

DU BIG DATA AU SMART-GOUVERNEMENT : DE LA VALEUR STRATÉGIQUE À LA CRÉATION DE VALEUR PUBLIQUE

par Nicolas SPATOLA, Dr, Chercheur
Josefina GIMENEZ, Directrice Recherche et Innovation

Le Big Data, sphère d'information en constante évolution générant de fortes attentes des acteurs publics et privés, est définie comme un ensemble très volumineux de données dépassant les capacités des outils classiques de gestion de bases de données ou de l'information. Dans le monde entier, différents opérateurs sont dédiés à la production et l'exploitation des données pour en extraire de l'information. Et ce dans des domaines multiples et variés comme la météorologie, la démographie, la géographie, la médecine, etc.

Mettre de l'ordre et tirer du sens de ces données est un défi à la fois technologique et social.

Se basant sur les 5V du Big Data mais aussi les 7 lois de l'information proposées par Moody et Walsh, nous proposons un modèle intégratif décrivant les courants évolutifs de la donnée. Dans celui-ci, le partage, l'utilisation, la véracité, l'association, l'autogénération, ainsi que la variété agissent en interaction pour transformer la donnée en information et lui conférer une valeur stratégique (voir schéma page 7).

Le Big Data est au cœur des enjeux de souveraineté européenne. Une stratégie de fluidification de la donnée a ainsi été mise en place dans le cadre du Data Governance Act. Cette réglementation vise à donner aux entreprises et aux startups un meilleur accès à davantage de données dans le but de développer de nouveaux produits et services, mais considère l'ensemble des échanges entre gouvernements, citoyens et entreprises. Elle comprend trois volets principaux :

- Les conditions en vue desquelles les organismes du secteur public pourront consentir à la réutilisation de données détenues par les acteurs publics qui ne font pas l'objet du droit à la protection des données personnelles, de la propriété intellectuelle ou de la confidentialité.
- Les données à caractère personnel ou non personnel avec les règles applicables aux activités de partage de données entre acteurs privés.
- L'altruisme concernant le partage de données entre organisations privées et à but non lucratif pour favoriser l'émergence d'applications d'intérêt général.

Cependant, en prenant en compte la faible intégration des technologies digitales dans les entreprises européennes et la grande disparité dans ce domaine entre les différents pays européens, la présentation des objectifs de la commission européenne semble optimiste et dénote des faiblesses dans l'approche systémique de la production, de la gestion et des capacités d'exploitation de la donnée.

Il est important de garder en tête que toutes les données ne sont pas vouées à être partagées, le concept de secret professionnel en est une preuve. Les données personnelles sont régulées par différents acteurs à travers les nations, notamment par le règlement général sur la protection des données (RGPD) dans l'Union Européenne. Ainsi, établir une stratégie de la donnée européenne revient à réfléchir à l'équilibre entre protection des données personnelles, des données de la zone UE et des modalités de partage entre pays membres de la zone UE mais également avec les partenaires extérieurs. Les enjeux géopolitiques sont forts.

C'est sur cette compréhension des Big Data qu'a été défini le concept de smart-gouvernement. Le smart-gouvernement a pour but de connecter les environnements physiques, numériques, publics et privés afin de déterminer de nouveaux modèles de prestation de services. Ce type de gouvernement entreprend une numérisation de ses process par le biais des technologies de l'internet et fait de l'exploitation de la donnée un de ses objectifs principaux. Il se place ainsi aux côtés de ses organisations du secteur public comme moteurs du cycle de création de la donnée, responsables de générer et partager la connaissance, en développant des méthodes d'intégration de l'exploitation de la donnée dans leurs habits.

Šiugždinienė et collègues ont proposé un cadre d'analyse du smart-gouvernement en tant que construction, en opérationnalisant quatre dimensions clés :

- **La dynamique stratégique** qui s'appuie autant sur la sensibilité stratégique du gouvernement que sur sa flexibilité organisationnelle
- **La collaboration intersectorielle** entre différents acteurs et secteurs, qu'ils soient privés, publics, gouvernementaux ou non
- **La collaboration inter-institutionnelle** car la collaboration, coordination et coopération des acteurs publics est un prérequis du smart-gouvernement
- **L'engagement des citoyens**, à travers leur participation dans les discussions politiques et retours sur les mesures existantes.

L'ensemble de ces dimensions va à la fois représenter des pratiques politiques dont l'information et la digitalisation (i.e. la facilitation du flux de la donnée) sont des outils pour des processus de prise de décision plus efficaces et, de manière générale, plus collaboratifs.

La question qui se pose alors concerne les modalités de développement d'un smart-gouvernement : quelles sont les bases fondamentales en termes de collecte, de partage, de génération de la donnée et de prise de décision associées à un tel processus ?

Pour définir un cadre normatif, on peut en premier lieu considérer que :

- **La gestion de la donnée doit être légalement acceptable** : le cadre législatif doit être spécialement préparé dans ce but.
- **La gestion de la donnée doit être éthiquement acceptable**.
- **La gestion de la donnée doit être socialement acceptable**, sans perte de légitimité dans sa transition, indispensable pour la conservation d'un climat de coopération.

Une fois le cadre normatif identifié, l'implémentation des principes fonctionnels d'un smart gouvernement peut se définir sur les dimensions de la gouvernance, de la technologie, de la culture et des compétences. L'ensemble de ces dimensions est important, et une implémentation pérenne présume le développement de chacune d'entre elles.

Dans ce rapport nous explorerons le concept de Big Data recoupant les évolutions de flux de la donnée mais aussi leur utilisation potentielle. Nous vous présenterons les principaux éléments de compréhension concernant la donnée en tant qu'objet et la manière dont l'information stratégique peut être produite à partir de celle-ci, élément capital de nos jours pour les preneurs de décision.



SOMMAIRE

I. Le Big Data comme valeur stratégique	06
• Le processus de cyclage de la valeur stratégique de la donnée	07
• Des caractéristiques modératrices de la valeur.....	09
◦ La vitesse	10
◦ Le volume	10
◦ La véracité et la qualité, la problématique des biais	11
• Les acteurs de la donnée	14
II. Entre ouverture et protection des données personnelles	16
• Les principes de gestion	16
• Des modèles d'échange et de protection de la donnée	18
◦ La protection des données ...	19
◦ Une disparité des acteurs qui suppose une disparité des applications	20
III. Vers un smart-gouvernement	21
• La valeur stratégique de la donnée dans le secteur public : le cas de la valeur publique.....	21
• Les dimensions du smart gouvernement	25
◦ La dimension de dynamique stratégique.....	26
◦ La dimension de collaboration intersectorielle	27
◦ Le dimension de collaboration interinstitutionnelle	28
◦ La dimension de l'autonomie et de l'engagement des citoyens	29
• Le développement de smart-gouvernement	29
◦ Le cadre normatif	30
◦ Le cadre d'implémentation	31
IV. La nécessité de fournir un cadre d'évaluation de la maturité de la gestion des données	33
• Un modèle de maturité	33
• L'exemple du Data Maturity Assessment Framework au Royaume-Uni.....	34
V. Conclusion	36
VI. Références	37

I. LE BIG DATA COMME VALEUR STRATÉGIQUE

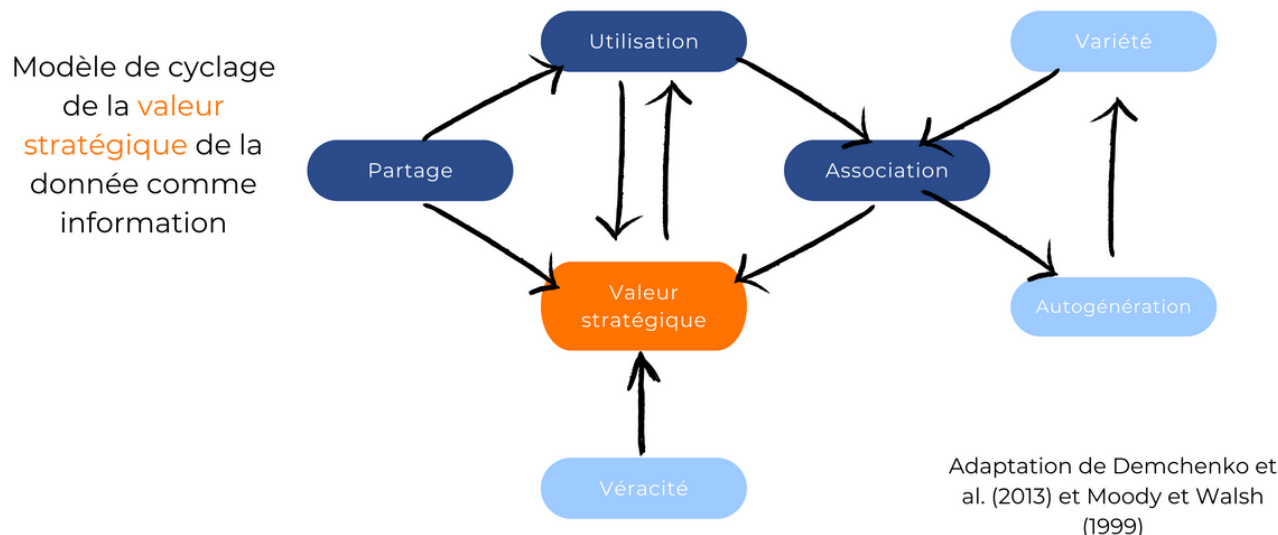
Le Big Data est un potentiel outil pour transformer la gouvernance traditionnelle en un smart-gouvernement¹¹. Dans un premier temps, nous développerons le Big Data comme concept illustrant la valeur stratégique de la donnée. Nous proposerons un modèle de cyclage de cette valeur stratégique afin de mieux comprendre les caractéristiques de la donnée. Nous discuterons des éventuels biais de celle-ci et des acteurs pouvant intervenir dans ce modèle de cyclage. Dans un second temps, nous nous intéresserons au lien entre la donnée et la gestion de la vie privée des usagers. Enfin, nous développerons le concept de smart-gouvernement en nous interrogeant sur la transition entre valeur stratégique de la donnée et production de valeur publique, puis les dimensions du smart-gouvernement et finalement les dimensions essentielles à son déploiement.

Malgré sa définition polymorphe, on peut considérer que le Big Data désigne un ensemble très volumineux de données dépassant les capacités des outils classiques de gestion de bases de données ou de l'information. A titre d'exemple, chaque seconde, 29.000 Giga-octets (Go) d'information sont publiés dans le monde, soit 2,5 exaoctets par jour, ou 912,5 exaoctets par an. En comparaison, un film en 4k sur Netflix pèse environ 7 giga-octets. On produit donc l'équivalent en données de 4143 films 4K par seconde. Ces données proviennent des messages que nous envoyons, de notre consommation d'outils numériques, d'internet, des textes, de la musique ou des vidéos que nous publions, de l'ensemble des capteurs que nous avons créés qu'ils soient à l'entrée d'une bouche de métro ou dans un satellite en orbite autour de la terre.

En France, des opérateurs sont dédiés à la production et l'exploitation de données, comme l'Institut national de la statistique et des études économiques, l'Institut géographique national, l'Institut national de la santé et de la recherche médicale, Météo France, l'Institut national d'études démographiques et l'Office national de l'eau et des milieux aquatiques. Organiser et tirer des enseignements de cette quantité d'informations est un défi technologique et social.^[3]

LE PROCESSUS DE CYCLAGE DE LA VALEUR STRATÉGIQUE DE LA DONNÉE

Différents cadres théoriques proposent une approche de la donnée.



Les 5V du Big Data englobent 5 principes clé qui définissent le Big Data : la valeur, la vitesse, la variété, le volume et la véracité. On peut en dresser un parallèle avec les sept « lois » (7L) de l'information proposé par Moody et Walsh.

Afin de simplifier la compréhension de ces différents modèles,^[2] Perspectives a défini un modèle intégratif, se basant sur les 5V et 7L, pour dresser une taxonomie du Big Data dans une approche d'outil informationnel, et pour définir la donnée par sa valeur stratégique.

Figure 1. Ce modèle consiste en un cyclage de la valeur stratégique de la donnée comme information, et en des caractéristiques modératrices de ce cyclage propres aux données - ©Artimon Perspectives

Au centre de ce modèle, la donnée est conceptualisée comme un élément de valeur stratégique. Ce qui nous intéresse ici n'est pas la valorisation de la donnée mais de sa valeur en tant qu'information. Cette représentation se base sur les processus de transformation de la donnée par l'utilisation^[4].

Dans la taxonomie des 5V, la **valeur** est définie par la valeur ajoutée que les données collectées peuvent apporter à un processus, à une activité ou à l'analyse, la compréhension ou la comparaison de phénomènes. Cette valeur va être influencée par des facteurs comme le **niveau d'utilisation de la donnée**. En effet, plus une donnée est utilisée, plus sa valeur est élevée. Le niveau d'utilisation est un indicateur de l'intérêt qui est porté à l'information. On peut considérer quatre types d'analyses principales qui, en tant que production d'information, vont produire la valeur stratégique^[5]:

- **L'analyse descriptive** utilise des données pour décrire ce qu'il s'est passé lors de l'analyse de questions complexes.
- **L'analyse diagnostique** va plus loin en explorant et en triangulant les données pour expliquer pourquoi un

problème spécifique s'est produit, identifier ses causes profondes et déchiffrer les tendances structurelles sous-jacentes.

- **L'analyse prédictive** utilise des données et des algorithmes pour prédire ce qui est le plus susceptible de se produire, souvent en utilisant l'apprentissage automatique.
- **L'analyse prescriptive** propose ce qui doit être fait pour provoquer ou empêcher que quelque chose se produise.

Cela implique donc un second axiome : la valeur stratégique d'une donnée et son niveau de **partage** sont positivement corrélés. En effet, contrairement à d'autres ressources sécables diminuant leur valeur à mesure qu'elles se fractionnent, les données peuvent être partagées sans perdre leur valeur informative mais, à l'inverse, multiplier leur valeur stratégique par le développement de leur **utilisation** (ce qui n'est pas nécessairement le cas de la réplication). Le partage de données permet de nouvelles exploitations (nouvelles analyses et perspectives) sur celles-ci et plus d'opportunité **d'association** avec d'autres données.

Par exemple des données d'utilisation d'un service et des données de satisfaction sur ce même service sont des sources d'information de valeur. Combinées, elles permettent d'accéder à un nouveau niveau d'analyse.

Une conséquence de cette association est de permettre l'**autogénération** de nouvelles données et donc de nouvelles informations. Que ce soit par l'utilisation unitaire ou la mise en comparaison ou combinaison avec d'autres données, l'exploitation de la donnée comme ressource permet de générer de nouvelles données et donc des ressources. Cette autogénération va permettre d'accroître la **variété** des données et, rétroactivement, influencer le potentiel d'**association**. Une seconde rétroaction est ; plus la **valeur** stratégique de la donnée est importante, plus son **utilisation** augmente.

Une fois ce processus défini, on peut cibler les vecteurs d'action les plus intéressants pour les services afin d'accroître la valeur stratégique de leur donnée : partager et associer.

DES CARACTÉRISTIQUES MODÉRATRICES DE LA VALEUR : LA VITESSE, LE VOLUME, LA VÉRACITÉ ET LA QUALITÉ

Après avoir présenté la valeur de la donnée et l'importance du cyclage, il est important de considérer les caractéristiques du flux de la donnée qui vont modérer le processus de cyclage en accélérant / freinant, facilitant / inhibant la génération de valeur à partir de la donnée. Un flux par définition est caractérisé en premier lieu par le temps entre la séquentialisation de ses composants et le nombre desdits composants. Nous discuterons donc les caractéristiques de vitesse (ou ^[2-3]vélocité) et de volume. Deuxièmement, on peut caractériser le flux par les propriétés de ses composants. Dans le cas de la donnée, deux propriétés sont à considérer comme fondamentales : la véracité de la donnée par rapport au phénomène qu'elle représente et la qualité de celle-ci. Ces deux propriétés sont les plus importantes car elles réfèrent aux ^[2-3-6]problématiques de biais.

LA VITESSE

La vitesse (ou vélocité) est la mesure de la rapidité du flux de la donnée. Par exemple, 1074 photos sont téléchargées sur Instagram chaque seconde. Le flux du processus présenté précédemment varie ainsi selon la vitesse de circulation de la donnée, qui dépendra de facteurs comme le nombre d'utilisateurs, l'architecture numérique sur laquelle la donnée circule et le niveau d'ouverture. De manière opérationnelle, cela implique d'allouer des moyens pour assurer une stabilité du flux et une gestion de la véracité et de la qualité de la donnée.

LE VOLUME

Le volume des données est un facteur crucial à prendre en compte. Le volume représente la quantité de donnée générée dans une période définie. Par exemple, selon le rapport de l'International Data Corporation (IDC), en 2020, 64.2ZB de données numériques ont été créés ou répliqués dans le monde. Selon le modèle de l'IDC la création et la réplique mondiale de données connaîtront un taux de croissance annuel de 23 % au

cours de la période 2020-2025. Selon le rapport de l'IDC, en 2020 seulement 10% du volume des données créées a été sauvegardé^[7]. Cela démontre que beaucoup de données sont temporaires (remplacées ou supprimées).

De manière positive, le volume permet d'accroître la fiabilité des modélisations, des analyses et des algorithmes se basant sur l'exploitation de données. De manière négative, volume ne signifie pas qualité. Et cela engendre des problématiques importantes de stockage qui interrogent sur les espaces et la consommation d'énergie (à relativiser avec l'optimisation des outils de stockage). Pour exemple, selon les estimations, les centres de données utiliseraient environ 200 TWh en 2020. Ils consomment entre 1 et 1.5 % de la demande mondiale d'électricité^[8].

Tant dans une perspective économique qu'écologique, cela implique de mettre en place des pratiques dans les organisations de différentes manières :

- Identifier la valeur stratégique des données pour définir si elles doivent être stockées.
-

- Identifier le cycle de vie des données pour ne pas conserver des données obsolètes (une des caractéristiques identifiées par Moody & Walsh, 1999).
- Développer l'utilisation des outils de gestion de cycle de vie des données.
- Privilégier le stockage objet, c'est-à-dire le stockage d'une seule copie facilement accessible et identifiable.
- Mettre en place un processus de déduplication à grande échelle.

LA VÉRACITÉ ET LA QUALITÉ, LA PROBLÉMATIQUE DES BIAIS

Les facteurs influençant la **qualité** et la **véracité** (ou fiabilité) de la donnée sont multiples. La donnée se **déprécie** avec le temps (le rythme dépendant de la donnée en question). Des critères peuvent être émis pour caractériser la qualité de la donnée comme :

- L'**exactitude** et la précision de ce que la donnée est censée représenter.
- L'**exhaustivité** des données.
- La **cohérence** et l'**intégrité** des données afin d'éviter les phénomènes de contamination.

- La **conception** de métadonnées pour assurer un suivi et une meilleure gestion.

Nous avons établi que le Big Data pouvait être considéré comme une source de production d'information pour objectiver des phénomènes du réel. Cependant, cet axiome n'est valide que si la donnée représente ces phénomènes sans biais et avec précision.^(FIG 2) C'est le critère de véracité de notre information. La véracité ne souffre pas seulement des informations considérées comme « fausses » mais aussi des biais présents dans les jeux de données et de la précision de ces dernières, ce qui pourrait conduire à une réplification des biais dans le cas d'une prise de décision basée sur ces informations.

Un biais peut être défini comme une erreur systématique, et la précision comme le niveau de dispersion des données. Il est fondamental de prendre en compte ces critères car si l'on considère la donnée comme source d'information sur le réel, cette représentation du réel peut être erronée.

Dés, lors, les décisions prises sur la base de l'interprétation de ces informations seraient ^[6] intrinsèquement inadéquates.

Voici quelques exemples de biais cognitifs observés lors de l'utilisation de données :

- Biais de confirmation : la tendance à collecter ou traiter l'information d'une manière qui confirme ses idées préconçues, ses croyances ou ses hypothèses.
- Biais d'observation : la tendance à collecter de l'information dans des situations ou contextes favorisant les résultats escomptés.
- Biais d'intérêt ou de financement : la tendance inconsciente à orienter les modèles, la collecte de données ou les interprétations des données d'une manière qui favorise les objectifs d'un financeur ou d'un employeur.

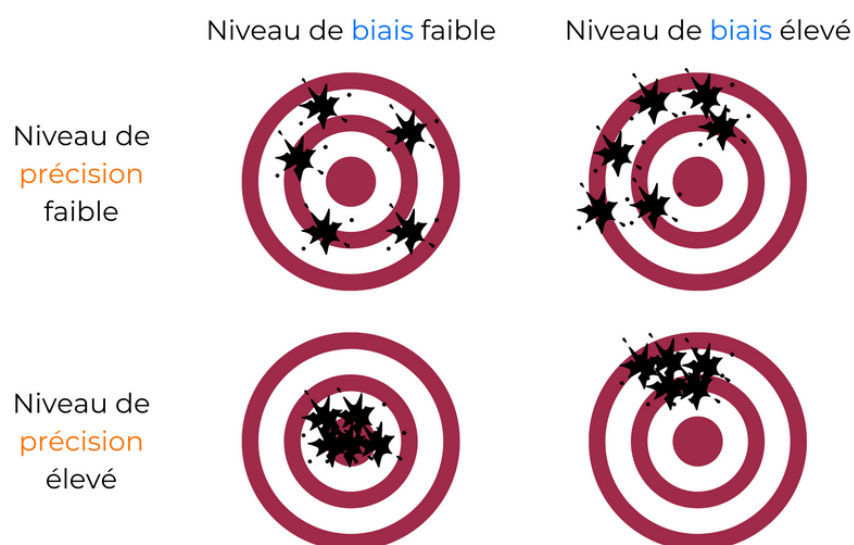


Figure 2. Cadre d'évaluation de la donnée en deux axes précision/biais -©Artimon Perspectives

Pour contrer ces problèmes de précision et de biais, un contrôle des sources, de l'information, du contexte de production (temporel et géographique), de la méthode de production et d'acquisition

sont nécessaires. Cependant, il est important de considérer un fait simple : malgré toutes les initiatives humaines et technologiques pour « débiaiser » des données, l'idée de « données débiaisées » est oxymorique.

Les biais émergeant du fonctionnement de la société et la donnée étant un proxy, il est aujourd'hui impossible d'accéder à la mesure totale d'un phénomène social (ce qui nous intéresse dans le cadre des services et institutions publics) et donc à des données fondamentalement complètes, précises et objectives.

« L'occurrence de biais involontaires codés dans les algorithmes et les applications d'IA découle du biais présent dans les données sous-jacentes. Les ensembles de données biaisés créent des modèles biaisés et conduisent à des prédictions et à des décisions biaisées. Il peut y avoir de graves implications en matière de discrimination, affectant des groupes disparates, sur la base de facteurs discriminatoires ou d'indicateurs de discrimination, tels que le sexe, l'âge ou d'autres caractéristiques inhérentes »^[9].

Il existe cependant des outils et méthodes statistiques permettant de réduire les biais que l'on peut trouver au niveau statistique ou au niveau cognitif. Dans une stratégie structurée autour de la donnée, cela implique des équipes formées à ces outils et méthodes, comme des data-

scientists ou des statisticiens pour les biais statistiques, secondés par des chercheurs en psychologie sociale ou en sociologie pour l'identification des biais cognitifs relatifs au contexte de production de la donnée.

Certains éléments sont fondamentaux à identifier et à analyser pour évaluer la qualité de la donnée. On peut citer en premier lieu la prise en compte d'informations comme le format, les motivations de l'obtention et de la publication des données, les enjeux financiers et conflits d'intérêts, le contexte éthique, légal et social d'utilisation. Ces différents éléments définissent a priori des indices sur l'évaluation des biais potentiels au niveau statistique ou cognitif qu'il sera nécessaire de corriger dans la mesure du possible.

Au niveau statistique :

- On peut tenter de **compléter l'information** même si cela comprend un risque^[10]. Ce processus revient à identifier un manque de représentativité dans les données et à compléter lesdites données avec l'information manquante. Par exemple, si j'étudie la

mobilité des emplois de cadres et que mes données ne considèrent que les cadres féminins, je peux rechercher les données des cadres masculins et les ajouter à ma base.

- Une autre possibilité est le **redressement des données**, qui vise à corriger l'échantillon enquêté de ses éventuelles déformations par rapport à la population cible de l'enquête.
- Le **rééchantillonnage** est une des méthodes permettant d'estimer la précision des statistiques d'échantillon en définissant des sous-ensembles pour s'assurer de leur similarité. Il consiste aussi à sélectionner de manière randomisée des points de données parmi l'ensemble des données pour créer de nouveau sous-ensembles.
- L'**ajout de données** pour compléter l'information et s'assurer que les effets observés entre une cause et un effet ne sont pas dus à une variable non contrôlée.

Considérer les dérives qui correspondent à l'évolution des données à travers le temps. Une mesure des valeurs managériales dans une entreprise en 2012 sera évidemment différente en 2022, démontrant

l'obsolescence intrinsèque des données.

Au niveau cognitif, Il faut aussi considérer que la cognition humaine est constituée de biais d'analyse à prendre en compte en amont et en aval de la collecte et l'utilisation de la donnée, pour identifier des biais à la fois chez les émetteurs et les utilisateurs de la donnée. En effet, une même information peut être interprétée de manière différentes selon les objectifs, les intentions ou même la culture. Des compétences pour identifier l'influence des contextes peuvent être pertinentes dans ce cadre (ex. sociologues, économistes, psychologues sociaux).

LES ACTEURS DE LA DONNÉE

Si nous avons présenté la donnée, sa valeur et son cyclage, il est aussi important de mentionner les acteurs de l'écosystème. On peut distinguer des grandes familles d'acteurs de la donnée.

Ici nous en présentons 6 :

- **la personne concernée qui représente l'entité à propos de laquelle les données sont générées ou collectées.** Il peut s'agir de personnes, d'animaux, d'objets (ex. capteurs, compteurs d'électricité, feux de signalisation, toits, etc.)
- **le producteur/fournisseur de données,** qui va agréger la donnée à la fois de lui-même mais aussi d'autres entités en utilisant un système ou en étant monitoré par un système
- **le responsable du traitement** assume en dernier lieu la responsabilité légale des modalités de stockage, de partage et d'utilisation des données. Selon le Règlement Général sur la Protection des Données (RGPD) il est responsable de :
 - La collecte de données à caractère personnel
 - La définition des fondements juridiques de cette collecte de données
 - La définition des finalités pour lesquelles les données à caractère personnel doivent être traitées
 - La définition des catégories de données à caractère personnel à traiter
 - La détermination des destinataires de ces données

- La communication des droits des personnes et la manière dont ces droits peuvent être exercés (Charte ou Déclaration Vie Privée)
- La détermination des délais de rétention des données (conservation, sécurisation, effacement)



- **le processeur des données** comprend l'ensemble des organisations s'attelant à la gestion et à l'analyse de données, comme des serveurs de stockage de données, la gestion de bases de données, des logiciels d'analyse et des ressources de cloud computing.
 - **l'analyste des données,** comprend les entités procédant à l'agrégation et à l'analyse de la donnée pour créer de l'information.
 - **l'utilisateur final,** qui représente le preneur de décision, en d'autres termes la traduction comportementale de l'utilisation de la donnée devenue information.
-

II. LE BIG DATA ENTRE OUVERTURE ET PROTECTION DES DONNÉES PERSONNELLES

Puisque la donnée peut être considérée comme une valeur stratégique, de nombreuses initiatives fleurissent pour accroître cette valeur. Comme démontré par notre modèle de cyclage, le meilleur moyen de créer de la valeur est de rendre la donnée disponible au plus grand nombre en temps réel. Si ce postulat semble applicable à certains types de données, comme le nombre de passages aux portiques à la gare Saint-Lazare pour prévoir les flux et adapter le service, d'autres s'avèrent plus problématiques à ouvrir. Par exemple, rendre disponible les données médicales de l'ensemble des citoyens d'un pays en temps réel pourrait contrevenir à la protection de la vie privée, et cela même si l'ouverture permet de créer de la valeur publique. Il existe donc un ratio à trouver entre l'ouverture des données et la protection de la vie privée.

PRINCIPES DE GESTION

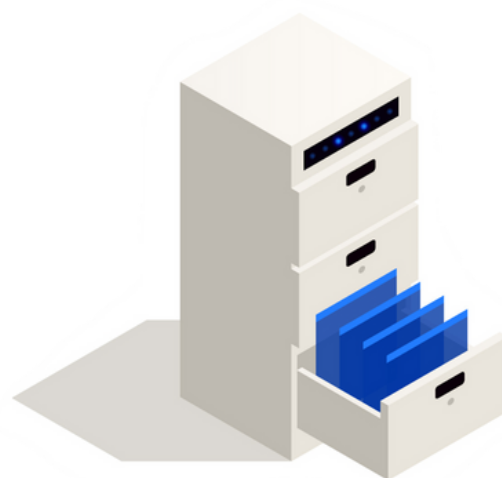
L'OCDE met en avant des principes autour de la gestion de la donnée pour assurer que la collecte, le stockage et l'utilisation soient justifiés au regard des différentes contraintes sur la protection des données personnelles⁽¹⁾:

- **Principe de limitation de collecte** : il doit y avoir des limites à la collecte de données à caractère personnel, et celles-ci doivent être obtenues par des moyens légaux et équitables et, le cas échéant, avec la connaissance ou le consentement de la personne concernée.
 - **Principe de qualité des données** : les données à caractère personnel doivent être pertinentes au regard des finalités pour lesquelles elles sont utilisées, exactes, complètes et tenues à jour.
-

- **Principe de spécification des objectifs** : les finalités pour lesquelles les données à caractère personnel sont collectées doivent être précisées au plus tard au moment de la collecte des données. L'utilisation ultérieure doit se limiter à la réalisation de ces finalités ou d'autres compatibles, avec des précisions à chaque nouvel objectif de collecte de données.
- **Principe de limitation** : les données personnelles ne doivent pas être divulguées, mises à disposition ou utilisées à des fins autres que celles spécifiées, sauf en cas de consentement de la personne concernée ; ou par l'autorité de la loi.
- **Principe des garanties de sécurité** : les données personnelles doivent être protégées par des mesures de sécurité raisonnables contre des risques (perte, accès non autorisé, destruction, utilisation, modification ou divulgation de données).
- **Principe d'ouverture** : existence de dispositions générales sur l'ouverture, les développements, les pratiques et les politiques en matière de données personnelles. Des moyens devraient être facilement disponibles pour établir l'existence et la nature des

données à caractère personnel, ainsi que les principales finalités de leur utilisation, l'identité et la résidence habituelle du responsable du traitement.

- **Principe de participation individuelle** : les individus devraient avoir le droit d'obtenir d'un responsable de traitement la confirmation qu'il détient ou non des données les concernant, et d'avoir eu au préalable la communication sur les données éventuelles dont il dispose. Tout refus devrait être motivé et contestable.
- **Principe de responsabilité** : un responsable de traitement est responsable du respect des mesures qui donnent effet aux principes énoncés ci-dessus.



DES MODÈLES D'ÉCHANGE ET DE PROTECTION DE LA DONNÉE

Si l'on s'intéresse aux institutions de l'Union européenne, l'objectif est de fluidifier les échanges de données à travers l'émergence d'un marché unique des données, reposant sur quatre piliers^[2]:

→ **Pilier 1** : un cadre de gouvernance intersectoriel pour l'utilisation des données et de l'accès au travers d'un cadre législatif pour la gouvernance des espaces européens communs de données. Il existe une législation sur les ensembles de données de grande valeur, sur l'analyse de l'importance des données dans l'économie numérique et un cadre politique dans ce domaine.

→ **Pilier 2** : des investissements pour renforcer les capacités et les infrastructures de l'Europe pour leur hébergement, traitement, utilisation et interopérabilité des données (cloud européen).

→ **Pilier 3** : l'autonomie des individus, l'investissement dans les compétences et dans les PME pour utiliser les outils numériques de manière plus efficiente, et leur permettre de mieux gérer leurs données.

→ **Pilier 4** : des espaces européens communs de données dans les secteurs stratégiques et les domaines d'intérêt public en mesurant les flux de données et en estimant leur valeur économique en Europe, et en dehors de l'Europe.

Ce marché devrait se développer dans le cadre réglementaire du **Data Governance Act** (DGA), une législation qui vise à donner aux entreprises et aux startups un meilleur accès à davantage de données dans le but de développer de nouveaux produits et services. Le règlement comporte trois volets principaux :

- Les données détenues par les acteurs publics qui ne font pas l'objet du droit à la protection des données personnelles, de la propriété intellectuelle ou de la confidentialité. Le texte définit les conditions techniques et juridiques selon lesquelles les organismes du secteur public pourront consentir à la réutilisation de telles données protégées.
 - Les données à caractère personnel ou non personnel avec les règles applicables aux activités de partage de données entre acteurs privés.
-

- L'altruisme en matière de données, qui s'intéresse au partage de données entre des organisations privées et des organisations à but non lucratif pour favoriser l'émergence d'applications d'intérêt général.

Ce marché est au cœur de la stratégie européenne, avec des rapports prévoyant une augmentation du PIB de la zone pouvant atteindre les 14% cumulés d'ici à 2030,^[13] interrogeant ainsi la balance entre protection des données personnelles, protection des données de la zone UE, et modalités de partage des données.

LA PROTECTION DES DONNÉES

Concernant la protection des données, il existe des risques d'usages quant à la protection des droits fondamentaux des citoyens et de certaines valeurs comme la dignité humaine et l'autonomie personnelle ou la confidentialité et protection des données.^[14] Ce risque est à relativiser dans l'UE car la zone est en avance par rapport aux autres régions du monde quant à la protection des données et des citoyens, notamment grâce au **Règlement général sur la protection des données** (RGPD).

Le RGPD s'applique à toute forme de données personnelles, définies comme toutes les données qui, en elles-mêmes, ou lorsqu'elles sont combinées avec d'autres auxquelles le détenteur peut probablement accéder, peuvent identifier un individu. Les données doivent être corrigées ou effacées sur demande de son émetteur. Le texte s'applique à toute organisation basée ou non en Europe, qui collecte, stocke ou traite les données personnelles des résidents de l'UE. Ainsi, chaque pays doit produire ses propres réglementations et services pour surveiller l'application du règlement.

Les Etats-Unis n'ont pas de règlement de protection des données des citoyens unifié. Pourtant, le pays a adhéré au Cross-Border Privacy Rules, une série de règles adoptée par les pays de l'A.P.E.C. (Asia Pacific Economic Cooperation). Ce dernier établit des réglementations minimales sur les modalités et les finalités du recueil de données.

En Chine, le China's Cybersecurity Law couvre toutes les formes de données personnelles appartenant aux citoyens chinois. Le texte intègre une interdiction du stockage de données personnelles à l'étranger, sauf dans certains cas spécifiques avec une évaluation de sécurité. Paradoxalement, toutes ces données doivent être examinées par leurs détenteurs dans l'objectif de faire remonter aux autorités chinoises des informations sur de potentielles activités prohibées. Pris ensemble, ces deux dispositions illustrent une considération de la donnée par le gouvernement chinois, non en termes de protection de la vie privée mais bien comme une valeur stratégique en termes d'information.

UNE DISPARITÉ DES ACTEURS QUI SUPPOSE UNE DISPARITÉ DES APPLICATIONS

Dés lors, la présentation des bénéfices de ces dispositions par la commission européenne semble tout de même très optimiste. En effet, un rapport de l'European Court of Auditors a mis en avant que l'utilisation des technologies innovantes n'était que partielle dans les entreprises européennes.^[15]

Sur la base d'un indice incluant l'intégration des technologies digitales dans les sociétés européennes, leur étude classifiait en 2018 l'Union européenne à la 11ème place, derrière la Corée du Sud, le Canada ou les USA, avec une légère amélioration en 2020. De plus, on peut noter une très grande disparité entre les pays de l'UE. Toujours en 2020, la France se retrouve à la 16ème place parmi les pays européens, en dessous de la moyenne. Ces éléments démontrent le besoin d'investissements pour acculturer l'ensemble des sociétés européennes à l'utilisation des nouvelles technologies.

La disparité se trouve aussi dans la taille des entreprises : 54% des entreprises de grande taille sont hautement digitalisées contre 17% pour les PME avec des investissements décroissants.^[16] On peut aussi considérer les problématiques liées à la domination sur le marché du Cloud et la disponibilité des semi-conducteurs.^[17] Dès lors, si les avantages sont surestimés, les désavantages s'en trouvent, en termes relatifs, sous-estimés.

III. VERS UN SMART-GOUVERNEMENT

LA VALEUR STRATÉGIQUE DE LA DONNÉE DANS LE SECTEUR PUBLIC : LE CAS DE LA VALEUR PUBLIQUE

En sa qualité de ressource stratégique, la donnée est associée à une valeur financière. Selon les analyses prospectives, le développement des écosystèmes basés sur les données ouvertes pourrait résulter en une impulsion pour l'économie allant d'environ 1 à 1,5 % du PIB en 2030 dans l'Union européenne, le Royaume-Uni et les États-Unis, et jusqu'à 4 à 5 % en Inde.^[18] La commission européenne propose une estimation de la valeur de la donnée au niveau européen à 829 milliards d'euros en 2025, en comparaison des 301 milliards (2.4% du PIB de l'union européenne) en 2018.^[19] Cette prospective doit être considérée comme un indicateur. Cependant, la pertinence de concevoir la donnée uniquement dans sa valeur financière pourrait être aisément remise en

question notamment dans le cadre du secteur public. La capacité des administrations à utiliser les investissements liés à la production, au partage et à l'utilisation des données ne peut pas être uniquement évaluée en termes financiers, mais plutôt dans les retombées politiques et sociales bénéficiant au public:^[20] la création de valeur publique.



La valeur publique est définie comme la valeur qu'une organisation apporte à la société.^[21] Dans ce concept réside l'objectif de maximisation de la valeur pour le public. La valeur est donc définie par l'évaluation des citoyens de leurs administrations et services.^[22] Pour Moore, lorsqu'il s'agit de valeur publique, trois domaines de performance sont primordiaux : la prestation de services réels, l'obtention de

résultats sociaux et le maintien de la confiance et de la légitimité. Dans le cadre de la problématique de la donnée et de son utilisation, ces domaines relèvent de trois catégories d'utilisation de la donnée : la gouvernance anticipative, la conception et prestation, ainsi que la gestion des performances.^[23]

Ainsi, le secteur public étant un des secteurs produisant le plus de données, si cette donnée peut être partagée, elle semble surtout utile à améliorer les activités du secteur public. Un secteur public axé sur la gestion, le partage et l'utilisation des données transforme la conception, la mise en œuvre et le suivi des politiques et services publics. L'utilisation des données renforce la conscience qu'a le secteur public de lui-même et, ainsi, sa capacité à développer des politiques et des services adaptés à un environnement changeant. L'idée est de faciliter la création de valeur publique qui peut être intégrée dans un cycle.

La notion de cycle de valeur publique des données a été introduite dans un rapport de l'OCDE pour envisager les phases par lesquelles les données doivent passer pour optimiser le fonctionnement

des administrations et créer de la valeur publique .^[27] Si nous mentionnons ce modèle c'est parce qu'il s'accorde avec le modèle présenté précédemment par une logique itérative, mais aussi qu'il rajoute un élément qui ne fait pas partie de l'approche du modèle de cyclage : le stockage des données. Depuis la donnée brute jusqu'à l'utilisation, le modèle propose quatre phases :

- **La collecte et la génération avec des données publiées** (données ouvertes), des données de capteurs, des données demandées (formulaires) ou des données administratives
 - **Le stockage, la sécurisation et le traitement** avec les activités de gestion de la qualité, de catalogage et de nettoyage
 - **Le partage, la curation et la publication** sur la base de demandes et d'accords de partage, de plateformes de partage de données, de sites web permettant l'accès aux données
 - **L'utilisation et la réutilisation** avec les différents types d'analyses statistiques à des fins diagnostiques, descriptives, prédictives et prescriptives.
-

Les deux premières phases réfèrent à la gestion de la donnée dans le secteur public en termes de collecte, traitement et stockage des données. Les deux dernières font référence à la distribution et l'utilisation permettant le cyclage.

L'exploitation des données dans le secteur public doit être envisagée comme l'exploitation d'un actif stratégique permettant : ^(FIG 3)

- **Une meilleure évaluation et un suivi optimisé** pour mesurer l'efficacité des processus, leur qualité et la satisfaction des agents et des usagers (ex. reporting d'activité, profilage démographique).

- **Une meilleure anticipation dans la conception des politiques**, des évolutions possibles et de la prévision des besoins (ex. exploration de systèmes de causes-effets).
- **Une meilleure prestation des services** en utilisant la donnée pour améliorer la réactivité et l'individualisation automatique des prestations (ex. aide à la décision pour le développement de nouvelles infrastructures ou pour les besoins en investissement).^[24]

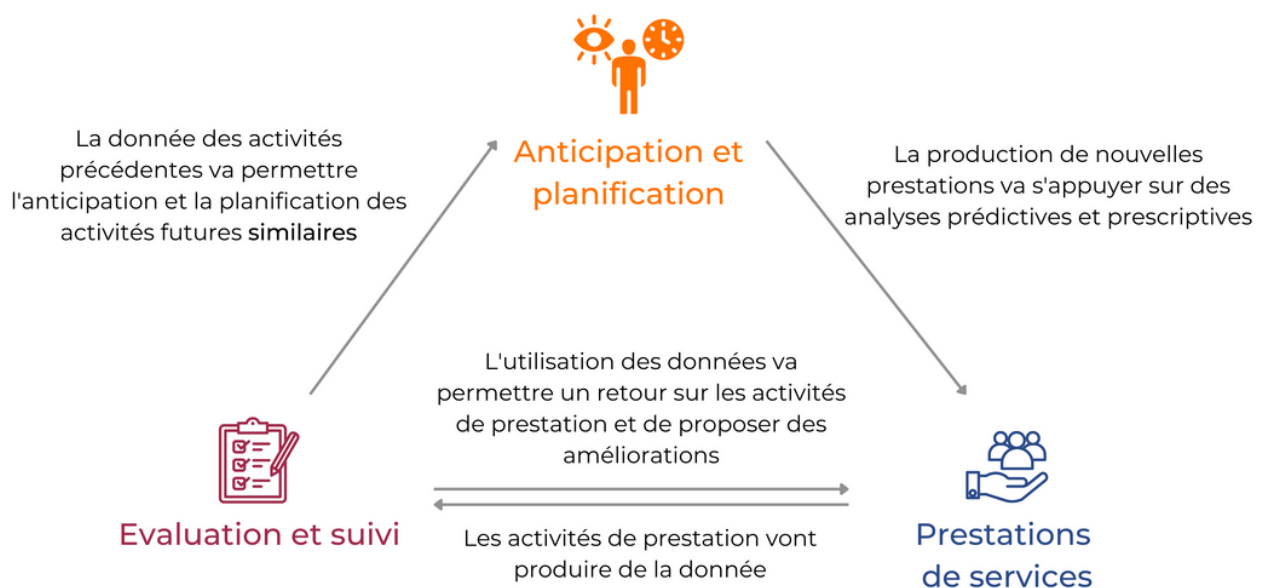


Figure 3. Cycle d'exploitation de la donnée comme information stratégique - ©Artimon Perspectives

C'est sur cette exploitation des Big Data qu'a été défini le concept de smart-gouvernement. Le smart-gouvernement a pour but de définir de nouveaux modèles de prestation de services en connectant et en intégrant des environnements physiques, numériques, publics et privés^[25 - 26].

Pour mieux comprendre le smart-gouvernement, on peut d'abord définir l'e-gouvernement. L'e-gouvernement se base sur la numérisation des processus utilisant les technologies de l'Internet comme levier. Comparativement, le smart-gouvernement intègre l'exploitation de la donnée au cœur de son processus de décision. Cela peut être effectif aux niveaux planification-opérationnel, stratégique-politique, ou aux deux. Les données peuvent provenir du gouvernement, de l'extérieur ou des deux. Cette approche va donc plus loin que les efforts de transformation digitale de l'e-gouvernement en interrogeant la relation entre l'administration publique et ses acteurs au travers des « smart » technologies d'information et de communication (TIC).

Dans cette perspective de smart-gouvernement la donnée joue donc un rôle central.

L'idée générale est de développer des méthodes de collecte de données de différentes sources, de transformation et de standardisation de ces données, d'utilisation de l'information extraite et ce de manière continue et cyclique (comme présenté précédemment). Considérant le gouvernement et des organisations du secteur public comme responsables de générer et partager la connaissance, ils ont dans ce système un rôle central et moteur pour promouvoir ce cyclage, mais aussi pour développer des méthodes d'intégration de l'exploitation de la donnée dans leurs habitus (ex. par l'utilisation d'algorithme d'aide à la décision ou à l'évaluation des processus).

A Singapour par exemple, l'exploitation des Big Data a permis d'améliorer la prestation de services en offrant aux Singapouriens des horaires et des services améliorés. Le Singpass est un service d'identité digitale qui permet l'accès et l'utilisation de 1400 services gouvernementaux et privés, de manière digitale ou physique. Le SingPass a pu réduire le temps de

déplacement et le temps de traitement des demandes de renouvellement des documents d'identité passant de 20 minutes à 10 secondes corollairement à une réduction des coûts de traitement et de la quantité de document non-numériques. Sur un autre plan, l'agence gouvernementale Housing and Development Board s'est aussi appuyée sur l'exploitation des Big Data pour mieux comprendre la composition démographique des quartiers, comme le niveau d'instruction et le revenu des ménages. Cela a permis d'identifier les vecteurs d'attraction par type de ménage, d'orienter les nouveaux projets de logement et les allocations de ressources en fonction de ces vecteurs.

En Australie, l'Australian Electoral Commission utilise la donnée de 16 millions d'australiens pour simuler et investiguer la fraude électorale, à travers la détection de la récurrence d'anomalies. Concernant la politique de transport, le projet Data61 du Commonwealth Scientific and Industrial Research Organisation a permis d'améliorer l'efficacité du système, en partenariat avec Transports for New South Wales. Basées sur la collecte de l'information du trafic en temps réel (ex. détecteurs de réseaux

routiers, base de donnée Opal, appareils GPS, feux de circulation et bus, données météo), ces organisations ont développé un outil de gestion de la congestion des routes. Le système modélise l'impact des changements sur les réseaux et réassigne automatiquement les trajets aux opérateurs de transport et aux usagers permettant d'optimiser le réseau en temps réel.

LES DIMENSIONS DU SMART-GOUVERNEMENT

Šiugždinienė et collègues ont proposé un cadre d'analyse du smart-gouvernement en opérationnalisant quatre dimensions clés et leurs sous-dimensions à travers des indicateurs empiriques : la dynamique stratégique, la collaboration intersectorielle, la collaboration interinstitutionnelle, et l'autonomie des citoyens.

L'ensemble de ces dimensions va à la fois représenter des pratiques politiques dont l'information (i.e. la donnée) et la digitalisation (i.e. la facilitation du flux de la donnée) sont des outils pour des processus de prise de décision plus efficaces et, de

manière générale, plus collaboratifs. Cela fait écho au modèle de cyclage de la donnée où le partage et l'utilisation catalysent la valeur stratégique de cette dernière. Sur la base du modèle proposé, il est ainsi possible de définir des critères d'évaluation pour chacune des dimensions présentées ci-dessous.

LA DIMENSION DE DYNAMIQUE STRATÉGIQUE

Dans le cadre du smart- (FIG 4) gouvernement, la **dynamique stratégique** représente la capacité d'un gouvernement à anticiper et s'adapter de

manière proactive à des événements internes ou externes afin d'éviter les crises (i.e. changements brutaux et non volontaires) avec une approche organisée et temporellement pertinente.

C'est ce qu'on peut qualifier de **perspicacité et sensibilité stratégique**.^[27 - 28] Vient ensuite la capacité à mobiliser les ressources internes et externes aux organisations de manière fluide et de les allouer de manière optimum pour conduire des projets de court, moyen ou long terme. Cela implique une part de **flexibilité organisationnelle** et d'autonomie.^[28 - 29]

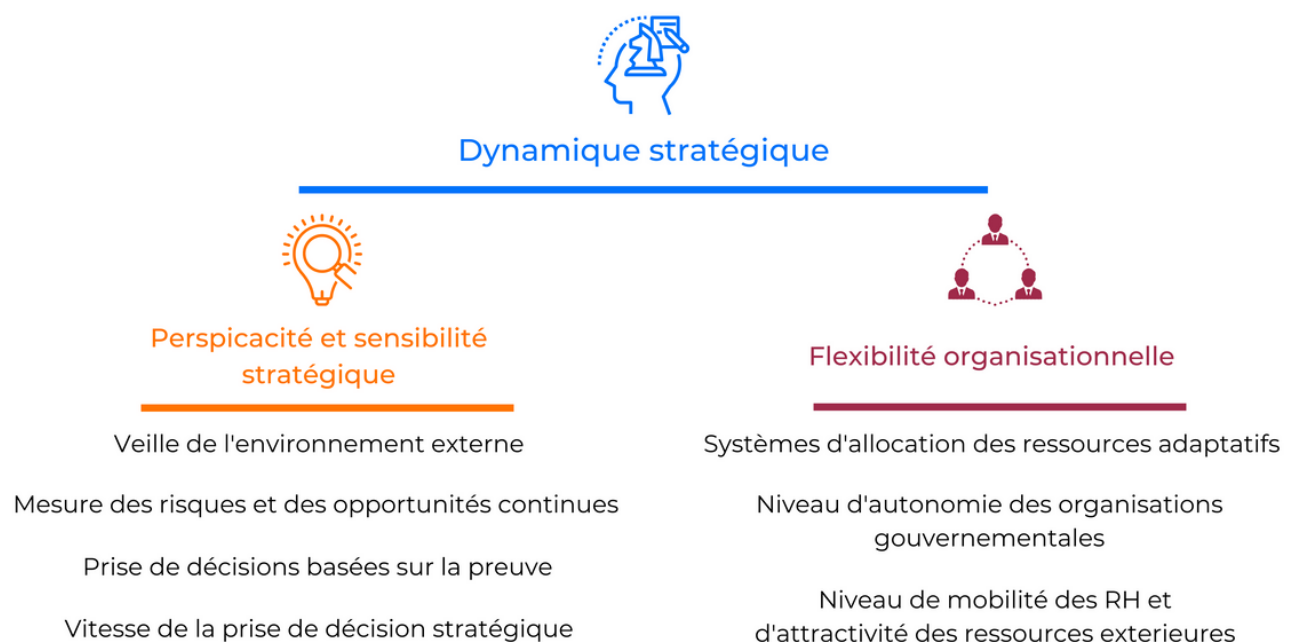


Figure 4. Dimensions stratégiques du smart-gouvernement -
©Artimon Perspectives

LA DIMENSION DE COLLABORATION INTERSECTORIELLE

La dimension de **collaboration intersectorielle** procède de l'interaction entre différents acteurs et secteurs qu'ils soient privés, publics, ^(FIG 5) gouvernementaux ou non. Le cœur de cette dimension est le partage d'information et de ressources pour des objectifs congruents bien que potentiellement différents. On peut identifier le type de gouvernance de ces écosystèmes collaboratifs comme un facteur important.

La gouvernance facilitante recouvre cette idée d'un besoin de gestionnaire de collaboration pour définir des règles, développer la confiance, le dialogue, guider les négociations et promouvoir ces objectifs congruents dans l'écosystème.

Ce développement de la confiance, de l'engagement et de la compréhension des problèmes passe par la mise en place de pratiques ^[30] collaboratives au travers de plateformes, mais aussi par la définition d'une interdépendance positive au travers, notamment, du partage des responsabilités. ^[31]



Figure 5. Dimension de collaboration intersectorielle du smart-gouvernement - ©Artimon Perspectives

LA DIMENSION DE COLLABORATION INTERINSTITUTIONNELLE

La collaboration, coordination et coopération entre les ministères, les départements, les strates gouvernementales est un prérequis du smart-gouvernement.^(FIG 6)

Dans le cadre du smart-gouvernement, cette tâche est dévolue aux institutions gouvernementales qui doivent développer des **plateformes d'interaction** pour permettre une **collaboration**

interinstitutionnelle

efficace,^[32] mais aussi pour promouvoir le développement de **compétences collaboratives**. Les compétences collaboratives incluent la capacité des institutions à comprendre les besoins et l'intérêt de la collaboration interinstitutionnelle, la disponibilité de formation pour le développement de ces compétences ou la disponibilité d'outils de partage d'informations (ex. systèmes d'information communs).



Figure 6. Dimension de collaboration interinstitutionnelle du smart-gouvernement - ©Artimon Perspectives

LA DIMENSION DE L'AUTONOMIE ET DE L'ENGAGEMENT DES CITOYENS

Cette dimension définit le modèle d'interaction entre le gouvernement et les citoyens, transitant de modalités d'information et de consultation à des modalités de coopération et de participation.^(FIG 7)

Ancrer la participation des citoyens dans les prises de

décision et permettre la production et la prise en compte des retours de ces derniers sont des caractéristiques du smart-gouvernement.

Ces dispositions visent à favoriser la transparence et la responsabilité des actions gouvernementales auprès des citoyens pour assurer que ce dernier, et les différentes administrations, définissent leurs actions par et pour la [32 - 33] création de valeur publique.

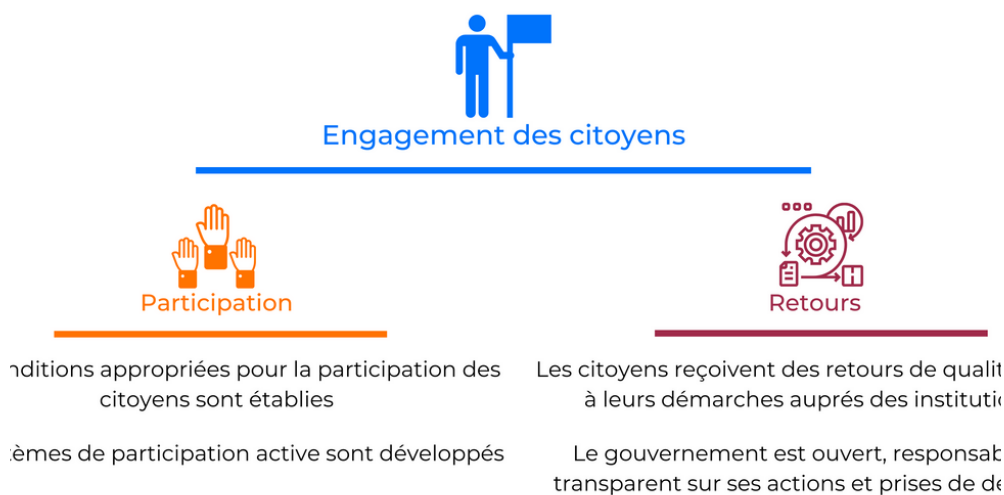


Figure 7. Dimension de l'engagement des citoyens du smart-gouvernement - ©Artimon Perspectives

LE DÉVELOPPEMENT DU SMART-GOUVERNEMENT

La question qui se pose maintenant concerne les modalités de développement d'un smart-gouvernement. Ici nous allons nous intéresser à la base sur laquelle le

développement doit se fonder en termes de collecte, de partage, de génération de la donnée et de la prise de décision associée.

LE CADRE NORMATIF

On peut en premier lieu considérer trois axiomes pour fixer notre base^[23] :

- La gestion de la donnée doit être légalement acceptable
- La gestion de la donnée doit être éthiquement acceptable
- La gestion de la donnée doit être socialement acceptable

Pour permettre le développement d'un smart-gouvernement, la gestion de la donnée doit d'abord répondre à des normes légales. Au niveau de la collecte, de la génération et de l'utilisation de la donnée un cadre réglementaire définit par exemple les modalités de rétention. En ce qui concerne le partage, des lois existent sur les modalités d'ouverture de la donnée, la transparence ou la propriété. En termes d'utilisation et de réutilisation de la donnée, l'accessibilité des données d'autres organisations est soumise à des conditions réglementaires. Enfin dans les processus de prise de décision on peut noter des nécessités d'interprétabilité et de transparence quant à des décisions ou des aides à la décision algorithmiques par exemple.

En parallèle du cadre légal se tient le cadre éthique. Il existe des normes concernant la gestion des données des citoyens et leur transférabilité. La donnée est supposée ouverte par défaut et l'utilisation de la donnée se définit dans des utilisations dites « positives ». En d'autres termes, on suppose l'utilisation éthique de la donnée à la fois en termes de pratiques mais aussi d'objectifs. Enfin, dans le processus de prise de décision, celui-ci doit être identifiable, s'adjoindre d'une responsabilité par un acteur, même (et surtout) si cette décision a été prise automatiquement.

Enfin, dans le cadre des normes, l'acceptabilité sociale du modèle de gestion de la donnée est un prérequis au smart-gouvernement. Cela passe par la protection de la vie privée quant à la collecte, la génération ou l'utilisation de la donnée, mais aussi par une participation des citoyens dans le débat autour des pratiques. Le partage doit être évidemment consenti au-delà du cadre légal : légalité et légitimité ne doivent pas être confondues. Pour ce qui est du smart-gouvernement, la génération d'information et l'utilisation de la donnée ne

doivent pas s'appuyer sur une sollicitation au partage permanent de données par les citoyens mais plutôt dans une démarche de simplification du partage d'information auprès des services. Cette simplification ne sous-tend pas une perte de contrôle des citoyens sur leurs données. Enfin, par principe, toute décision doit être expliquée et cette explication doit être communiquée au citoyen. C'est une mise en application des principes éthiques et légaux autour de la prise de décision dans un schéma socialement acceptable : la communication de l'information par défaut.

LE CADRE D'IMPLÉMENTATION

Une fois le cadre normatif identifié, l'implémentation des principes fonctionnels d'un smart gouvernement se définit sur 4 dimensions :

- La gouvernance
- La technologie
- La culture
- Les compétences

Toutes ces dimensions sont importantes. Une implémentation pérenne suppose le développement de chacune d'entre elles à partir du cadre normatif défini précédemment.

Au niveau de la gouvernance, on peut en premier lieu appuyer l'implémentation sur la présence d'un chief data officer ou directeur des données dont la responsabilité sera d'évaluer l'état des données (ex. en prenant en compte leur véracité ou variété) mais aussi l'application du critère d'« open by default » promu par l'OCDE^[34 - 35]. Ensuite, le partage et l'utilisation doivent s'appuyer sur une stratégie « quelle donnée pour quel usage ? » et donc de définir des guides (et des équipes support) pour la gestion des données personnelles et la définition de standards en termes de format et de métadonnées pour favoriser l'interopérabilité^[34]. Enfin, les méthodes de prise de décision sur la base d'analyse de données doivent être cultivées. Cela suppose l'émergence d'une catégorie de « preneurs de décision guidés par la donnée ». En effet, comme nous l'avons mentionné, les biais et les limites des algorithmes et des bases de données doivent pouvoir être identifiés afin de pondérer le suivi d'une information à des critères préalables à leur production. En d'autres termes, il est nécessaire d'inclure dans les processus de décision des personnes capables d'évaluer la

valeur stratégique d'une information.

Pour ce qui est de la technologie, différentes solutions vont permettre de répondre à la fois aux critères normatifs de légalité et d'acceptabilité éthique et sociale. On peut citer des solutions de cloud (ex. européens) pour le stockage, des plateformes d'interopérabilité, l'identification des sources fiables, le développement d'algorithmes et d'API de traitement ou de visualisation. Cela passe également par la mise en conformité par rapport aux critères de transparence sur les performances des outils.

Au niveau de la culture, un besoin de partage, d'explication et de transparence est nécessaire pour permettre à l'ensemble des organisations publiques de s'approprier ce nouveau processus de prise de décision. En effet, le processus d'acculturation suppose de modifier les représentations, les normes, valeurs et croyances d'individus avec des risques d'anticipations négatives quant à l'utilisation de la technologie, notamment si ces dernières sont mal comprises et/ou mal présentées et expliquées. Aussi l'introduction de ces pratiques et outils doit être contextualisée. Chaque

service peut être considéré comme une entité à part entière dont la culture, les attitudes et les comportements vont diverger. Dès lors, comme tout projet de transformation, supposer une recette universelle reviendrait à mettre en échec le processus d'acculturation et d'implémentation.

Finalement, au niveau des compétences, faisant écho à la gouvernance, la technologie et la culture, le développement s'appuie sur des formations et des recrutements permettant d'orienter la stratégie d'implémentation d'un smart-gouvernement. De plus, ces compétences sont requises pour l'ensemble des tâches de gouvernance (ex. chief data officer), du développement des outils (ex. data scientist, computer scientist) ou de l'acculturation (ex. sociologue, psychologue). Considérer que les seules compétences nécessaires sont techniques serait une erreur.

L'intégration d'une technologie se fait dans le cadre de comportements humains, il est donc nécessaire de comprendre à la fois la technologie et le comportement humain pour favoriser la réussite de l'implémentation de ces nouveaux outils.

IV. LA NÉCESSITÉ DE FOURNIR UN CADRE D'ÉVALUATION DE LA MATURITÉ DE LA GESTION DES DONNÉES

UN MODÈLE DE MATURITÉ DU SMART GOVERNEMENT

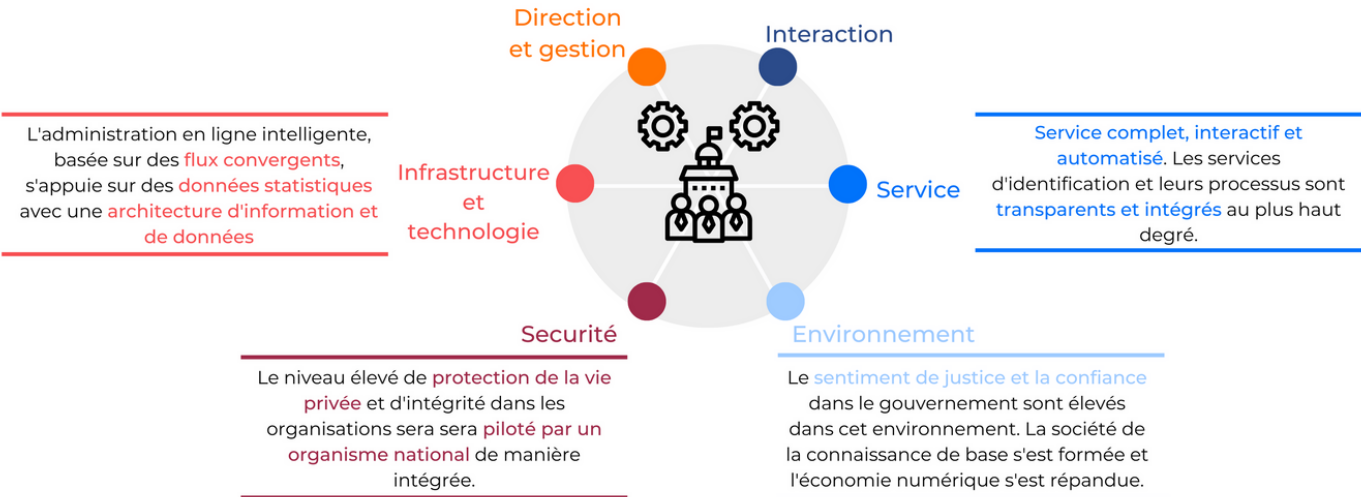
Suivant une revue de la littérature, des chercheurs ont proposé un modèle en 6 composantes^[36] pour définir la maturité d'un gouvernement comme « smart government »^(FIG 8). La complexité des gouvernements modernes et la nécessité d'une gestion et d'une gouvernance efficaces sont devenues de plus en plus

évidentes à l'ère des technologies avancées. Afin de gérer correctement et d'adopter des politiques de prise de risque, il est important d'avoir une perspective qui va au-delà de la technologie. C'est là que ce modèle multidimensionnel de gouvernement intelligent a été proposé. Il vise à couvrir tous les aspects du gouvernement intelligent. Le modèle peut être utilisé pour décrire le concept de la maturité pour chaque dimension.

Disposer de **structures en réseau** et de des réponses à multiples facettes dans un **écosystème transparent** avec productivité maximale, avec une gestion intelligente, une grande **flexibilité** et l'innovation, une approche **centrée sur le citoyen**

Interopérabilité organisationnelle et technique élevée, **coopération organisationnelle** à 360 degrés, dialogue global entre le gouvernement, le citoyen et les administrations, **coordination des processus**

Figure 8. Modèle de maturité du Smart Government - ©Artimon Perspectives



L'EXEMPLE DU DATA MATURITY ASSESSMENT FRAMEWORK AU ROYAUME-UNI

Le cadre d'auto-évaluation de la maturité des données (DMA) pour le gouvernement est une méthode permettant aux organisations du secteur public de comprendre leurs capacités, leur efficacité et leur état de préparation à l'utilisation des données.^[37] Le DMA contient six thèmes pertinents pour le contexte gouvernemental et les différentes organisations du secteur public.

Chaque ligne de l'évaluation décrit les caractéristiques ou les comportements associés à la progression d'une maturité faible à élevée pour chaque sujet et thème. L'évaluation de la maturité ne donne pas un score unique mais fournit une vue matricielle de la maturité à travers les sujets et les thèmes. L'examen des résultats permet de hiérarchiser les ressources sur la base d'éléments concrets et encourage une approche équilibrée du développement et du progrès.



Figure 9. Niveau de maturité dans la gestion de la donnée du Data Maturity Assessment Framework - ©Artimon Perspectives

Les thèmes inclus dans le DMA permettent d'examiner le niveau de maturité actuel d'une ligne ou d'un sujet et sont utiles pour comprendre les moteurs d'une maturité élevée ou faible. Les thèmes comprennent :

- l'engagement avec les autres,
- les compétences et connaissances en matière de données,
- les systèmes adéquats,
- la connaissance des données, la prise de décisions grâce aux données, ^(FIG 10)
- la gestion et l'utilisation éthiques des données,
- la gestion des données produites,
- la protection des données produites,
- la définition de l'orientation en matière de données
- la prise de responsabilité à l'égard des données.

En utilisant le DMA, les organisations peuvent identifier les forces et les faiblesses de leur écosystème de données, et ainsi résoudre les domaines de faible maturité des données qui présentent un risque pour la réalisation des priorités de l'organisation, tandis que les domaines de bonne ou de haute maturité qui sont essentiels à la réussite de la réalisation peuvent être maintenus.

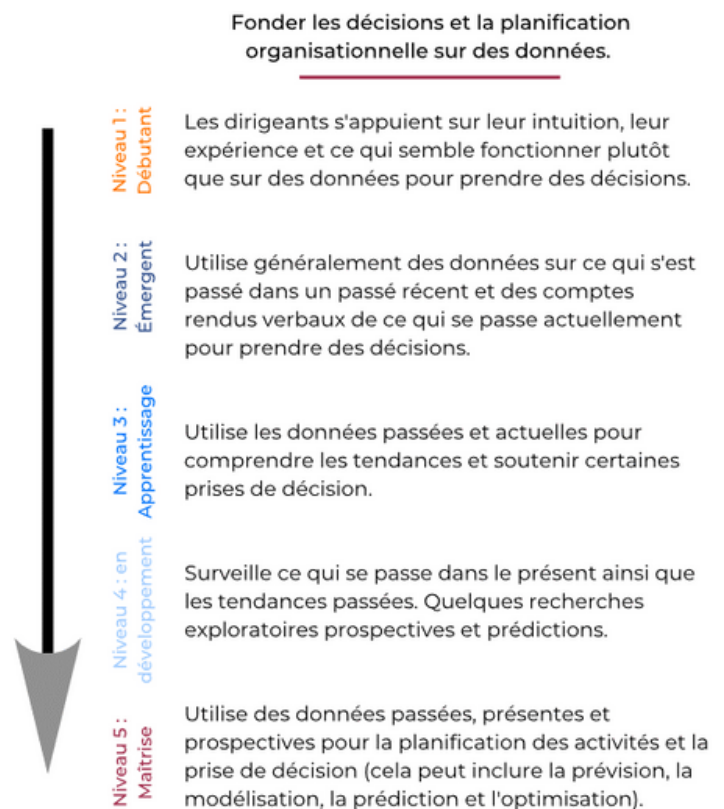


Figure 10. Exemple de développement par niveau dans le thème « prise de décisions à l'aide de données » sur la dimension « Fonder les décisions et la planification organisationnelle sur des données. » - ©Artimon Perspectives

CONCLUSION

Dans un monde de plus en plus complexe et dans lequel les informations s'échangent à des vitesses intangibles pour l'homme, l'utilisation de la donnée dans les prises de décisions humaines est devenu une nécessité pour son adaptation. Le Big Data fait office de concept recoupant ces évolutions de flux et leur utilisation potentielle.

Comprendre la donnée en tant qu'objet, et comment de l'information stratégique peut être produite à partir de celle-ci, est primordial pour les preneurs de décision.

En effet, les modèles de prise de décision 2.0, c'est-à-dire hautement assistés par les technologies de l'information et de la communication, changent radicalement les temps de réaction, la compréhension de l'environnement et des enjeux nationaux et internationaux, l'analyse des options d'action ou les conséquences de ces dernières. C'est dans ce cadre qu'émergent les smart-gouvernements, avec l'idée que les enjeux des années à venir pour nos sociétés doivent être évalués dans des systèmes complexes dynamiques (ex. le système écologique) qu'il est difficile à l'esprit humain de manipuler.

Dès lors, l'utilisation de données massives associées à d'autres données massives pour produire de l'information accessible à la cognition humaine, devient un outil pour les décideurs, leur permettant de restaurer une possibilité d'action sur l'environnement.

En cela, la donnée est considérée comme une valeur stratégique et, dans le cadre des gouvernements, permettant la création de valeur publique. Cela passe par un accroissement des capacités d'adaptation des stratégies gouvernementales, l'augmentation des capacités d'interaction intersectorielle et interinstitutionnelle, ainsi que par une nouvelle approche de la prise de décision à travers le développement d'outils de participation citoyenne. L'émergence d'un smart-gouvernement est un processus au long-cours qui a déjà commencé. Une question se pose néanmoins : le smart-gouvernement est-il un concept viable pour prendre de meilleures décisions politiques fondées sur le réel ou sera-t-il l'instrument d'une abstraction et technocratisation de ces décisions plus profondes ?

RÉFÉRENCES

- [1] M. N. I. Sarker, M. Wu, and M. A. Hossin, "Smart governance through bigdata: Digital transformation of public agencies," 2018 International Conference on Artificial Intelligence and Big Data, ICAIBD 2018.
- [2] Y. Demchenko, P. Grosso, C. de Laat, and P. Membrey, "Addressing big data issues in Scientific Data Infrastructure," Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013.
- [3] D. Moody and P. Walsh, "Measuring the Value Of Information - An Asset Valuation Approach," in European Conference on Information Systems (ECIS'99), Frederiksberg, Denmark, 1999.
- [4] B. Azevedo, R. Bastardo, R. Tortosa, and J. Bártolo, "Information Visualization: An analysis On The Data Glut And The Emergency To Rethink And Design New Communicative Paradigms," in 9th Conference of the International Committee for Design History and Design Studies, Editora Edgard Blucher, 2014.
- [5] C. Santiso and B. Roseth, "Data Disrupts Corruption," Stanford Social Innovation Review, 2017
- [6] S. M. Tam and J. K. Kim, "Big Data ethics and selection-bias: An official statistician's perspective," Stat J IAOS, 2018
- [7] International Data Corporation, "Worldwide Global DataSphere and Global StorageSphere Structured and Unstructured Data Forecast, 2021–2025," 2021.
- [8] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," Science (1979), 2020.
- [9] L. Lechardoy, A. Sokolyanskaya, and F. Lupianez-Villanueva, "Analytical paper on the structure of the online platform economy post COVID-19 outbreak," 2021.
- [10] J. Chen and J. Shao, "Nearest Neighbor Imputation for Survey Data," Journal of Official Statistics, 2000.
- [11] OECD, "The OECD Privacy Framework," 2013.
- [12] "European data strategy | European Commission", 2022.
- [13] European Parliamentary Research Service, "Digital transformation," 2022.
-

[14] AIDA, "AIDA Working Paper on AI and Competitiveness", 2022.

[15] European Court of Auditors, "Digitising European Industry: an ambitious initiative whose success depends on the continued commitment of the EU, governments and businesses. Special report", 2020.

[16] European Investment Bank, "Digitalisation in Europe 2020-2021: evidence from the EIB investment survey", 2021.

[17] E. Young, "Study on the economic detriment to small and medium-sized enterprises arising from unfair and unbalanced cloud computing contracts," 2018.

[18] O. White et al., "Financial data unbound: The value of open data for individuals and institutions," 2021.

[19] "European data strategy | European Commission."

[20] A. M. Cresswell, G. B. Burke, and T. A. Pardo, "Advancing Return on Investment Analysis for Government IT A Public Value Framework," Albany, NY, 2006.

[21] M. H. Moore, *Creating Public Value Strategic Management in Government*. Cambridge: Harvard University Press, 1997.

[22] C. Talbot, "Paradoxes and prospects of 'public value'", 2011.

[23] C. van Ooijen, B. Ubaldi, and B. Welby, "A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance," 2019.

[24] K. C. Desouza and B. Jacob, "Big Data in the Public Sector: Lessons for Practitioners and Scholars", 2014.

[25] K. Schedler, A. A. Guenduez, and R. Frischknecht, "How smart can government be? Exploring barriers to the adoption of smart government," Information Polity, 2019.

[26] "Smart Governance: A Roadmap for Research and Practice," 2014

[27] J. Šiugždinienė, E. Gaulé, and R. Rauleckas, "In search of smart public governance: the case of Lithuania," *International Review of Administrative Sciences*, 2019.

[28] Y. Doz and M. Kosonen, for the Future: *Building the Strategic and Agile State | Semantic Scholar*. Sintra Studies Publications, 2014.

[29] J. Siugzdiniene and I. Kirstukaite, "Dealing with Wicked Problems: Government Priority System as a Tool for Engaging Leadership and Focusing on Changes," Soc Sci, 2015.

[30] C. Ansell and A. Gash, "Collaborative Governance in Theory and Practice," Journal of Public Administration Research and Theory, 2008.

[31] M. Bartenberger and V. Grubmüller-Régent, "The Enabling Effects of Open Government Data on Collaborative Governance in Smart City Contexts," JeDEM - eJournal of eDemocracy and Open Government, 2014.

[32] D. Reforgiato Recupero et al., "An Innovative, Open, Interoperable Citizen Engagement Cloud Platform for Smart Government and Users' Interaction," Journal of the Knowledge Economy 2016

[33] J. Torfing and P. Triantafyllou, "What's in a name? grasping new public governance as a political-administrative system," International Review of Public Administration, 2013.

[34] OCDE, "The OECD 2019 Open Useful Reusable Data (OURdata) Index: 2020".

[35] OCDE, "The OECD Framework for digital talent and skills in the public sector | Documents de travail de l'OCDE sur la gouvernance publique," 2021.

[36] M. R. Zynoddini et al., "The Smart Government Maturity Model," 2023.

[37] UK Central Digital and Data Office, "Data Maturity Assessment for Government," 2023.

Artimon Perspectives est l'institut de Recherche et Innovation d'Artimon, créé en 2018 et dédié à l'étude de l'impact des nouvelles technologies sur les organisations. Le travail de l'Institut Perspectives est pluridisciplinaire et transversal, mobilisant le corpus des sciences humaines et sociales, ainsi que les observations d'experts métier.

Cette approche vise à apporter des éclairages scientifiques à l'intégration des nouveaux outils technologiques, à l'observation et l'analyse des impacts des technologies sur les activités humaines. Pour ce faire, la technologie est étudiée de manière systémique, considérant les questions cognitives, sociales, environnementales, organisationnelles voire philosophiques.

CONTACTEZ-NOUS

8, rue de la Victoire,
75009 Paris
+33 (0)1 53 20 89 89
<https://artimon.fr/>

Josefina GIMENEZ

Directrice Recherche et
Innovation
Artimon Perspectives
jgimenez@artimon.fr

SUIVEZ-NOUS SUR



Artimon



Artimon Perspectives



@ArtimonConseil



@artimon5307