stockées plutôt que des flux en temps réel. La période de traitement sera volontairement limitée en évitant des risques de violation de données protégées. Les prototypes data au sein d'un partenariat nouveau sont souvent spectaculaires car le croisement de données qui jusqu'alors n'étaient pas partagées produit rapidement des connaissances nouvelles. Ceci vaut pour des projets modestes en datascience. C'est ce que montrent les travaux de Résovilles avec des données inédites de Pôle emploi dans les quartiers de la politique de la ville. Et ceci vaut pour des prototypes très sophistiqués, comme celui du Conseil régional d'Occitanie utilisant de l'intelligence artificielle pour modéliser, à partir également de données de Pôle emploi, les besoins en compétences des entreprises de son territoire.

La construction de ces POC est l'occasion d'amener des partenaires à se connaître. Elle impose souvent la signature de conventions pour permettre le partage et les échanges de données. Ces conventions préfigurent les règlements ou les chartes à venir. Ou servent de test à des chartes en cours d'élaboration.

Mais ces prototypes peuvent être aussi source d'erreurs et compromettre durablement la confiance en construction. Au motif du caractère expérimental d'un projet, certains acteurs vont emprunter des raccourcis méthodologiques et s'affranchir de règles élémentaires que le cadre de confiance est supposé garantir. Les membres de l'Observatoire Data Publica ont vu à de nombreuses reprises des prototypes reposant sur des choix technologiques contraires aux principes affichés, sur des jeux de données trop sommairement anonymisés, sur des transmissions de données non sécurisées ou pire sur des consentements bafoués, au motif qu'il s'agit d'une expérimentation. Ces

⁴⁷. *Proof of concept* ou preuve de concept.

écarts commis par un acteur isolé exposent le fautif à un rappel à l'ordre (et la CNIL est intervenue à plusieurs reprises ces dernières années pour rappeler que le RGPD s'applique en toutes circonstances, y compris dans un prototype). Mais si ces écarts sont commis au sein d'un partenariat de données, ils peuvent compromettre la construction de la confiance.

C'est la raison pour laquelle il convient de contrebalancer la grande latitude d'essais et d'erreurs des phases de POC par de rigoureux processus de documentation des prototypes. Cette rigueur, qui devrait être inhérente à toute démarche expérimentale, permettra aussi de dégager des enseignements pour le cadre de confiance. Et s'il faut pour cela éprouver les limites du pacte en construction, ce doit être en toute transparence et dans les limites du cadre juridique existant. Il est intéressant de noter que la Charte métropolitaine de la donnée de Nantes encadre les expérimentations data conduites dans l'espace public et impose un conventionnement rigoureux aux acteurs publics et privés concernés. Si la preuve n'est pas faite du respect des principes et valeurs de la charte, l'expérience n'est pas autorisée.

DES CONTRÔLES

Des objectifs, des règles, des outils, de la méthode... les éléments constitutifs du cadre de confiance s'articulent les uns autres suivant des temporalités diverses. Nantes a commencé par élaborer et adopter une charte, Rennes a structuré une démarche partenariale expérimentale sur plusieurs thématiques partagées, l'Occitanie a créé une association avec les premiers partenaires intéressés, Brest a consulté les citoyens...

Mais tous ces territoires ont très vite été confrontés à la nécessité d'organiser et d'afficher l'existence de moyens de contrôle. Le contrôle, interne ou externe, est indissociable de la bonne application des règles du jeu au sein d'un cadre de confiance. Les partenaires sont tenus à une triple exigence : la redevabilité qui implique la transparence des actions menées et la preuve du respect des engagements pris, la responsabilité dans la mise en œuvre de tous les moyens nécessaires à leur bon accomplissement et l'efficacité des mesures prises au regard des objectifs initiaux. Mais qui vérifie que ces exigences sont respectées ? Dans un cadre de confiance plusieurs systèmes peuvent cohabiter. Les délégués à la protection des données des partenaires sont amenés à coopérer et à coordonner leurs interventions. La réalisation d'études d'impact, obligatoire dans

l'hypothèse où des traitements de données personnelles particulièrement innovants seraient engagés, peut relever de la responsabilité d'un partenaire ou de plusieurs. Les responsables de la donnée, *chiefs data officers* ou chefs de projet impliqués, auront aussi à coopérer et à opérer mutuellement des contrôles sur les usages des données mises en commun. Ils seront le cas échéant dotés d'outils de pilotage et d'orchestration des données à travers des plateformes dédiées.

Le contrôle peut aussi être diligenté par des acteurs externes, on pensera bien sûr à la CNIL.

Mais s'agissant de cadres de confiance conçus au service de l'intérêt général et des politiques publiques territoriales, il est un autre acteur important à associer : le citoyen.



LE RÔLE DES CITOYENS

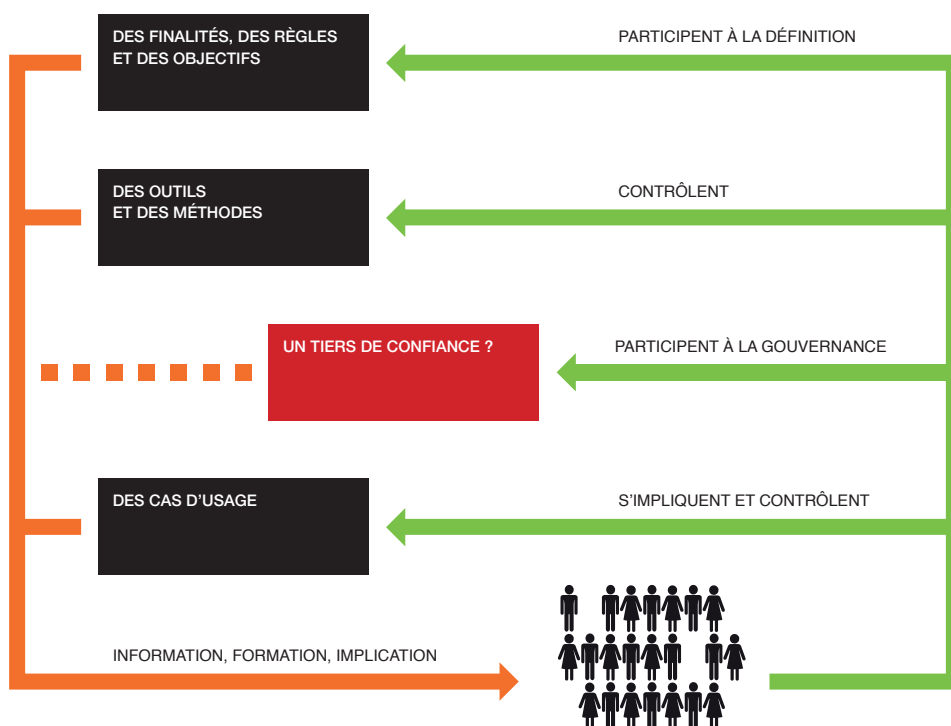
LE sondage et les panels de citoyens conduits par CIVITEO et Harris Interactive en septembre 2019 pour le projet Occitanie Data⁴⁸ ont détaillé les mécanismes de la confiance pour le citoyen face à des dispositifs complexes de partage, d'échange ou de traitement des données impliquant des acteurs publics et des acteurs privés. L'affirmation d'un motif d'intérêt général et la confiance dont bénéficie *a priori* les acteurs publics territoriaux dans l'utilisation qu'ils font des données ne suffisent pas.

Pour avoir confiance dans des dispositifs susceptibles d'exploiter massivement, même anonymisées, leurs données personnelles, les citoyens réclament des garanties. Ils souhaitent d'abord de la transparence sur l'utilisation des données à travers « une information claire et pédagogique » (76 % des sondés). Ils souhaitent que des contrôles indépendants soient possibles (73 %) diligentés par exemple par des comités éthiques associant des citoyens (71 %). Ils approuvent l'adoption de chartes éthiques (69 %).

Les expériences en cours dans les territoires français testent des solutions diverses en réponse à ces enjeux qui interrogent les fondements démocratiques d'un pilotage accru des politiques publiques par la donnée. Cette préoccupation ira probablement croissante dans les communes et EPCI qui généralisent des projets de ville intelligente.

Parmi les réalisations de collectivités pionnières, quelques exemples illustrent la variété des approches possibles. Nantes Métropole a eu recours à un panel citoyen pour définir des règles d'usage des données et écrire sa charte. La collectivité organise également des séquences de sensibilisation et de formation des jeunes citoyens à la protection des données personnelles. Occitanie Data a également souhaité associer des citoyens et prépare notamment leur implication dans des ateliers de prospective inspirés de travaux conduits à Montréal sur l'intelligence artificielle. En 2021, Brest Métropole a été le premier territoire français à organiser une conférence de consensus sur l'utilisation des données des habitants au service de l'intérêt général. Durant plusieurs semaines des citoyens se sont formés et ont auditionné des experts avant de rendre un avis très étayé sur les conditions d'acceptabilité d'usages nouveaux des données. Le produit de ce travail sera intégré à une charte qui sera également mise en concertation sur une plateforme publique. À Dijon, territoire pionnier dans le déploiement technologique d'une *smart city*, une délibération a été votée annonçant la création d'un comité éthique sur la donnée, incluant la participation de citoyens.

⁴⁸. *Les Français et le big data, le défi de la confiance*, Harris Interactive, CIVITEO, Cap pour Occitanie Data, septembre 2019.



*La place des citoyens dans un cadre de confiance pour la gestion des données au service de l'intérêt général.
Observatoire Data Publica*

L'association des citoyens n'est pas réservée aux acteurs publics. En témoignent de nombreuses initiatives de médiation et de data littératie organisées dans des tiers lieux. En témoigne aussi l'exposition « Souriez, vous êtes captés ! » organisée en septembre 2020 par des entreprises privées pour sensibiliser les citoyens et les aider à comprendre les enjeux de la captation des données dans l'espace public. Entre pédagogie et mise en garde, cette expérience présentait de façon concrète et réaliste les usages les plus innovants des territoires connectés et montrait pourquoi ils doivent être accompagnés d'une grande transparence vis-à-vis des citoyens⁴⁹.



⁴⁹. Le contenu de l'exposition est disponible sous forme d'un livre gratuit sur le site <https://souriezvoussetescaptes.fr/>

ENTRE CONSTRUCTION RATIONNELLE ET EXPÉRIENCE SUBJECTIVE : PRENDRE DU TEMPS

LA construction du cadre de confiance repose sur un processus complexe qui agrège des éléments techniques, méthodologiques, juridiques, éthiques... Mais la confiance ne s'appuie pas uniquement sur des garanties factuelles et tangibles, elle est aussi subjective. Dans ses travaux sur la confiance, l'Open Data Institute opère une distinction entre le fait d'être fiable et digne de confiance (*being trustworthy*) et le fait de bénéficier de manière effective de la confiance (*being trusted*)⁵⁰. Le premier se construit, le second se mûrit.

L'élaboration de la Charte métropolitaine de la donnée de Nantes a pris de longs mois. Au-delà du panel citoyen déjà évoqué, il y a eu des ateliers internes et des rencontres avec les acteurs publics et privés appelés à partager et mettre en commun leurs données. Le processus d'élaboration de la charte éthique a probablement fait tout autant pour la confiance que la charte elle-même. La construction d'Occitanie Data et sa transformation en GIP est le fruit de plusieurs années de collaboration entre des acteurs publics et des acteurs privés qui partageaient au démarrage une vision commune mais qui ont construit des habitudes de travail à plusieurs niveaux : entre laboratoires d'innovation, entre responsables de projets data, entre juristes et DPO, etc. Le processus engagé à Brest, qui débouchera sur la création d'un tiers de confiance (voir ci-après) s'inscrit dans le temps. L'étape de la concertation citoyenne est posée comme un préalable durant lequel l'appropriation des enjeux du partage des données au service de l'intérêt général progresse, y compris auprès des décideurs publics ou privés impliqués. La construction de la plateforme de données RUDI à Rennes, qui associe des acteurs publics, des entreprises, des associations et des citoyens, est également un processus construit sur un temps long. La préfiguration du Climate data hub du Centre Val-de-Loire a été préparée sur une année.

Se connaître, tisser des liens, créer des habitudes de travail, connaître et comprendre les intentions de chacun, sont aussi des éléments constitutifs du cadre de confiance. Il ne faut pas les négliger.



50. Open Data Institute, Designing trustworthy data institutions, 2020.

LES TIERS DE CONFIANCE

Il reste un point essentiel à décrire pour retracer dans ce cahier l'état des réflexions engagées en France par quelques territoires pionniers pour construire un cadre de partage des données au service de finalités d'intérêt général. Certains d'entre eux (l'Occitanie, la Région Centre Val-de-Loire, Brest Métropole...) considèrent que la confiance sera plus sûrement établie si une structure tierce se voit confier la responsabilité de la gestion des données. Nous allons décrire en détail le processus qui conduit à l'émergence de ce tiers.

« TIERS DE CONFIANCE » OU « ACTEUR DE CONFIANCE » ?

MAIS il faut auparavant revenir sur une spécificité française : une conception très exigeante et extensive de la notion de service public, l'attachement des Français à ce service public et la confiance qui lui est faite. Chaque année, plusieurs études montrent cet attachement et corrélient confiance et service public de proximité⁵¹. En matière de gestion des données ceci se traduit par un niveau de confiance *a priori* élevé⁵² qui conduit nombre de collectivités à considérer que « le tiers de confiance c'est la collectivité ». Cette formulation est impropre puisque la collectivité ne peut pas être à la fois l'acteur principal du processus de gestion, de la collecte jusqu'au traitement des données, et être considérée comme un « tiers ». En revanche elle traduit une réalité : des collectivités très avancées dans des processus de partage et d'échange de données, portant par exemple des projets de territoires intelligents, font le choix de conserver, seules, la maîtrise des données. Elles sont à la fois actrices des processus mais aussi, au nom du service public et de l'intérêt général, garantes du respect de règles établies. Sans recours à un tiers, ces collectivités veillent alors à demeurer des « acteurs de confiance ». Ce choix est celui de Dijon Métropole. Territoire pionnier de la *smart city*, Dijon a confié à un consortium d'entreprises privées (Bouygues, Suez, Citelum et Cap Gemini) la construction d'un hyperviseur et d'outils de pilotage innovants pour plusieurs services urbains essentiels. La Métropole a souhaité prendre la pleine responsabilité des données produites et se pose en garante des usages qui en sont faits. Ce choix est partagé et revendiqué par Paris qui est allé plus loin encore en construisant son propre *data center* ; par Rennes y compris avec le développement d'une interface d'accès aux données ouvertes aux partenaires privés et aux citoyens ; par Nantes dont la charte éthique conforte le rôle central de la collectivité publique ; par Toulouse qui intègre au sein de ses services une « agence de la donnée » ; par la région Grand Est dont la plateforme DataGrandEst est co-gérée avec l'État. De nombreux territoires s'inscrivent dans cette vision, car il apparaît naturel, en France, que l'acteur public joue ce rôle.

51. Voir notamment les enquêtes annuelles de l'Association des Maires de France ou le baromètre de l'Association des administrateurs territoriaux (<https://www.administrateurs-territoriaux.asso.fr/actualites/barometre-aatf-ipsos-2020-les-francais-et-le-service-public-local/>).

52. Voir les éléments détaillés précédemment, issus de l'étude *Les Français et le big data, le défi de la confiance*, Harris Interactive, CIVITEO, Cap pour Occitanie Data, septembre 2019.

D'autres collectivités ont pourtant fait un choix différent. Elles ont pour caractéristique commune d'anticiper de futurs échanges de données très importants entre acteurs publics et acteurs privés et/ou entre acteurs publics. Elles souhaitent équilibrer à la fois les attentes et les craintes des uns et des autres. Souvent, elles anticipent aussi la possibilité d'utiliser des données des habitants et veulent encadrer les usages qui en seront faits. S'il est légitime que l'acteur public se préoccupe fortement du respect des priorités d'intérêt général, il est légitime aussi que les habitants comme les entreprises privées veillent à ce que leurs données ne servent qu'aux finalités d'utilité publique qui motivent leur mise à disposition. Vu sous cet angle, l'acteur public n'est pas obligatoirement le plus légitime à agir.

C'est ainsi que naissent des tiers de confiance, c'est-à-dire des acteurs choisis ensemble par les partenaires pour gérer tout ou partie du partenariat de données. Le tiers peut être un simple intermédiaire ou le gestionnaire d'une plateforme d'échange. Mais il peut aussi être le porteur neutre du projet pour lequel les acteurs d'un territoire décident de mutualiser des données. Il peut se voir déléguer la définition des règles, l'écriture des processus, le choix des outils techniques, le contrôle des usages, le portage économique du projet. Il peut intégrer un volet réservé aux données personnelles de citoyens impliqués et s'assurer qu'aucun partenaire (public ou privé) n'en fasse un usage détourné.

Le recours à un tiers de confiance pour gérer des données au service de politiques publiques territoriales reste, à ce jour en France, un procédé peu courant. Avant de présenter quelques exemples, il est intéressant de regarder comment les choses se passent ailleurs.



LES DATA TRUSTS

Depuis quelques années, des experts anglo-saxons se penchent sur la constitution de nouveaux modèles économiques et juridiques favorisant le partage et la circulation des données, en confiance et au service de finalités d'utilité publique.

Ces organisations peuvent être de statut public, privé ou public-privé. La fondation britannique Nesta⁵³, propose une grille d'analyse pour identifier le mode de partenariat le plus approprié en fonction de deux critères : le caractère contraignant ou non du partage des données concernées et la valeur d'intérêt général de ces données.

Selon ce schéma, des données dont le partage est obligatoire et dont l'ouverture peut profiter à tous auront intérêt à être traitées dans un cadre de confiance totalement public. C'est le cas des nombreux portails *open data* gérés directement par les collectivités. Des données librement partageables mais dont l'intérêt collectif est limité pourront être gérés à travers des dispositifs de gestion des données personnelles. Entre les deux, il existe une variété de choix organisant la collaboration d'acteurs publics et privés (individus, associations, entreprises) : coopératives de données, espaces de partage publics dédiés à la recherche, *data trusts* associant public et privé, cadres de partage uniquement entre acteurs privés, etc.

Pour les projets d'intérêt général conduisant à la gestion partagée de données publiques et privées, ce schéma conduit à privilégier le *data trust*, que l'on peut traduire par fiducie de données. L'Open Data Institute décrit ce tiers comme une structure mutualisée, dotée d'une personnalité juridique autonome, utilisant une banque de données et travaillant dans un cadre de référence fait de règles définies par ceux qui décident de sa création au service d'objectifs partagés⁵⁴. Dans l'hypothèse d'un tel dispositif appliqué à la gestion des villes, des citoyens peuvent être impliqués à la fois comme producteurs de données, comme utilisateurs et comme parties prenantes de la gouvernance. On parle alors de « *civic data trusts* », une notion qui a été au cœur de nombreux débats et de quelques expérimentations à Toronto au Canada dans le contexte du projet de « ville Google » porté par Sidewalk Labs⁵⁵.

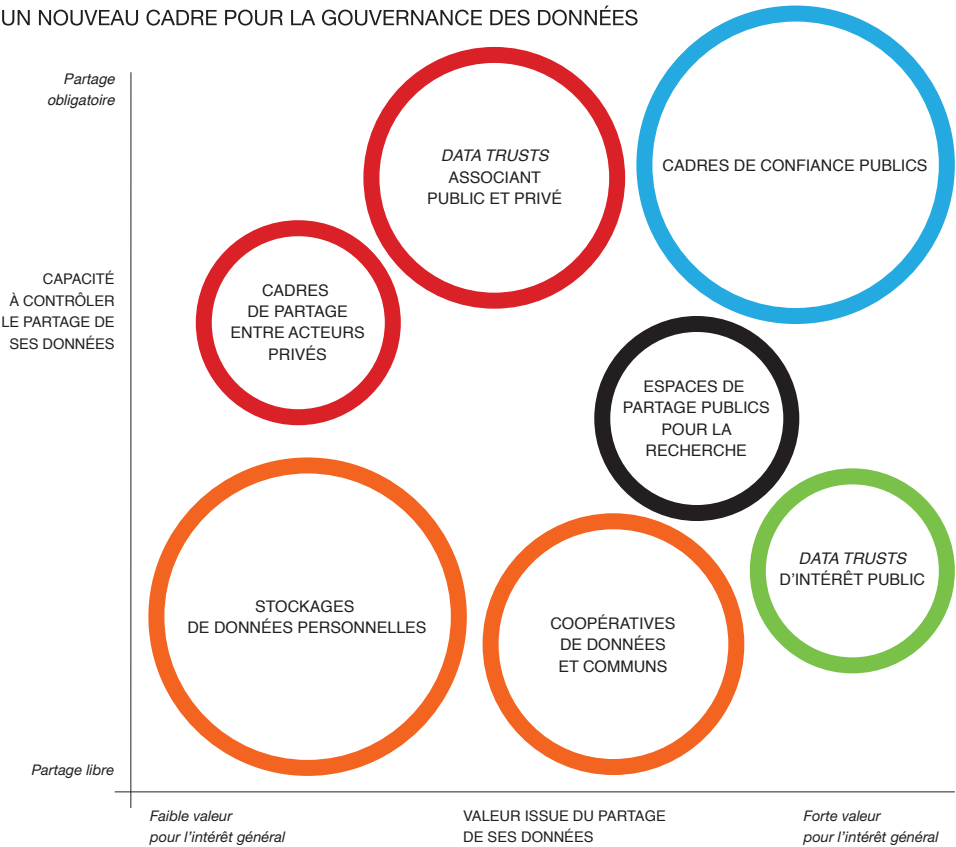


54. *Data Trusts: Lessons from Three Pilots*, Open Data Institute, 2019.

55. Jacques Priol, *Ne laissez pas Google gérer nos villes !*, Les Éditions de l'Aube, 2020.

53. <https://www.nesta.org.uk/blog/new-ecosystem-trust/>

UN NOUVEAU CADRE POUR LA GOUVERNANCE DES DONNÉES



Source : Fondation Nesta.

La création de ces *data trusts* permet donc de garantir, en confiance, aux partenaires fondateurs que leurs données seront utilisées uniquement à des fins d'intérêt général et suivant les règles qu'ils auront définies en commun. La diffusion de ce modèle est devenue un sujet de préoccupation mondialisée en raison du développement des usages de l'intelligence artificielle. Le contrôle et l'explicabilité de l'utilisation des données par de l'IA complexifie la quête de la confiance. Le Partenariat Mondial sur l'Intelligence Artificielle⁵⁶ a instauré un groupe de travail dont l'objectif est de définir les conditions du partage massif des données pour l'apprentissage d'Intelligences artificielles au service de missions d'utilité publique. Le modèle mis en avant est celui des *data trusts* ou des fiducies de données dont le fonctionnement est précisé. Une structure tierce est créé pour « *aider les producteurs de données à mettre en commun leurs données (ou leurs droits sur les données) dans le but de négocier collectivement les conditions d'utilisation avec les utilisateurs potentiels de données, par le biais de la supervision de fiduciaires indépendants, avec des obligations fiduciaires, et dans un cadre d'interventions techniques, juridiques et politiques qui facilitent l'utilisation des données et fournissent des garanties solides contre les utilisations abusives*⁵⁷ ».

Les experts, notamment européens et français, de ce groupe de travail sont confrontés à une difficulté importante : ce modèle de *data trust* repose sur une conception anglo-saxonne du droit et semble difficilement transposable de la *common law* vers le droit continental (et plus encore vers le droit public français).

⁵⁶. Voir notes 1 et 2.

⁵⁷. The GPAI Data Governance Working Group (2021), "Understanding Data Trusts", <https://ceimia.org/wp-content/uploads/2021/07/2021-07-09-GPAI-summary-understanding-data-trusts-updated.docx.pdf>

3 questions à anne-sophie hulin

Anne-Sophie Hulin, postdoctorante en droit à l'ANITI, Artificial and Natural Intelligence Toulouse Institute et spécialiste de la fiducie et gouvernance des données, nous apporte un éclairage sur le sujet.

Vous êtes experte des fiducies de données qui se développent au Canada et notamment au Québec. Pouvez-vous nous expliquer ce dont il s'agit ?

L'accès à des jeux de données pertinents et suffisamment volumineux est une nécessité pour offrir ou améliorer l'offre de biens et services numériques. Toutefois, derrière cette forte demande d'accès et d'exploitation des données par les organisations publiques comme privées, se cache un enjeu social et juridique majeur : garantir le respect des droits et des intérêts des personnes qui génèrent des données personnelles par leurs usages numériques (ex : respect de la vie privée). C'est dans ce contexte que le concept de fiducies de données a récemment émergé. Techniquement, la fiducie de données est un outil juridique grâce auquel des données sont confiées à un administrateur (un fiduciaire) qui a pour mission de veiller à un usage des données préalablement défini et soumis au respect de finalités précises par les organisations. S'il ne se conforme pas à sa mission, cet administrateur des données est susceptible d'engager sa responsabilité juridique. À travers cet outil, il est donc possible de mettre en œuvre une gestion « responsable » des données (y compris personnelles), et de créer un cadre de confiance au partage des données pour l'ensemble des parties prenantes (organisations, personnes).



Vous connaissez très bien les droits européens et français, Pouvez-vous nous dire si ce principe est transposable en France ?

La fiducie de données telle qu'entendue au Canada et au Québec n'est pas transposable à l'identique en France pour des raisons conceptuelles et juridiques majeures. En revanche, d'autres outils et structures juridiques en droit français sont capables de remplir un rôle quasi analogue. Par exemple, le groupement d'intérêt public (GIP) constitue la structure juridique plébiscitée pour la mise en œuvre de projets d'envergure de mutualisation des données comme celui du Health data hub. Si le GIP offre des garanties notables en matière de confiance dans les données à l'instar des fiducies de données, il n'est pas à proprement parler une institution fiduciaire ce qui change beaucoup l'approche du point de vue de la gouvernance des données et du contrôle qu'ont les personnes sur leurs données à l'issue de leur mutualisation.

De quels grands principes est-il (néanmoins) possible de s'inspirer pour un partage en confiance des données ?

La raison pour laquelle la fiducie de données intéresse tant les juristes réside dans son potentiel à atténuer les asymétries de pouvoir sur le marché des données personnelles. Ces asymétries jouent au détriment des personnes physiques. Tout le droit des fiducies s'articule autour du concept de confiance qu'il concrétise à travers des obligations strictes de loyauté, de diligence et de transparence à la charge des fiduciaires. C'est de ces leviers juridiques pour bâtir un cadre de confiance qu'il faut s'inspirer, transposer et/ou les renforcer en droit français. D'ailleurs, la proposition de règlement européen sur la gouvernance des données pourrait répondre prochainement à ces enjeux.



CONSTRUIRE UN TIERS DE CONFIANCE

Retour en France et examinons les démarches entreprises par quelques territoires engagés dans la voie de la création d'un tiers de confiance. La décision de créer un tiers peut être le fruit d'échanges entre partenaires publics et privés (Occitanie Data), une proposition de la collectivité faite pour engager la discussion avec des partenaires (Climate data hub Centre Val-de-Loire ou « mutuelle de la donnée » brestoise) ou encore une proposition d'entreprises à l'adresse de la collectivité (projet de tiers de confiance au sein de la *smart city* d'Angers Loire Métropole).

L'OBJET

La construction du tiers de confiance passe par plusieurs étapes. Elles peuvent être menées de façon successive ou concomitamment. Il convient que les partenaires s'accordent sur les finalités d'intérêt public des futurs échanges et traitements de données. Occitanie data a été créé pour « favoriser un développement éthique et responsable de l'économie de la donnée » avec une déclinaison à travers des cas d'usages dans différents domaines : santé, environnement, espace, énergie... Le Climate data hub Centre Val-de-Loire veut « construire un outil de mesure des trajectoires de lutte contre le réchauffement climatique [et] mettre à disposition des données pour permettre de l'innovation au service de la lutte contre le réchauffement climatique ». La « mutuelle de la donnée » de Brest veut « faire de la donnée d'intérêt général un bien commun, une ressource partagée et maintenue collectivement par une communauté ; établir des règles dans le but de la préserver [...] et de permettre à ses membres de l'utiliser ».



LE FONCTIONNEMENT

En Occitanie comme en Région Centre Val-de-Loire, des commissions de travail ont été instaurées pour travailler sur la forme juridique du tiers de confiance, sa gouvernance, ses règles et son cadre éthique, ses outils et ses choix technologiques et son modèle économique. Présidées par des partenaires différents, ces commissions ont formulé des recommandations qui peuvent se résumer ainsi : quelles orientations ou quels choix faire pour que chacun des membres puisse mettre des données à disposition avec la garantie qu'elles seront utilisées pour les finalités prévues et uniquement celles-là ? Et avec la garantie qu'elles seront exploitées en conformité avec l'ensemble des règles convenues (règles éthiques, conditions financières, transparence...) ?

Le rôle du tiers de confiance ressemble fort à celui du *data trust* ou de la fiducie de données. Mais avec une différence capitale : le fiduciaire est un opérateur indépendant et neutre à qui est confiée une mission par ses mandants ; les tiers de confiance territoriaux suivis dans notre étude, sont des structures tierces certes, mais dont les mandants sont parties prenantes. On notera au passage que dans ces projets, l'acteur public apporte au cadre collectif son « capital confiance ». Il peut aussi avoir un rôle légitime d'animateur car il représente, avant que cette mission ne soit confiée au tiers, l'intérêt général.

La construction des projets prend du temps. Il faut l'organiser. En Occitanie, une association de préfiguration a été créée. C'est elle qui a porté l'ensemble des travaux des commissions. Elle sera dissoute lors de la création de la structure finale (renommée EKITIA). En Centre Val-de-Loire, la préfiguration a été prise en charge par le Conseil régional et les services de l'État mais les commissions ont été animées par des acteurs variés. À Brest, la démarche est innovante à plus d'un titre et la préfiguration est portée par l'agence d'urbanisme ADEUPa qui présente la caractéristique d'intervenir sur un périmètre qui dépasse les limites de la métropole et couvre une large part de l'Ouest breton.

brest métropole

La Métropole de Brest a engagé une démarche de création d'une « mutuelle de la donnée ». Ce tiers de confiance a été imaginé comme l'acteur clef d'une stratégie locale d'un service public de la donnée visant à faciliter le partage et la réutilisation des données produites par les acteurs du territoire, en lien avec les territoires voisins de l'Ouest breton, au service de projets d'intérêt général.

L'idée de prendre appui sur les valeurs mutualistes pour créer ce tiers de confiance est originale. Le projet de Brest transpose au monde de la data des principes de solidarité et d'entraide, sans but lucratif et pour faire de la donnée un bien commun. La structure en cours de définition pourrait offrir plusieurs grandes familles de services : un socle de services innovants et opérationnels pour les usagers, une aide à la gestion des services pour les collectivités territoriales et un support au développement d'usages innovants pour les partenaires du territoire⁵⁸.

La « mutuelle de la donnée » est imaginée pour agir de façon solidaire. Des services développés à l'échelle de la métropole auront vocation à bénéficier aux territoires voisins. Des données mais aussi des outils et des compétences (la datascience) seront accessibles à des collectivités de taille modeste. C'est l'une des raisons du choix de confier la préfiguration du tiers de confiance à l'Agence d'urbanisme ADEUPa, association loi 1901 qui traite déjà d'enjeux de données avec de nombreux territoires du Finistère et des Côtes-d'Armor.

En revanche, la construction du cadre éthique ne lui est pas confiée. Brest Métropole a fait le choix d'organiser en amont de la construction du tiers, une concertation citoyenne inédite. Une première étape en ligne a eu lieu en juillet 2021 pour identifier des questions que se posent les habitants sur la gestion des données (et tout particulièrement de leurs données) par les acteurs publics. Ces éléments ainsi de quelques travaux d'experts ont alimenté ensuite une conférence de consensus. Durant plusieurs semaines des citoyens se sont formés aux enjeux de la donnée (aspects juridiques, méthodes de travail, enjeux politiques et stratégiques...) et ont auditionné des experts. Ils ont formulé des recommandations fortes à l'aide d'une méthode de travail supervisée par des garants (un représentant associatif et un représentant de la CNDP⁵⁹). Ces recommandations vont servir à rédiger une charte éthique qui sera elle-même remise en concertation publique.

Le futur tiers de confiance devra intégrer les principes de la charte dans son processus de travail. Il n'en sera pas l'émetteur mais le garant.

Le statut final du tiers brestois n'est pas encore défini. Après différents travaux exploratoires avec le concours d'experts mutualistes, la collectivité a fait le constat que le Code de la mutualité française n'autorisait pas la création d'une véritable « mutuelle de la donnée ». La future structure portera néanmoins les valeurs mutualistes, sous un statut à définir.

⁵⁸. Source : conférence de presse de lancement du service public de la donnée, Brest 3 juin 2021.

⁵⁹. Commission nationale du débat public.

CHOISIR UNE FORME JURIDIQUE

À ce jour, les tiers de confiance construits en France pour partager localement de la donnée publique et de la donnée privée au service de l'intérêt général sont soit associatifs, soit sous forme de groupements d'intérêts publics. D'autres formes sont sans doute possibles. L'idée de créer des coopératives de données est explorée dans différents domaines (projet Mobicoop pour des données de mobilité territoriale, projets de coopératives de données de santé...). Occitanie Data, le Climate data Hub ou plus récemment Brest Métropole ont exploré la piste mutualiste, coopérative (SCIC) ou celle d'une structure d'économie mixte (SPL, SEM, SEMOP...).

En l'absence de statut de fiduciaire utilisable en France pour les sujets data⁶⁰, il est possible que diverses formes nouvelles soient testées dans les prochaines années. Y compris pour mettre en œuvre les futures dispositions européennes qui inciteront à la création d'organismes « altruistes » de données. Le modèle économique de chaque tiers aura une incidence sur son statut. Financé par ses seuls membres en rémunération des services rendus il pourra être une association ou un GIP. Financé par des services offerts à l'extérieur du périmètre, il pourra être une société (ou également un GIP).

Le choix du modèle dépendra aussi sans doute du contexte d'exercice du tiers : a-t-il une compétence territoriale limitée ? A-t-il une finalité thématique unique ? A-t-il une durée limitée ou s'inscrit-il dans une stratégie à long terme ? Un tiers de confiance territorial annoncé en 2020 retient l'attention. Choisi pour mettre en œuvre le projet de territoire intelligent d'Angers Loire Métropole, le consortium d'entreprises piloté par ENGIE a proposé que La Poste joue le rôle de tiers de confiance numérique et assume notamment la responsabilité de l'hébergement des données. Le projet d'Angers est très ambitieux, des investissements conséquents sont prévus (178 millions d'euros sur 12 ans) et de nombreuses fonctions urbaines seront pilotées par la donnée : énergie, eau, déchets, environnement, bâtiments, mobilité, stationnement, sécurité... Le cahier des charges précisait que la gestion des données devrait offrir des garanties importantes à la collectivité en matière de souveraineté publique et aux habitants en matière de protection des données personnelles. La proposition de faire appel à La Poste, institution publique et de proximité reconnue des Français, est une réponse inédite. Elle préfigure peut-être l'arrivée de nouveaux acteurs, opérateurs de confiance.

60. La « fiducie à la française » a été créée en 2007 pour de la gestion patrimoniale (familiale ou d'entreprise).

DATA TRUST OU DATA HUB ?

À côté des quelques tiers de confiance territoriaux présents dans cette étude, se développent de nombreux *hubs* de données. Les deux termes sont confondus et parfois employés l'un pour l'autre.

Les « *hubs* » renvoient à des réalités diverses. Parmi ceux qui ont une finalité d'utilité générale, certains sont nationaux : le Health data hub, l'Education data hub ou les plateformes transports.data.gouv.fr et datatourisme par exemple. Ils peuvent avoir été construits par des filières économiques (NumAlim ou Agdatahub dans l'agro-alimentaire). D'autres projets sont régionaux : Occitanie Data et APIDAE, le Climate data hub en Centre Val-de-Loire ou le Data hub de la région Normandie. D'autres encore sont métropolitains : plateforme et *hub* de données de La Rochelle, RUDI à Rennes Métropole, plateforme OnDijon, etc.

Le *hub* est un lieu d'accès et de circulation de la donnée. Il peut s'agir d'un lieu de stockage, d'une plateforme technique accueillant des données thématiques, d'un entrepôt ou d'un lac de données, d'une interface... Certains offrent des outils de traitement (algorithmie, cartographie, datavisualisation...). Parfois les *hubs* ne sont que des regroupements d'outils existants. Nous voyons ainsi se multiplier les espaces de consolidation de différents observatoires publics.

Si un *data hub* n'est pas nécessairement un tiers de confiance, la plupart des tiers de confiance intègre ou s'appuie sur un espace de partage.



CONCLUSION

LA confiance ne se décrète pas, elle s'acquiert et se conforte par la mise en place de dispositifs de gestion rigoureux et fiables. Faisant appel à un tiers construit avec des partenaires ou reposant sur la légitimité forte d'un acteur public, un cadre de partage et d'utilisation de données au service de l'intérêt général doit offrir de multiples garanties : organisationnelles, méthodologiques, techniques, juridiques, économiques...

La démarche n'est pas aisée et elle prend du temps. Les exemples français décrits dans le présent cahier montrent que le processus de construction d'un cadre de confiance contribue sans doute autant à la confiance que le cadre lui-même. L'égrégore d'un collectif partenarial mérite sans doute beaucoup plus d'attention et d'énergie que ne l'imaginent au départ les porteurs de projet.

Il n'existe pas de modèle « clef en main ». Nous n'avons pas assez de recul et les initiatives abouties sont trop peu nombreuses pour en proposer. Certes, des briques sont disponibles et peuvent permettre aux uns et aux autres de gagner du temps : la charte éthique de Nantes, le label EKITIA, la technologie de Dawex... Mais chaque contexte et chaque projet est unique.

Par ailleurs, des questions nouvelles émergent. La redevabilité et la transparence qui sont au cœur des principes de la confiance nécessiteront notamment que les partenaires des projets d'utilité publique s'accordent sur la transparence des algorithmes qu'ils utilisent. L'enjeu n'est pas mince, *a fortiori* avec le développement du recours à l'intelligence artificielle. Il y a fort à parier que d'ici peu chaque cadre de confiance en construction (et chaque tiers de confiance installé) soit tenu d'adopter des méthodologies d'étude d'impact des algorithmes pour lutter contre d'éventuels biais (et de possibles discriminations⁶¹).

Dans cette attente, tous sont d'ores-et-déjà soumis à des obligations juridiques de transparence algorithmiques. Elles sont trop souvent ignorées.

Ce sera le thème du cahier n°2 de l'Observatoire Data Publica. À suivre.



61. Défenseur des droits et la CNIL, *Algorithmes : prévenir l'automatisation des discriminations*, mai 2020.

REMERCIEMENTS

L'équipe de l'Observatoire Data Publica tient à remercier tout particulièrement les collectivités territoriales ainsi que nos amis de la 27^{ème} Région qui ont participé à la journée d'étude du 23 janvier 2020, point de départ des travaux qui sont rassemblés ici. Cette journée d'étude avait été organisée sous la règle de *Chatham House* permettant à chacun de partager dans le détail ses réalisations, ses interrogations et ses éventuelles difficultés.

Les collectivités représentées étaient les suivantes : Agglopolys, Brest Métropole, la région Bretagne, la région Centre Val-de-Loire, Dijon Métropole, la ville d'Évry-Courcouronnes, le département de l'Eure, celui du Finistère, Grenoble Métropole, Nantes Métropole, la Région Occitanie, Orléans Métropole, Grand Paris Sud, Grand Paris Sud Est Avenir, La Rochelle et la Ville de Paris.

Étaient également représentés : la Banque des Territoires, la CNIL et ETALAB.

Nous tenons aussi à remercier nos partenaires canadiens : le Pr Christophe Abrassart de l'Université de Montréal qui avait participé à la journée de janvier 2020, Anne-Sophie Hulin pour sa contribution au présent cahier, ainsi que l'équipe de Nord Ouvert dont les travaux résonnent outre-Atlantique avec ceux de l'Observatoire Data Publica.

Nous remercions enfin les territoires dont les réalisations illustrent ce cahier.



L'OBSERVATOIRE DATA PUBLICA

Créé en janvier 2020, l'Observatoire Data Publica est une association loi 1901. Il s'appuie sur l'expertise de ses membres fondateurs, les cabinets de conseil CIVITEO, DATACTIVIST et INNOPUBLICA et le cabinet PARME Avocats. Pionniers de la gestion publique des données en France, ils ont souhaité mettre en commun et rendre disponibles des savoir-faire acquis auprès de collectivités et d'administrations publiques dans un cadre ouvert et non lucratif.

L'Observatoire Data Publica a été créé pour observer les pratiques nouvelles de gestion publique des données : émergence de « services publics locaux de la donnée », chartes éthiques, formes innovantes de gouvernance et de management de la donnée, prototypes de datascience et usages inédits d'algorithmes, recours à l'intelligence artificielle, etc.

L'Observatoire produit des récits et des analyses, il facilite les retours d'expérience et propose des enseignements sur ces nouveaux usages de la donnée. Ses fondateurs veulent faire de cette connaissance un bien commun.

Des études sont publiées dans Les Cahiers de l'Observatoire qui traitent de sujets souvent inédits ou proposent une grille de lecture nouvelle sur des problématiques connues. Les cahiers s'inscrivent dans la continuité des recherches les plus récentes au service d'une mise en œuvre opérationnelle et pragmatique.



