

OPEN DATA : LE TEMPS DE CHANGER D'ÈRE



INTRODUCTION

La *MIT Technology Review* établit chaque année sa liste des dix technologies de rupture, l'un des classements les plus attendus des entreprises de technologie. Cette année, elle a choisi de placer les *data trusts* dans cette liste, un modèle de gouvernance de la donnée qui permet la création d'alliance entre des acteurs publics et privés pour créer de la valeur en commun. Issus du rapport Pesenti Halle qui a défini la stratégie industrielle pour l'Intelligence Artificielle au Royaume-Uni, ces *data trusts* ne sont pas des technologies. Ce sont d'abord des types de contrat et des écosystèmes. La Commission européenne s'en inspire dans sa stratégie européenne pour la donnée à travers le concept de *european data spaces*.

La gouvernance des données ne doit plus être déterminée par une frontière hermétique entre la donnée propriétaire et inaccessible et le monde de l'open data. L'enjeu central pour la puissance publique est de favoriser un continuum de solutions en assurant une juste valorisation de l'effort de collecte et de traitement des données, tout en poussant au maximum pour la constitution d'écosystème ouverts.

La France a pris le tournant de l'open data depuis une dizaine d'année et la création de data.gouv.fr en 2011. Elle l'a fait pour des raisons militantes en même temps que démocratiques. Elle a pris une longueur d'avance sur le sujet, reconnue à l'échelle européenne : en 2020, l'étude *Open Data Maturity* conduite par le *European Data Portal* la positionnait ainsi dans le trio de tête à l'échelon européen, dans la catégorie des avant-gardistes.

Cette dynamique marque le pas principalement parce qu'elle s'est focalisée sur un horizon trop étroit : la bascule de jeux de données existants vers un modèle ouvert ; le cœur du sujet devient aujourd'hui d'accélérer les dynamiques de collecte des données et de créer des opportunités de partage. Le principe d'« ouverture par défaut », consacré par la loi pour une République numérique de 2016 est encore loin d'être appliqué. Mais plus encore, la formidable surface de contact et d'interaction des administrations publiques et opérateurs n'est pas utilisée pour ce qu'elle est : l'un des éléments les plus essentiels du patrimoine de l'État, comme acteur de confiance auprès des citoyens.

La publication du rapport Bothorel marque une rupture de temporalité. Ce rapport a été remis au Premier Ministre fin décembre 2020. Réalisé avec l'appui du préfet Renaud Vedel, coordonnateur national pour l'intelligence artificielle, et de Stéphanie Combes, directrice du *Health Data Hub*, il a réalisé un tour d'horizon de rythme et de méthode. Faute de quoi la France risque de perdre son avance et sa souveraineté technologique.

Les enjeux sont de taille. Pour les citoyens, la transparence est la première brique de la confiance. L'*open data* donne une capacité restaurée aux citoyens pour se saisir des politiques publiques, les évaluer et tester des options alternatives. Pour l'administration, l'*open data* permet de se confronter à un écosystème ouvert d'intelligence autour des politiques publiques et à construire des politiques publiques fondées sur les preuves (*evidence based policies*). Pour les milieux économiques et académiques, l'*open data* stimule l'innovation, favoriser la croissance et la recherche.

Les retombées économiques sont très fortes. L'étude *The Economic Impact of Open Data* a examiné près de 15 travaux de référence. L'analyse a permis d'établir un impact économique médian de l'*open data* à 1,19% du PIB actuel. Cela représente 184,45 milliards d'euros à l'échelon européen (UE27), soit une valeur de 27,63 milliards d'euros pour la France.

Pour suivre ces évolutions nous avons souhaité contribuer à la réflexion du rapport Bothorel. Ce document reprend la contribution que nous avons réalisée pour ces travaux. Nous souhaitons aussi nous projeter autour de plusieurs convictions. **Trois axes nous paraissent prioritaires.**



AXE 1

L'INTEROPÉRABILITÉ DES DONNÉES PUBLIQUES RESTE UN CHAMP DE BATAILLE

L'exploitation du patrimoine de données publiques reste freinée par un grand morcellement. Les bases de données sont de qualité, de granularité et de périmètres hétérogènes, et par conséquent difficile à mettre en commun et réutiliser.

A l'origine de ce phénomène, plusieurs explications : les données sont générées dans des systèmes sources eux-mêmes hétérogènes, ayant chacun leur format et leur finalité spécifique ; les réflexes associés à une « culture de la donnée » – la capacité à envisager une multiplicité d'usages, mais aussi à anticiper les prérequis du partage et du croisement de données – demeurent encore trop rares chez les acteurs susceptibles de la produire et de la réutiliser ; les circuits de remontée de ces données ne sont pas industrialisés ou ne suivent pas un tempo suffisant.

L'enjeu est double pour les acteurs publics. D'une part, consolider le patrimoine de données à des échelons intermédiaires – comme celui d'un hôpital, d'une ville ou d'un rectorat – selon un format « interopérable », garant du respect de

la vie privée et de la sécurité des données. D'autre part, assurer une « mise en qualité » de ces données.

L'interopérabilité implique une sémantique et syntaxe partagées, une structuration commune, une doctrine de métadonnées et un catalogage. Par « mise en qualité », il faut entendre lisibilité, complétude, exactitude, récente et mise à jour régulière, cohérence, accessibilité ou encore maîtrise de la confidentialité. Deux engagements que la mission Bothorel invite à tenir pour accélérer et passer à l'échelle. Elle propose notamment la création d'un label de « service producteur de la donnée » et d'une politique d'interopérabilité des données à l'échelon interministérielle, conditions sine qua non pour réussir le développement d'intelligences artificielles.

Autant d'impératifs fixés par la norme FAIR (*Findable, Accessible, Interoperable, Reusable*) qui conditionnent notamment l'accès aux dispositifs de financement européens.

RÉUSSIR UNE DÉMARCHE D'OPEN DATA : LE CAS DE LA LOMBARDIE

Pour comprendre les enjeux d'ouverture et d'interopérabilité du patrimoine de données, l'« approche lombarde » est particulièrement éclairante. Dès 2012, la région italienne a lancé son portail Open Data. En 2014, le gouvernement régional a publié une première liste de jeux de données recommandées pour l'ouverture auprès des autorités régionales. Toutefois, en 2017, seules 18 municipalités sur 1507 avaient fourni les jeux de données.

Pour améliorer cette situation, les autorités régionales ont lancé un programme de financement pour couvrir les coûts engagés par les municipalités. Celles-ci devaient en contrepartie s'engager à mettre en place un flux automatique de publication de 10 ou 25 jeux de données dans une liste de 50 établie par la région. Le cofinancement couvrait une partie des coûts de mise à disposition – de l'ordre de 2000 € par municipalité, incluant à la fois les ressources humaines nécessaires pour préparer la donnée au sein de chaque municipalité, mais aussi les solutions technologiques requises pour publier les données. Un montant unitaire faible, mais qui ouvrait la possibilité de massifier les investissements technologiques, et de développer des solutions réutilisables entre les différentes municipalités. En complément du financement, la région a travaillé avec les parties prenantes locales à l'établissement de normes et standards, alignés avec les lignes directrices européennes.

Les résultats ont été spectaculaires : en une année, le nombre de municipalités engagées dans une démarche d'open data est passé de 18 à 136, avec près de 95 municipalités qui ont publié les jeux de données prioritaires identifiés. Il y a aujourd'hui 1275 jeux de données sur le portail, celles-ci sont mises à jour régulièrement, et suivent les standards de qualité requis pour permettre une réutilisation ultérieure. Le plan a eu particulièrement de succès auprès des plus petites municipalités, qui sont aujourd'hui près de 60% à avoir publié leurs données.

Sur la base de ce retour d'expérience, quatre facteurs s'avèrent essentiels pour développer, au niveau local, l'usage de l'open data : une impulsion politique forte, une incitation économique permettant de rétribuer les coûts engagés, la constitution collaborative de normes et standards, et la mutualisation des moyens technologiques.



AXE 2

FACILITER LE PARTAGE DES DONNÉES : LE RÔLE DÉCISIF DE L'ÉTAT

Explorer de nouveaux modèles de partage

La gouvernance des données ne doit plus être déterminée par une frontière hermétique entre la donnée propriétaire et inaccessible et le monde de l'*open data*. L'enjeu central pour la puissance publique est de favoriser un continuum de solutions en assurant une juste valorisation de l'effort de collecte et de traitement des données, tout en poussant au maximum pour la constitution d'écosystème ouverts.

Le partage de donnée répond à des enjeux multiples, aux premiers rangs desquels figurent le pilotage des politiques publiques par la donnée, la transformation des organisations publiques et privées ainsi que la compétitivité de la France dans la course globale à l'intelligence artificielle. Il peut prendre des formes très diverses selon les secteurs, les finalités et les acteurs impliqués : portails publics d'*open data*, *data hubs*, *marketplaces* ou encore approches dites '*self data*'.

Le positionnement de l'État peut varier selon les domaines, mais l'objectif reste le même : éviter la dispersion de moyens dans le développement d'infrastructures et de « tiers de confiance » facilitant l'émergence d'une chaîne de valeur de l'information et assurant les meilleurs standards de gouvernance de la donnée. La notion de tiers de confiance est cœur pour fluidifier les échanges au sein d'un écosystème d'acteurs, qu'ils soient publics, privés, ou public/privé. Concernant le premier modèle, le rapport Bothorel propose de repenser le service data.gouv.fr pour améliorer visibilité des données publiques et l'exploration de l'offre, mieux cibler l'utilisation d'APIs, et être davantage à l'écoute des ré-utilisateurs.

Le concept de *hub*, qui connaît un essor ces dernières années en France, permet notamment d'orchestrer le partage de données en apportant les garanties de gouvernance et sécurité requises pour accéder à des données personnelles ou propriétaires et nécessitant à cette fin des traitements préalables (anonymisation, pseudonymisation, croisement...).

Ces plateformes, que le rapport Bothorel qualifie de « puissant vecteur de décloisonnement de la donnée », recouvrent toutefois une réalité protéiforme. Elles peuvent par exemple prendre la forme d'une *marketplace* permettant l'échange de données tels le *AgDataHub* en agriculture, ou encore, un modèle plus centralisé comme celui du *Health Data Hub*, qui joue le rôle de tiers de confiance dans le partage de données sensibles.

Enfin, un autre mécanisme de partage prend de l'ampleur. Connue sous le nom de 'self-data', ce modèle laisse aux producteurs de la donnée – les individus dans le cas de données personnelles – le choix de partager leurs données à la demande et d'en contrôler les usages, en consentant / retirant les autorisations à leur gré. Ce modèle a notamment démontré son potentiel aux États-Unis sous la forme d'une initiative [Blue button](#), mise en place dès 2010 par l'administration Obama et inspirée par l'administration des Vétérans pour permettre aux citoyens américains d'accéder à l'ensemble de leurs données médicales, telles qu'elles sont collectées chaque jour par les acteurs du système de santé, et d'autoriser leur partage dans un format standardisé avec un écosystème approuvé de services de santé. L'initiative en a depuis inspiré d'autres, notamment le *Green button*, lancé par des acteurs de l'énergie en Californie.

Au-delà du choix de modèle, de nombreuses questions non triviales se posent au cas par cas pour assurer la pérennité d'un *hub* : qui est pertinent pour agir comme tiers de confiance entre les acteurs publics et privés partageant la donnée ? Quel modèle de gouvernance permettra de convenir d'éléments clés comme les formats de partage ou le type d'usages autorisés ? Et quel modèle économique permettra d'assurer un réel partage de la valeur ?

Données d'intérêt général

Les « données d'intérêt général » sont la nouvelle frontière du partage des données. En la matière le rapport distingue les « B2G », données du secteur privé utiles à l'administration, et les « B2B », initiative des partages de données au sein d'un secteur de l'économie. Il s'agit d'« inciter les acteurs économiques à la mutualisation de données » tout en respectant la liberté d'entreprendre, le droit de propriété et le secret des affaires. La puissance publique veut se placer comme animateur et financeur des comités stratégiques de filière pour inciter l'émergence d'écosystèmes organisés par filière. Du reste, si cette approche par secteur doit se poursuivre, la mission Bothorel invite à ne pas recréer des « silos sectoriels » et à réfléchir à leur interconnexion.

Ces travaux devront s'articuler avec l'ambition actuelle des instances européennes de créer un marché unique de la donnée à l'échelon continental. Annoncé dans son Livre Blanc sur l'Intelligence Artificielle en février 2020, la Commission prévoit d'impulser la création d'espaces de partage et valorisation de la donnée (*data space*) dans multiples secteurs – santé, mobilité, énergie, industries manufacturières... L'initiative GAIA-X portée par l'industrie franco-allemande poursuit à cet égard des objectifs similaires.

Une ambition que viennent accélérer le *Data Governance Act* et le futur *Data Act*, projets réglementaires contraignant allant dans le sens d'un partage croissant de données interopérables entre acteurs publics et privés à l'échelon européen.

De nouvelles obligations sont donc à prévoir demain pour les acteurs publics et privés.

DATA GOVERNANCE ACT : QUELLES IMPLICATIONS POUR LES PRODUCTEURS DE DONNÉES ET LES PLATEFORMES DE PARTAGE ?

Le 25 novembre dernier, la Commission européenne a présenté la première version du *Data Governance Act*. Actuellement en cours de négociation entre le Parlement et le Conseil européen (après quoi il entrerait en vigueur sous 12 mois), le projet de réglementation vise à faciliter l'accès aux données du secteur public.

Plus précisément, le chapitre II du *Data Governance Act* prévoit que les organismes publics doivent créer les conditions pour que les données qu'ils collectent puissent être partagées et réutilisées, y compris les données à caractère personnelle et celles protégées pour raison de propriété intellectuelle, ou d'autres obligations de confidentialité – excepté les informations pouvant porter atteinte à la sécurité nationale. Ce partage devra être permis sous la forme d'un accès à des API dédiées. Le paiement d'une redevance pourra être demandé par l'État pour couvrir le coût de mise à disposition des données, à condition de publier sa méthode de calcul et de prendre des mesures incitant la réutilisation de ces catégories de données par les TPE/PME et les usages non-commerciaux. Là où le partage par défaut n'est pas possible, l'organisme devra aider les « ré-utilisateurs » à obtenir le consentement des individus ou entités légales auxquels le partage pourrait porter préjudice, dans la mesure où un tel dispositif n'engendre pas des coûts « disproportionnés » pour le secteur public. Chaque pays membre devra se doter d'une autorité nationale veiller au respect des dispositions de la réglementation et soutenir les acteurs publics dans leur implémentation.

Le chapitre III du règlement établit un régime juridique applicable pour l'ensemble des services de partage de données. Que l'on soit opérateur d'un *hub* de données, d'une *marketplace*, de coopérative ou de trusts, ou même simplement fournisseur de plateforme pour ces acteurs, il sera nécessaire de notifier ses activités auprès de l'autorité nationale mentionnée précédemment pour exercer ces services au sein de l'UE. Surtout, les prestataires de tels services auront interdiction d'utiliser les données qu'ils mettent à disposition à d'autres fins que l'amélioration du service de partage. Idem pour les métadonnées collectées dans le cadre de ces activités. Le prestataire devra également veiller à ce que la procédure d'accès à son service soit équitable, transparente et non discriminatoire à l'égard tant des détenteurs de données que des utilisateurs de données, y compris en ce qui concerne les prix.

Concrètement l'UE souhaite empêcher la capture de la valeur économique par le petit nombre d'acteurs impliqués dans leur partage, afin d'encourager son usage. Les fournisseurs devront garantir « un niveau de sécurité élevé pour le stockage et la transmission de données » et de « mettre en place des mesures techniques, juridiques et organisationnelles appropriées afin d'empêcher le transfert de données ou l'accès à celles-ci dans les cas où ils sont illicites au regard du droit de l'Union », une référence claire à la nécessité de respecter l'arrêt « Schrems II » de la Cour de Justice Européenne.

Le *Data Governance Act* sera bientôt rejoint par un deuxième projet de réglementation, le *Data Act*, qui apportera des obligations similaires au premier, cette fois-ci pour les données issues du secteur privé. Ensemble, ces initiatives visent à créer les conditions juridiques, économiques et techniques propices au partage de données entre acteurs : la confiance, l'interopérabilité, et un marché unique de 448 millions de citoyens et consommateurs capable de rivaliser avec les marchés chinois et américain.



AXE 3

FAIRE ENTRER L'OPEN DATA DANS UNE ÈRE INDUSTRIELLE

La difficulté du « passage à l'échelle »

Il faut le rappeler, la valeur des jeux de données ne se réalise que dans l'usage. Partir du bénéfice « métier » qu'il est possible de tirer de la donnée, pour chaque acteur public, contribuera à accélérer la politique d'ouverture des données. L'accent mis sur l'usage, plutôt que sur l'ouverture en tant que telle, permettra en outre de redonner du sens au partage des données publiques.

Le développement des usages sera facilité par le recours à l'expérimentation et par une ambition politique claire. La facilitation de l'expérimentation reste donc plus que jamais nécessaire. Le rapport propose par exemple de créer un « bac à sable » expérimental pour déroger aux textes existant en matière de réutilisation de données personnelles dans les jeux d'apprentissage. Par ailleurs, l'inscription, de la valorisation du patrimoine de donnée comme l'une des missions d'un ministère constitue une piste très prometteuse qu'ont pris certains pays comme l'Estonie, connu pour être à l'avant-garde en la matière.

Au-delà de l'expérimentation, rappelons toutefois que l'accent doit être mis sur le déploiement industriel des projets de data et d'intelligence artificielle de l'administration. Si les « preuves de concept » semblent avoir fait florès dans l'administration depuis une dizaine d'année, trop rare sont les cas d'usage véritablement passés à l'échelle. Alors qu'elles n'étaient que 38% en 2017, [plus de la moitié des organisations privées](#) ont aujourd'hui réussi à dépasser l'étape du démonstrateur et à passer des cas d'usages d'intelligence artificielle en production.

La tendance de fond a été globalement la même pour tous les secteurs et toutes les géographies, avec des premières expérimentations en 2016 et 2017, et un passage progressif vers la production en 2018 et 2019. Si les organisations du secteur public ont démarré en même temps que les autres, elles ont particulièrement du mal à passer les pilotes en production. En 2020, 60% en sont encore au stade du démonstrateur et seules 40% d'entre elles considèrent être en retard par rapport à leur plan initial.

Pour y parvenir, l'État doit aujourd'hui surmonter plusieurs obstacles. Il s'agit notamment de

mettre en œuvre un modèle opérationnel spécifique, par la création d'une capacité dédiée, qui permettrait de se doter des compétences adéquates, de mutualiser les moyens techniques et humains nécessaires, qui font souvent défaut, et d'agir concrètement sur les sujets d'éthiques et d'évolution juridiques.

Cela conduirait notamment à mettre en place des modèles industriels de développement de cas d'usage *Data Factory* articulés autour d'une plateforme technologique mutualisée et d'une ambition interministérielle affirmée de développement des compétences. Ces initiatives nécessitent également d'établir un portefeuille de politiques publiques « pilotes » - qui pourraient faire l'objet d'une expérimentation de « bout en bout » et de partenariats ouverts avec les territoires et les citoyens.

Talents

Pour accélérer en matière d'ouverture et de valorisation des données publiques, l'administration peine à se doter des talents nécessaires. Il s'agit de compétences encore trop rares sur le marché de l'emploi – la tendance naturelle est de se tourner vers le secteur privé plus attractif en matière de salaire et considéré plus proche de la frontière technologique. Le rapport invite premièrement à nommer un administrateur général de la donnée, des algorithmes et des codes sources (AGDAC)

positionné dans les services du Premier ministre, pouvant s'appuyer sur un réseau d'administrateurs ministériels des données, des algorithmes et des codes sources (AMDAC). Successeurs des Administrateurs Ministériels des Données (AMD), ces derniers incarneraient la politique d'ouverture des données à l'échelon ministériel et seraient dotés de moyens humains, techniques et financiers accrus. Compétents tant en matière technique que juridique, ils seraient à la fois chargés de définir la politique d'ouverture des données mais aussi de négocier le partage de données avec les autres administrations et les acteurs privés. Un nouveau rôle défini via une fiche de poste type et une lettre de mission signée par le ministre. Parmi les initiatives déjà mises en œuvre ces dernières années par l'État, la mission Bothorel invite notamment à pérenniser et davantage structurer le dispositif d'Entrepreneurs d'Intérêt Général, qui permet de doter temporairement des projets publics de profils *ad hoc* possédant les compétences attendues.

Enfin, la formation et de l'acculturation des acteurs publics (cadres dirigeants, fonctionnaires, élus...) demeure en outre très imparfaite. Des sessions de formation à la donnée et à l'intelligence artificielle devront être généralisées pour les principaux acteurs concernés afin de prendre conscience des cas d'usages au sein de leur propre administration et de la nécessité de poursuivre l'ouverture des données.

Nous sommes convaincus que la France doit agir sans tarder pour tirer pleinement partie de la donnée dans la transformation de l'action publique. Dans la suite de la publication du rapport Bothorel une première série de décisions ont été prises qui accéléreront l'accès aux données publiques. Mais le rôle de l'État doit aller au-delà de ce type d'actions et accompagner la structuration de nouveaux écosystèmes de données. L'avenir est plus que jamais à l'*open data*. Il est aussi aux *data trusts*.



A propos de Capgemini Invent

Capgemini Invent est la marque d'innovation digitale, de conseil et de transformation du groupe Capgemini qui aide les décideurs à concevoir et construire le futur de leurs organisations. Forte de plus de 6 000 collaborateurs, basés dans plus de 30 bureaux et 22 studios de création à travers le monde, Capgemini Invent conjugue une expertise en stratégie, technologies, data science et creative design, avec des connaissances sectorielles approfondies pour bâtir les solutions digitales et business models de demain.

Capgemini est un leader mondial, responsable et multiculturel, regroupant 270 000 personnes dans près de 50 pays.

Partenaire stratégique des entreprises pour la transformation de leurs activités en tirant profit de toute la puissance de la technologie, le Groupe est guidé au quotidien par sa raison d'être : libérer les énergies humaines par la technologie pour un avenir inclusif et durable. Fort de plus de 50 ans d'expérience et d'une grande expertise des différents secteurs d'activité, Capgemini est reconnu par ses clients pour répondre à l'ensemble de leurs besoins, de la stratégie et du design jusqu'au management des opérations, en tirant parti des innovations dans les domaines en perpétuelle évolution du cloud, de la data, de l'Intelligence Artificielle, de la connectivité, des logiciels, de l'ingénierie digitale et des plateformes. Le Groupe a réalisé un chiffre d'affaires de 16 milliards d'euros en 2020.

Get the Future You Want (Réalisez le futur que vous voulez)

Plus d'informations sur

www.capgemini.com/fr

Auteurs :

Pierre Demeulemeester

Vice-President, AI & Data Strategy

pierre.demeulemeester@capgemini.com

Léo Isaac-Dognin

Directeur, European Public Sector

leo.isaac-dognin@capgemini.com

Jean-Guillaume Messmer

Manager, Citizen Services

jean-guillaume.messmer@capgemini.com