

Collecte et gestion des données numériques pour le pilotage des politiques publiques

Vers un big data territorial

Novembre 2016

TACTIS



PARME
AVOCATS

Tél 01 49 57 05 05 43 rue des meuniers,
www.tactis.fr 94300 VINCENNES

Tél 01 56 33 12 12
www.parme-avocats.com

12 Boulevard de Courcelles,
75017 Paris

Sommaire :

I.	Introduction	3
II.	Etat des lieux des politiques publiques autour de la donnée	6
A.	Contexte général : le marché de la donnée et du big data	6
B.	Un contexte législatif et réglementaire plus offensif sur la donnée	20
C.	Secteur public d'Etat : de l'open data à la gouvernance par la donnée	31
D.	Collectivités : début de structuration sur la gestion des données	41
E.	Freins et facteurs clés de succès identifiés	60
III.	Vers une nouvelle gouvernance de la donnée	64
A.	Recouvrer plus de souveraineté sur la donnée pour dynamiser l'économie	64
B.	Politique à mettre en œuvre pour atteindre les objectifs à 5 ans	65
IV.	Recommandations	82
A.	Poursuivre la dynamique de la loi pour une République numérique	82
B.	Affirmer, dans les contrats, la maîtrise des collectivités sur leurs données	83
C.	Préparer la création de missions locales de service public de la donnée	86
D.	Créer un maillage national de structures de mutualisation informatiques	87
E.	Mettre les data sciences au service des politiques publiques locales	87
F.	Structurer nationalement tous les acteurs de la data	88
G.	Créer un observatoire des données publiques	89
H.	Renforcer la coopération au niveau européen	89
I.	Participer à la réflexion sur la protection des infrastructures sensibles	90
V.	Annexes	91
A.	Fiches notions	92
B.	Résultats de l'enquête flash	109
C.	Fiches collectivités	114
D.	Fiches entreprises	125
E.	Analyses juridiques	129
F.	Personnes et entités interrogées	158

I. Introduction

Dans une économie mondialisée, graduellement portée par le numérique, la donnée devient un élément de valeur ajoutée et de richesse, autant par les volumes traités que par la diversité des données accessibles. Les flux d'échanges et de communication en ligne atteignent des niveaux qui étaient encore inconcevables il y a une décennie. Et ils poursuivent une spectaculaire progression puisque le nombre de messages échangés quotidiennement sur la planète entre 2013 et 2015 est passé de 100 milliards à 150 milliards...

Cette révolution des modes de communication et des échanges transactionnels confère à la donnée une position centrale. Les traces laissées par des dizaines de millions d'utilisateurs, en consultant un site web ou en effectuant leurs achats en ligne, sont désormais captées et exploitées par des entreprises capables d'en tirer une information comportementale, commercialement utile. Cette faculté de lecture massive des besoins des utilisateurs, pour fournir un service adapté, a d'ailleurs donné lieu à des innovations de rupture au succès foudroyant, entraînant le déclin d'activités de services considérées jusque-là stables et prospères.

"Google en sait plus que l'INSEE sur la France" regrettaient il y a peu deux chercheurs, pour stigmatiser la transformation rapide des marchés et l'arrivée de nouveaux acteurs entrants qui les contrôlent. Une prise de position qui visait surtout à mettre en évidence la perte croissante de souveraineté de la France sur ses données, la privant du même coup d'une capacité d'exploitation donnant des avantages compétitifs.

Les sciences de l'analyse permettent en effet de réaliser des prouesses. Elles font apparaître de nouvelles tendances, jusque-là invisibles, ou prédisent avec une exactitude croissante les comportements des usagers, grâce aux traitements du "big data". A l'instar de Google ou d'Amazon, qui mettent de nombreux services à la disposition des internautes (moteurs de recherche et navigateur, services d'achats et de consultations d'objets culturels...) quelques entreprises numériques sont désormais en capacité d'exploiter à l'échelle mondiale les données générées par les internautes et de les convertir en atouts concurrentiels. On ne cite plus les places de marché comme eBay, les services à vocation touristique comme Booking, TripAdvisor, Airbnb ou les plateformes autour de la mobilité comme Uber... Avec une rapidité inédite, elles supplantent des activités anciennes incapables de s'adapter aux nouvelles attentes des consommateurs en raison de leur fragmentation.

Notre pays peine à faire émerger de tels champions numériques au point de voir certains analystes pointer un retard qu'ils considèrent déjà irréversible. Faut-il les suivre dans cette forme de déclinisme ? Pas si sûr, car la conjonction en France du déploiement de l'Internet très haut débit (THD), de l'explosion des objets connectés et de la croissance exponentielle des données produites par les administrations et les services urbains, pourrait bien représenter à son tour un nouvel "eldorado".

Les entreprises privées ont été les premières en France à amorcer le virage du big data, avec l'utilisation progressive, dans le courant des années 2000, d'outils de traitement massif des données. Le mouvement a d'abord impacté les télécoms, la grande distribution et la finance, puis il s'est étendu à d'autres secteurs plus inédits pour porter aujourd'hui aussi bien sur la maintenance prédictive, la détection temps réel des dysfonctionnements, que sur l'analyse des facteurs de risque de maladies par les laboratoires pharmaceutiques ou encore l'élaboration de stratégies marketing basées sur l'analyse comportementale des consommateurs...

L'usage, plus récent, de ces mêmes outils au sein des administrations publiques, couvre des finalités plus axées sur la performance ou sur la gestion de la relation citoyen. Ces dernières commencent à s'intéresser aux data sciences pour optimiser leurs dépenses, lutter plus efficacement contre la fraude, prédire des risques d'incendies ou de vols de biens, pratiquer l'analyse comportementale pour mieux

gérer les heures d'affluence sur l'espace public... L'objectif étant de fournir des services efficaces, personnalisés et plus proches des gens.

Ces champs d'action semblent prometteurs. Ils supposent toutefois une organisation et une structuration qui n'existe pas encore à grande échelle. La transition vers ces formes nouvelles de gouvernance de la donnée supposerait par exemple une association plus étroite entre les grandes entreprises d'utilités, dotées de systèmes d'information centrés sur la donnée, et les administrations elles-mêmes engagées sur l'exploitation de leurs actifs numériques. L'enjeu pour les territoires étant de favoriser la création d'écosystèmes administratifs et entrepreneuriaux capables d'exploiter la valeur ajoutée des données mises à leur disposition via des plateformes de services.

La loi pour une République Numérique récemment promulguée, va placer l'open data au rang des principaux leviers de transformation des politiques publiques de la donnée, tant au niveau de l'Etat que des collectivités. Certes la réforme ne couvre pas la totalité des données valorisables produites par les administrations – certaines informations, comme les données personnelles n'ayant pas vocation à être diffusées sans contraintes– mais elle devrait déjà contribuer largement au décroisement et à une meilleure circulation de l'information.

L'Etat joue un rôle moteur dans la structuration de ces politiques. Il impulse par exemple des formes d'action publique plus orientées sur la collaboration et l'échange de données entre administrations, avec la mise en place de programmes stratégiques comme "Dites-le nous une fois" pour la simplification ou comme France Connect, pour la gestion des identités numériques. Il poursuit dans le même esprit une stratégie d'ouverture de bases de données nationales et de référentiels incontournables pour les acteurs publics et privés, comme la base adresse nationale (BAN) ou le fichier SIRENE dédié aux données sur les entreprises françaises. La mission « Etalab », créée par décret le 21 février 2011, est chargée de porter cette politique ambitieuse, notamment à travers la diffusion des données publiques, via un portail national.

Par ailleurs, à côté des mises à disposition, l'Etat est engagé dans le déploiement d'une stratégie de plateformes afin de mettre en réseau les grandes bases de données détenues par les administrations. Il pose ainsi les fondamentaux d'une nouvelle architecture numérique fondée sur une interopérabilité fondée sur les interfaces de programmation (API)... La mise en place de ces dispositifs a pour vocation de faciliter la réutilisation des données, donc de simplifier et de stimuler la capacité d'innovation et de développement de services, prenant plus fortement appui sur les approches utilisateurs.

Les grandes collectivités sont aussi engagées dans cette transition. Productrices et gestionnaires de données, elles investissent dans des outils d'analyse : tableaux de bords décisionnels, outils de reporting, systèmes de monitoring urbain, plateformes de données ouvertes... Les réformes nationales sur la coopération intercommunale et l'open data pourraient même accélérer la fusion des systèmes d'information "ville-centre/EPCI" et leur recentrage sur la réutilisation, le contrôle et la diffusion des données publiques.

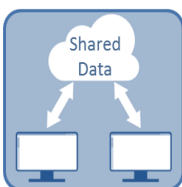
Toutefois, la mise en place de tels dispositifs est encore largement tributaire de la taille des entités publiques. En dehors des grands ensembles urbains, le bloc communal ne dispose pour l'instant, ni des capacités d'action ni des moyens pour financer de telles "info structures". La législation sur l'open data va toutefois les y conduire dans des délais relativement brefs. Aussi, pour accompagner cette évolution et éviter qu'elle ne soit trop lourde à supporter, une clarification du cadre de gouvernance local, de la problématique des échanges inter-administrations et des relations Etat/collectivités devra progressivement s'imposer afin d'organiser aux différentes échelles, un service de la donnée reposant sur la collecte, la diffusion et l'exploitation mutualisée des données.

Ce projet stratégique - du point de vue de la souveraineté et de la valeur ajoutée qu'il représente - reste à construire et à organiser à partir des territoires. Il constitue le principal objet de ce rapport.



Big Data ou Données Massives ou Mégadonnées

Des ensembles de données volumineux qui deviennent difficiles à traiter avec des outils classiques de gestion de base de données ou de gestion de l'information.



Shared Data ou Données partagées

Données échangées de manière restrictive entre organisation ou individus.

- Ex : échanges de données métiers entre différentes administrations ou de données nécessitant un niveau de sécurisation et de confidentialité élevés.



Open Data ou Données ouvertes

Définit les données structurées, sous licence libre, accessibles à tous. L'objectif est de favoriser la réutilisation des données pour générer de nouveaux usages.

II. Etat des lieux des politiques publiques autour de la donnée

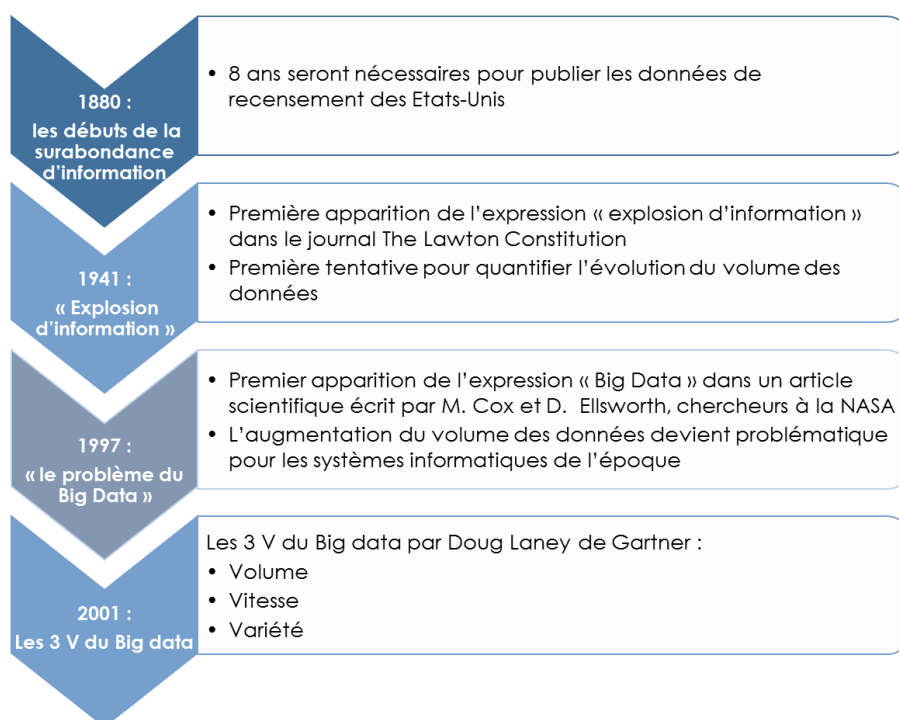
A. Contexte général : le marché de la donnée et du big data

1. Révolution numérique : révolution par la donnée

➤ *Explosion du volume et de la diversité des données*

La production de données connaît une croissance exponentielle qui se mesure aujourd'hui en exaoctets (10^{18} octets) et se comptera demain en zettaoctets (10^{21} octets). Dans le monde, 23 exaoctets de données sont produits chaque semaine et plus de 90 % du stock existant a été constitué au cours des deux dernières années. Cette croissance va se poursuivre et s'accélérer : une étude du cabinet IDC pour EMC¹, annonce un volume de données annuel multiplié par 10 d'ici 2020 atteignant ainsi 5,5 zettaoctets.

- 1 octet = 1 caractère
- 1 Mégaoctet = 10^6 octets = la capacité d'une disquette 3,5 pouces
- 1 Gigaoctet = 10^9 octets = 1 film en qualité TV
- 1 Pétaoctet = 10^{15} octets
- 1 Exaoctet = 10^{18} octets
- 1 Zettaoctet = 10^{21} octets



« 90 % du stock existant de données a été produit au cours des deux dernières années »

¹ Source : étude IDC pour EMC – Digital Universe, 2014
<http://france.emc.com/leadership/digital-universe/2014/view/executive-summary.htm>

Souvent qualifié de « data déluge », ce phénomène s'explique par des effets conjugués, et notamment :

- La mise à disposition généralisée d'outils numériques de plus en plus puissants et connectés (comme les ordinateurs et les smartphones),
- L'accès instantané à l'information par le plus grand nombre, notamment à travers les réseaux sociaux,
- Le mouvement de libération des données par les administrations engagées dans des démarches open data et la multiplication des objets connectés.

La nature des données se diversifie. Au-delà de la quantité, les données sont également de plus en plus variées : on parle aujourd'hui de données spatiales (vectorielles) et de données alphanumériques spatialement référencées dans les systèmes d'information géographiques, de données structurées et de données non structurées (cf. *fiche notion en annexe*), de données issues des réseaux sociaux et de vidéos, de données transmises par des capteurs, sans oublier les données de flux qu'elles soient en temps réel ou en temps différé (cf. *fiche notion en annexe*).

Face à cette prolifération, **la question de la qualité, de la fiabilité et de la valeur des données est posée** : sont-elles propres à l'utilisation et pour quels usages ? La présence de doublons, de données périmées ou erronées introduit des biais. Aussi, pour prévenir ce risque et extraire les données fiables et utiles, des outils spécifiques doivent être mis en place : on passe alors du big data au smart data. Mais le chemin pour exploiter les données pertinentes, est encore long : selon IDC², en 2013 seules 22% des données numériques étaient potentiellement exploitables – à conditions d'être correctement taguées et analysées. Dans la réalité, le cabinet estime que pas plus de 5% des données sont exploitées.

Les données non structurées vont croître cinq fois plus vite que les données structurées.

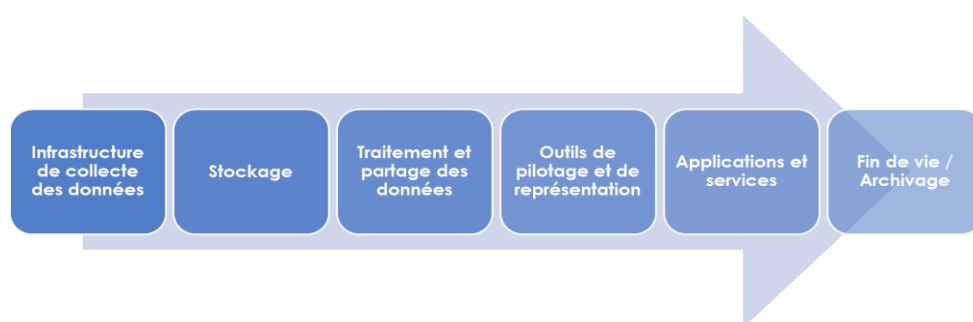
Aujourd'hui :

- 51% des données sont structurées,
- 27% sont non structurées
- 21% sont semi-structurées,



DANS : GUIDE DU BIG DATA 2015-2016 (SOURCE : IDC)

➤ De nouveaux enjeux sur le cycle de vie de la donnée



L'exposition du volume des données n'aurait aucune valeur sans innovations technologiques pour produire et gérer ensuite les flux. L'ensemble de la chaîne de valeur de la donnée se transforme avec la mutation des équipements et l'apparition d'outils et de logiciels dédiés au traitement de flux massifs de données. Utilisés aux différents stades de cette chaîne, ils augmentent la performance et la production de données pertinentes qu'il conviendra ensuite d'exploiter au mieux.

² Source : étude IDC pour EMC – Digital Universe, 2014
<http://france.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

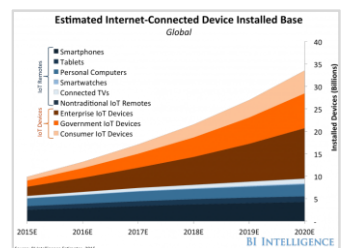
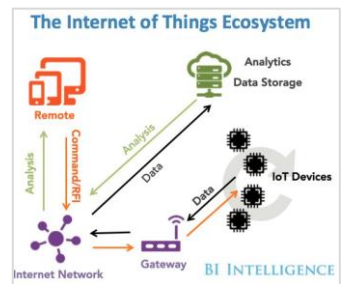
Collecte des données : la **prolifération des capteurs**, et plus généralement des **objets connectés** dans le monde, démultiplie les points de diffusion et donc de collecte de données. Organisés au sein d'infrastructures et de réseaux spécialisés, ils génèrent de nouveaux services vecteurs de performance dans les réseaux urbains comme dans la vie quotidienne des particuliers. La croissance de ces points de captation de données est phénoménale. D'ici 2020, l'explosion des objets connectés devrait porter leur nombre entre 34 milliards (estimation de BI Intelligence³) et 200 milliards (prévision de l'étude sur l'IoT d'Intel⁴). On notera au passage que **les administrations publiques** formeront, après les entreprises, le deuxième marché pour les objets connectés.

Stockage des données : la croissance exponentielle du volume de données rend obsolète les modalités de stockage traditionnelles. Les environnements cloud (informatique dans les nuages) sont aujourd'hui conçus pour absorber l'afflux croissant de données. L'introduction de la « **scalabilité** », c'est-à-dire de technologies d'ajustement progressif et continu de l'outil de stockage à la masse des données à traiter, permet de répondre aux besoins d'adaptation rapide des outils tant sur le plan du stockage que de la capacité de calcul. Le cloud permet en effet d'automatiser la scalabilité, en fonction de l'usage et de la croissance des besoins de traitement, en ajoutant ou enlevant dynamiquement du stockage et de la puissance de calcul, ce qui facilitera la construction des architectures big data.

Traitement des données : l'**apparition de nouveaux outils de traitement massif**, dans le courant des années 2000, a contribué à l'essor du big data. D'abord avec la création en 2004 de MapReduce, modèle de programmation adapté au traitement de très grandes quantités de données par Google, puis en 2005, avec l'arrivée d'Hadoop, une architecture spécifique de bases de données adaptées aux données non structurées. Sa mise à disposition en open source a contribué à sa généralisation et au développement du big data.

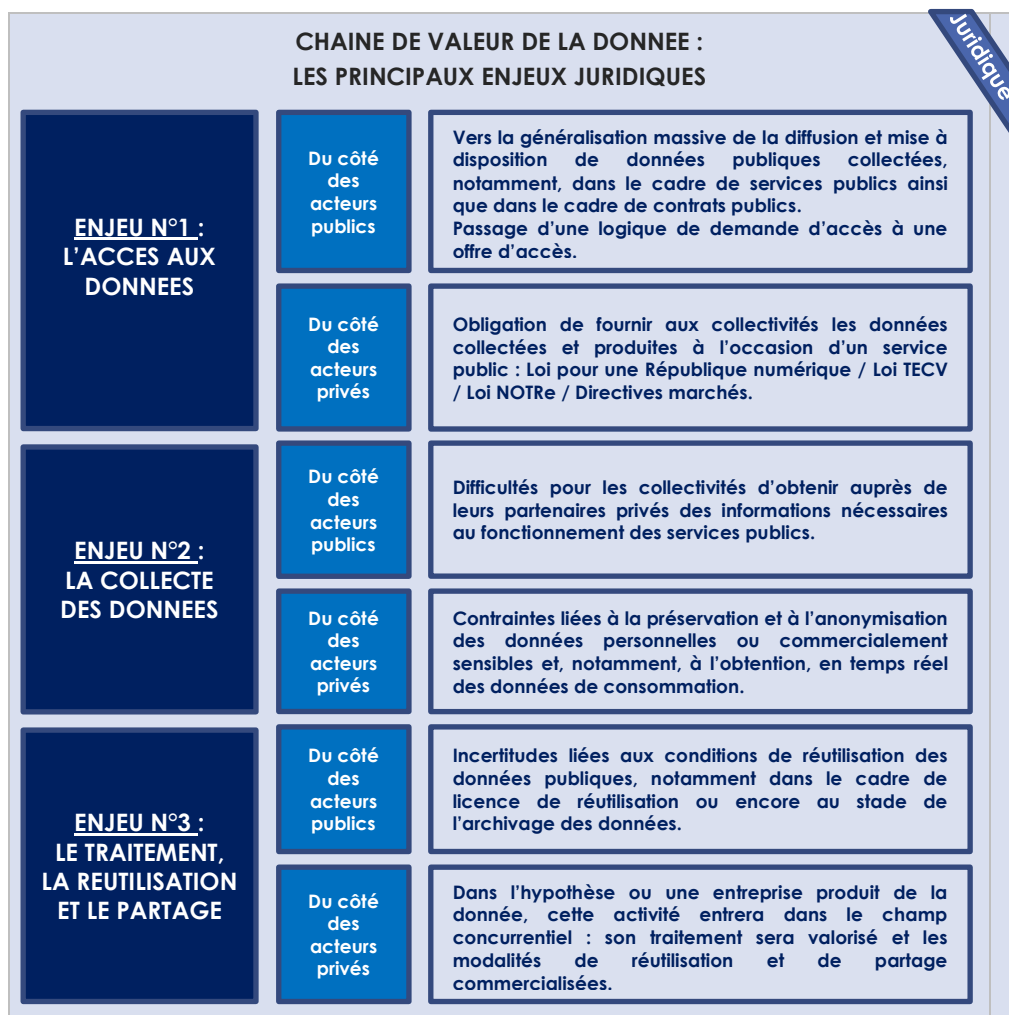
Les solutions big data sont aujourd'hui portées par les grands acteurs du stockage et de l'analytique (IBM, SAP, Oracle ...) ainsi que par des start-ups spécialisées (Hurence, Dataiku).

Représentation des données : au-delà du simple traitement, c'est la **visualisation des données** en contexte et de manière compréhensible par tous qui pousse certaines entreprises à développer des solutions de datavisualisation pour répondre aux nouveaux enjeux : permettre par exemple la visualisation de flux massifs en temps réel, dans une logique de monitoring, ou encore proposer des tableaux de bords (dahsboards) grâce à l'analytique en temps réel.



3 Source : BI Intelligence - Internet of Things report, 2015 <http://www.businessinsider.com/bi-intelligence-34-billion-connected-devices-2020-2015-11?IR=T>

4 Source : Prévisions INTEL <http://www.intel.com/content/www/us/en/internet-of-things/infographics/guide-to-iot.html>



➤ *La donnée : source d'innovation et de développement de nouveaux services*

Si l'utilisation de données à des fins de traitement et d'analyse est inhérente à tout pilotage d'organisation, le changement d'échelle en matière de données a permis le développement de nouveaux services grâce à la collecte et à l'analyse en temps réel de données massives et grâce à des croisements de données inédits (des données open data, des données non structurées, des données de géolocalisations, etc.).

C'est ce qu'on appelle **le data mining** (littéralement "forage des données") ou l'exploration massive de données qui permet d'analyser les informations issues de bases de données et d'établir d'éventuelles corrélations significatives et utilisables entre celles-ci. Si ces fonctions sont aujourd'hui très largement utilisées dans la connaissance client afin d'apporter une aide à la décision performante dans les stratégies marketing, elles pourraient l'être davantage dans la connaissance des besoins des citoyens (en préservant leur anonymat).

La donnée se transforme en information si on lui donne du sens afin de l'interpréter et de la rendre utilisable. Il faut donc traiter les données (parfois en les recoupant avec d'autres données) pour les rendre exploitables et les transformer par la même en information.

Le data mining pratiqué depuis les années 80, s'appuie désormais sur l'ensemble des outils et méthodes d'extraction des connaissances actuelles comme le crawling ou le scraping de données⁵ ou encore le machine learning⁶.

Ces nouveaux outils permettent le développement de trois types d'applications inédites :

- **Détecter et optimiser** : l'abondance et le croisement de données en temps réel apporte une compréhension fine de l'environnement pour une prise de décision facilitée et des activités pilotées plus efficacement ;
- **Tracer et cibler** : la granularité des données analysées assure la découverte et le suivi à un niveau très fin, par exemple de l'individu, dans le cadre d'une population ou d'un pays ;
- **Prévoir et prédire** : la masse de données disponibles sur un phénomène ou une population permet de construire des modèles prédictifs.

Comme le montrent les cas d'usage ci-dessous, tous les secteurs ont d'ailleurs vu leur approche de la relation client et de la création d'offres de services transformée par l'exploration de données.

Banque, Finance, Assurance, Mutuelle

Tracer et cibler /
Prévoir et prédire

Ce secteur a été l'un des premiers à utiliser l'analyse prédictive afin de **détecter des facteurs de risque clients**. L'analyse conjointe des mouvements bancaires effectués en ligne, des opérations effectuées dans des distributeurs, ainsi que des comportements d'achat sur internet permet notamment de détecter des anomalies et des risques de fraudes.

Grande distribution

Tracer et cibler

Data-driven marketing : le big data permet une meilleure appréhension des comportements d'achat des consommateurs grâce à l'analyse et au croisement à grande échelle de données issues de diverses sources, comme les historiques d'achat, les caractéristiques socio-économiques des consommateurs et les données issues des réseaux sociaux.

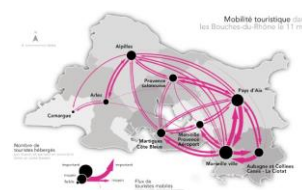
Détecter et optimiser

Optimisation de la gestion des livraisons (planification automatique des tournées grâce aux historiques de livraison)

Tourisme

Tracer et cibler

- En analysant le nombre de transactions effectuées par des étrangers avec leur carte bancaire, des photos et vidéos postées sur des réseaux sociaux ainsi que des tweets géo-localisés, des chercheurs du MIT ont pu mesurer l'attractivité de telle ou telle ville. Cette **analyse du comportement des touristes**, permet d'avoir une meilleure compréhension de leurs besoins.
- Analyse des flux touristiques à partir des traces mobiles : exemple solution Flux vision d'Orange Business Solutions (cf. annexe)



Mobilité touristique dans les Bouches du Rhône
Source : Flux Vision
Tourisme _ Bouches du Rhône Tourisme

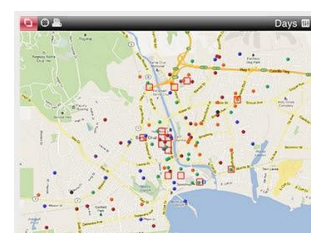
⁵ Crawling et Scraping sont deux techniques différentes d'acquisition automatique des données sur le Web qui permet de les convertir directement en données structurées. (cf. fiche notion)

⁶ Techniques permettant d'obtenir des analyses prédictives à partir de formes d'apprentissage de la machine par les données.

Industrie	
Détecter et optimiser / Prévoir et prédire	Dans le secteur de l'industrie, les progrès de la maintenance prédictive couplés à l'essor de l'internet des objets permettent de détecter des dysfonctionnements en temps réel afin de mieux les prévenir et les gérer. Depuis quelques années, General Electric propose par exemple à ses clients la maintenance prédictive des moteurs d'avion ou des pompes.
Energie	
Détecter et optimiser	• Télémesure de consommations électrique ou de gaz, • Mise en place d'outils de pilotage énergétique à distance des installations par Engie, Cofely. Ces installations mettent en œuvre des échanges de données sur les sites via Wireless M-Bus, puis la collecte ainsi que le contrôle à distance des installations est réalisée via des liaisons cellulaires 3G ou des liens ADSL ou grâce au réseau Sigfox.
Tracer et cibler	Le projet Bento a pour but de développer une " cartographie énergétique " du territoire de Rennes Métropole à l'échelle du bâtiment, ainsi qu'une application web permettant la visualisation et l'exploitation de celle-ci.
Sécurité	
Détecter et optimiser	Optimisation des tournées de surveillance des bâtiments pour prévenir le risque d'incendie. Le New York City Fire Department dispose d'un algorithme qui analyse 60 facteurs de risque d'incendie. Il établit un score de risque pour 330 000 immeubles et fournit une feuille de route.
Prévoir et prédire	La police d'Atlanta s'est dotée du système PredPol ("Predictive Police") qui lui permet de prévoir dans quels quartiers des crimes peuvent survenir pour rendre les interventions pertinentes.
Santé	
Détecter et optimiser	Développement de la chirurgie assistée par ordinateur (CAO) pour le traitement chirurgical des tumeurs à l'Ircad à Strasbourg.
Prévoir et prédire	Identification des facteurs de risque ou de réduction de certaines maladies, grâce aux techniques d'analyse prédictive, ce qui permet d'améliorer les politiques de prévention.
Tracer et cibler	Constitution facilitée de cohortes de patients correspondant le mieux aux critères d'essais clinique, grâce à l'analyse sémantique et intelligente de bases de données de recherche.
Mobilité / transport	
Détecter et optimiser	L'analyse des données de trafic et de déplacements permet d'optimiser les réseaux et de proposer des solutions alternatives en cas de pic. • Le projet « SENSEable City Lab » porté par le MIT et déployé à Rome vise à analyser et croiser les données de géolocalisation des usagers issues de leur téléphone mobile avec les informations GPS générées par le réseau de bus. • Le projet Optymod Lyon, centralise l'ensemble des données issues des différents modes de mobilité et permet la prévision du trafic urbain à une heure. Le projet BigCycleData, conduit par Rennes Métropole, vise à permettre aux



Vertuoz, le portail mis en ligne par ENGIE Cofely permettant à ses clients de suivre en ligne la performance énergétique de leur patrimoine



PredPol



Chirurgie assistée par ordinateur

	opérateurs de réseaux de vélos en libre-service de prédire la demande en temps réel grâce à l'analyse des données de localisation des vélos via la plateforme 3DEXPERIENCE.
Environnement	
Mesurer détecter	Les réseaux de capteurs environnementaux, que ce soit dans l'analyse des pollutions de l'air, de l'eau et des sols, ont fortement progressé. Dans le domaine de l'eau, les systèmes proposés (capteurs de pollution, débitmètres, compteurs électroniques) permettent : • D'améliorer la gestion en assurant une surveillance des réseaux à distance. • D'agir préventivement sur les risques de pollution et sur la détection des fuites (amélioration de la qualité et des rendements). Les autres domaines, notamment la pollution de l'air et des sols, bénéficient également des moyens renforcés : • Le projet de monitoring urbain environnemental de la métropole de Nice Côte d'Azur teste une nouvelle application qui vise à améliorer la vie et la santé des personnes allergiques aux pollens d'essences végétales
Anticiper sur les risques	Les expérimentations de systèmes préventifs et prédictifs en matière de risques naturels sont également variées • Sismologie : le RESIF (Réseau Sismologique et Géodésique Français) est un équipement d'excellence (programme Equipex) financé pour la période 2012-2018, qui vise à doter la France d'un outil d'observation sismologique performant. Un réseau de capteurs sera bientôt capable de mesurer à l'échelle de la France les déformations de la surface de la Terre sur une échelle de temps allant de la dizaine d'années à la milliseconde. • Prévention des inondations : Nice expérimente des systèmes à base de vidéos capables de détecter un risque d'inondation deux heures à l'avance.

2. Gouvernance de la donnée : définition et enjeux

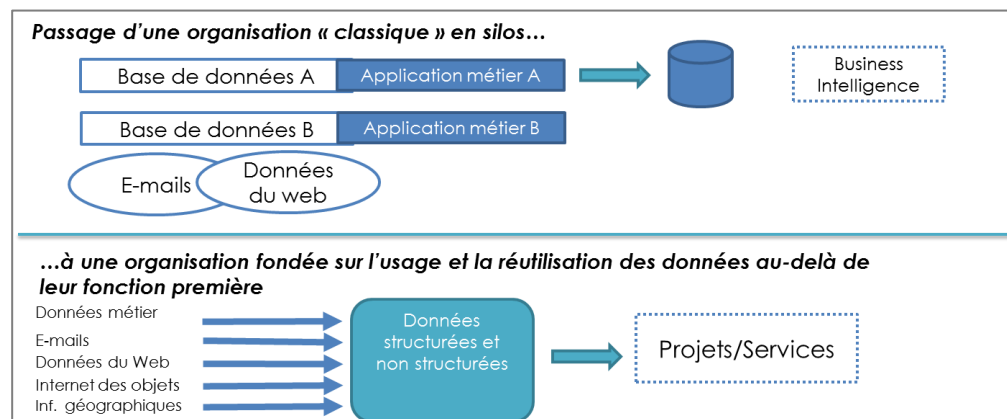
➤ *D'une organisation en silos à une organisation fondée sur l'usage et la réutilisation des données*

Ces dernières années, on note une évolution singulière dans la manière dont les entreprises et les administrations commencent à gérer leurs données.

Les applications métiers fonctionnent encore assez systématiquement en silos à partir de bases de données structurées dédiées aux métiers et généralement propriétaires. Certes les outils de business intelligence (veille d'affaires de haut niveau) sont en mesure d'analyser des données consolidées et structurées à partir d'extractions, mais généralement dans un contexte contraint et relativement limité.

Toutefois les systèmes d'information amorcent une évolution vers des **architectures orientées services** et axées sur la donnée, afin d'assurer la communication entre les différentes applications du Système d'information. Ces architectures se fondent sur l'usage et la réutilisation des données au-delà de leur fonction première. De fait cette "gouvernance de la donnée" constitue un renversement de paradigme dans le mode d'organisation de l'entreprise ou de l'administration dans le sens où la donnée

devient l'élément central. Les métiers sont amenés à travailler davantage en synergie et les temps de décisions sont potentiellement raccourcis grâce à l'actualisation et à la mise à disposition permanente de l'information. Les outils de type big data permettent alors d'analyser des données reflétant la complexité du monde réel (données structurées et non structurées) pour proposer de nouveaux services ou améliorer l'efficacité de ceux déjà en place.



➤ *Différentes familles de préoccupations à conjuguer : faire converger l'approche métier et les approches transversales*

Pour parvenir à mettre en place une architecture orientée services, deux grandes familles de préoccupations doivent être prises en compte :

- **Les préoccupations sur un domaine particulier**, le plus souvent sur un département ou un service support (données métier, données RH, données comptables, etc.). Dans ce cas, l'accent est mis sur l'exploitation du patrimoine d'informations avec un questionnement sur la qualité, l'usage opérationnel, voire l'enrichissement des données disponibles ;
- **Les préoccupations liées à la cohérence transverse de données clés pour l'organisation**. Il s'agit pour l'essentiel de données de référence massivement réutilisables (Master data), vers lesquelles il conviendra de faire converger les utilisateurs. Dans ce cas la performance résidera dans la mise en place de processus permettant d'accéder rapidement à ces données.

Les données de référence d'un service public de transport urbain seront par exemple la base de données des arrêts de bus, tramway et des gares, données clés pour construire les lignes de transport.

Ces préoccupations ne s'opposent pas, elles tendent plutôt à converger. Au moins sur les activités de description de la donnée et de son cycle de vie. Il convient donc de mettre en place une organisation capable de traiter toutes ces préoccupations : les problèmes quotidiens de la gestion de données, comme l'accompagnement de projets de transformation de l'entreprise qui vont impacter le cycle de vie des données afin de l'adapter au nouveau contexte.

Dans un tel schéma, la création concertée de référentiels de données devient cruciale puisque la représentation des connaissances va servir de socle aux développements d'applications métiers.

➤ Des enjeux en matière de circulation et de réutilisation des données

L'architecture orientée service, confère à la donnée une valeur liée pour l'essentiel à son potentiel de circulation et de réutilisation, que ce soit en interne (entre les métiers) ou en externe (avec des partenaires, ou en open data). Aussi, une gouvernance de la donnée suppose de prêter une attention particulière à un grand nombre d'éléments structurants constitutifs de la qualité de cette donnée :

Propriété	• Au-delà de la production, qui en est propriétaire ? • Quelles sont les responsabilités qui s'y rattachent ?
Confidentialité et sécurité	• Comment assurer la protection des données personnelles et confidentielles ? • Quel niveau d'anonymisation choisir (sans pénaliser la qualité du service) ? • Comment se prémunir contre les cyber-attaques ?
Qualité	• Comment éviter les doublons ? • Comment les qualifier du point de vue de leur véracité ?
Mutualisation des fonctions	• Quelles fonctions mettre en commun ? • Quel rôle transverse créer ?
Portage des projets et conditions d'échange et de partage	• Comment éventuellement inscrire dans les contrats les conditions de transfert des données en fonction de leur type ?

➤ Des gains attendus variés

Les dispositifs de gouvernance de données peuvent engendrer des gains de différentes natures :

- **Augmentation des revenus** de l'organisation. Ces recettes supplémentaires sont rendues possibles grâce à l'accélération de la collaboration des partenaires, à l'amélioration de la pertinence des analyses ou encore à la mise en place de nouveaux services ;
- **Réduction des coûts métiers et amélioration de la productivité.** La mise en place d'un référentiel transverse au sein d'un groupe de transport a par exemple permis de réduire les coûts métier (notamment ceux liés à la double-saisie) et de mettre en place un reporting groupe unifié ;
- **Diminution des coûts d'exploitation et de maintenance des systèmes d'information, grâce à la mutualisation et à la simplification de services de diffusion des données.** La mise en place d'une architecture orientée services à l'échelle de l'organisation peut conduire à gagner en agilité et donc en coût de mise en œuvre des futurs projets ;
- **Atténuation des risques** notamment ceux liés au non-respect des exigences réglementaires ou au déficit d'image en cas de pertes de données.

3. Focus sur le marché de la donnée et du big data

➤ *Une offre big data toujours en croissance et un décollage contrasté de la demande*

Le marché des fournisseurs de solutions big data connaît une forte croissance. Au cours des dernières années, de nombreux acteurs se sont positionnés sur cette filière :

- Des fournisseurs historiques de solutions IT : IBM, SAP, Oracle...
- Des acteurs du Web : Facebook, Google...
- Des spécialistes de solutions data : Teradata, MapR...
- Des éditeurs de solutions de business intelligence ou d'analytique big data : SAS, Datameer, Zettaset...
- Des intégrateurs : Cap Gemini, Accenture, ATOS...

Le cabinet IDC les regroupe en 3 catégories, en fonction de leur offre de services :

Infrastructures de stockage	• Serveurs • Hardware • Réseaux • Stockage • Sécurité des datacenters • Etc.
Logiciels	• Systèmes de gestion • Bases de données • Analytics • Logiciels métiers • Etc.
Services	• Conseil • Intégration

Côté utilisateurs, la demande connaît aujourd'hui un décollage contrasté selon les secteurs.

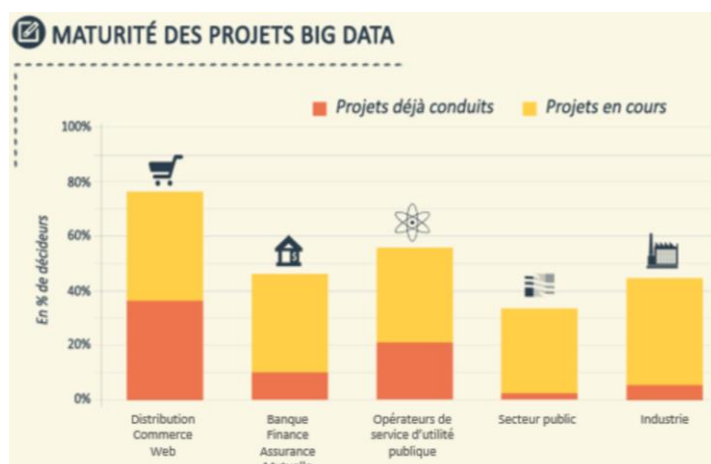
Les télécoms, la grande distribution et la finance ont été les premiers secteurs à se tourner vers le big data car le grand nombre de données captées apportait un avantage comparatif et facilitait le développement de nouveaux services.

La santé offre aujourd'hui les plus intéressantes perspectives, au niveau de la recherche et du ciblage des médicaments.

Le secteur public et les opérateurs de services d'utilité publique gagnent en maturité et de plus en plus de projets voient le jour, notamment dans l'objectif de réduire les coûts et de lutter contre les fraudes.

Enfin, d'après IDC⁷, parmi les débouchés les plus prometteurs figurent également le **secteur manufacturier** (2,1 milliards de dollars en 2014) et **l'industrie de process** (1,5 milliards de dollars en 2014).

⁷ Source : prévision IDC-2015 : <http://www.idc.com/getdoc.jsp?containerId=prUS40560115>



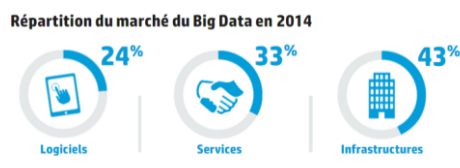
SOURCE : MARKESS, JUIN 2014

➤ Un marché en forte augmentation en France et dans le monde

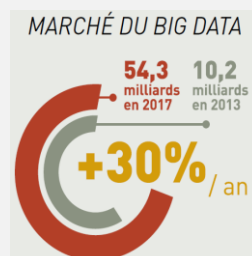
Si les différentes études menées sur le marché du big data peinent à converger sur une estimation de sa taille (vraisemblablement en raison des différences de périmètre), toutes s'accordent sur une croissance à deux chiffres dans les trois prochaines années.

Au niveau mondial, Mushroom Networks⁸ évalue le marché du big data à 54,3 milliards de dollars en 2017 (pour 10,2 milliards en 2013) avec une croissance d'au moins **30%** pour les années à venir tandis que, selon IDC⁹, le marché mondial des équipements, logiciels et services de big data devrait croître de **23,1%** par an en moyenne pour atteindre **48,6 milliards de dollars en 2019**.

Au niveau français, le marché du big data connaîtrait actuellement une croissance de **40%** par an et serait estimé à **652 millions d'euros en 2018** d'après IDC¹⁰. MARKESS¹¹ qui s'est intéressé plus largement au marché « des solutions d'analytique, de big data et de gestion des données ainsi que des services associés » estime quant à lui qu'il pesait déjà plus de 1,9 milliard d'euros à fin 2015 et qu'il devrait croître à un rythme annuel de plus de 12% entre 2016 et 2018.



SOURCE : « LA DYNAMIQUE DU BIG DATA EN FRANCE », INFOGRAPHIE HP, DONNEES IDC¹²



DANS : GUIDE DU BIG DATA 2015-2016
(SOURCE : MUSHROOM NETWORKS)

⁸ Source : Mushroom Networks - 2014 <http://www.mushroomnetworks.com/infographics/the-landscape-of-big-data-infographic>

⁹ Source : prévisions IDC - 2015 : <http://www.idc.com/getdoc.jsp?containerId=prUS40560115>

¹⁰ Source : prévisions IDC - 2014

¹¹ Source : <http://blog.markess.fr/2016/03/analytique-big-data-et-gestion-des-donnees-un-marche-de-pres-de-2-mds.html>

¹² En juin 2014, l'International Data Corporation a mené l'enquête auprès de 200 entreprises de plus de 500 employés issus de tous les secteurs d'activité. Les données et statistiques récoltées ont permis de réaliser une infographie sur la dynamique du big data en France, en partenariat avec Hewlett-Packard. <http://www.lebigdata.fr/infographie-idc-enjeux-dynamique-big-data-france>

La croissance du marché mondial est actuellement tirée par les ventes de logiciels de traitement et d'analyse par opposition aux services de stockage dont la proportion tend à s'amoindrir¹³ :

- +26,2% en moyenne par an d'ici 2019 pour les logiciels de gestion, d'exploration, d'analyse et d'application des données,
- +22,7% en moyenne par an d'ici 2019 pour les services de conseil et d'accompagnement dans la mise en place d'applications,
- +21,7% en moyenne par an d'ici 2019 pour les équipements d'infrastructure (dont stockage).

Au niveau des logiciels, on constate en effet un renouveau complet puisque les solutions traditionnelles (de reporting, décisionnels, de tableaux de bord...) sont en perte de vitesse tandis que de nouvelles générations de solutions, plus en phase avec les enjeux actuels arrivent.

Les enjeux prioritaires du moment pourraient porter sur quelques axes, comme le suggère une étude en cours portée par la Direction générale des entreprises (DGE)¹⁴ :

- Comprendre ce qu'on peut obtenir des données et étudier le besoin avant d'engager des actions
- Communiquer autour du big data pour rassurer le public
- Développer une éthique autour de l'usage de la donnée
- Monétiser les données pour préserver la souveraineté des propriétaires.

➤ *Focus sur les services à valeur ajoutée pour les collectivités apportés par quelques acteurs innovants du marché français*

Avec le bouleversement introduit par le data déluge et la complexification des données, de nouvelles solutions sont apparues sur le marché formant un nouvel écosystème composé à la fois de grands groupes et de startups innovantes. Ainsi, la librairie open source Hadoop créée par Yahoo en 2004 est utilisée par de nombreuses entreprises qui y apportent leur propre enrichissement.

Aussi, même dans un environnement ouvert, chaque acteur se positionne sur la chaîne de valeur et propose des solutions spécifiques, plus ou moins intégrées, intervenant sur au moins l'une des quatre dimensions suivantes :

- Développement et intégration de bases de données,
- Hébergement et maintenance de plateformes,
- Equipement et apport de la puissance de calcul,
- Utilisation des données via des applications.

¹³ Source : prévisions IDC - 2015 : <http://www.idc.com/getdoc.jsp?containerId=prUS40560115>

¹⁴ Etude en cours sur l'opportunité de créer un "observatoire des usages du big data" en association avec Tech'In France anciennement Association Française des Editeurs de Logiciels et Solutions Internet (achèvement prévu en fin d'année 2016)

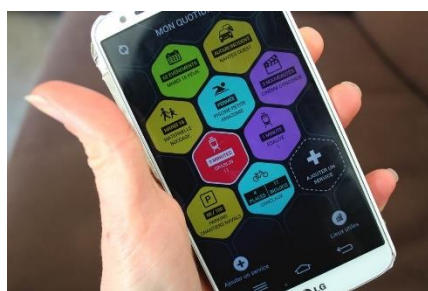
Les administrations et les collectivités n'ont pas toujours connaissance de ces solutions : elles doivent s'organiser pour mener une veille leur permettant de détecter les offres pertinentes pour leurs besoins actuels et futurs.

En effet, des services basés sur la donnée et à réelle valeur ajoutée pour le secteur public sont aujourd'hui disponibles.

Des plateformes privées clé en main dédiées à l'open data et à la smart data sont par exemple proposées aux collectivités en prévision des nouvelles obligations que leur impose la loi pour une République numérique. Des plateformes de services en marque blanche permettant d'agréger des données issues de différents canaux, y compris des données temps réel, existent déjà. En effet, certaines entreprises, comme Ozwillo et OpenDataSoft (cf. ci-contre), proposent des solutions capables de gérer des niveaux d'accès différenciés entre les données ouvertes au public et celles dont l'accès sera plus restreint. La plateforme OpenDataSoft est utilisée par certains de ses clients pour effectuer de l'analyse prédictive : prévoir les places disponibles aux bornes de vélopartage dans une heure, ou fournir un niveau de pollution de l'air en temps réel...

Au-delà de l'open data, la question est plus largement celle de la gestion des données et des échanges de données avec les prestataires. Certaines entreprises proposent aux collectivités d'externaliser cette activité. SUEZ Smart solutions, par exemple, participe activement à une expérimentation à Saint-Etienne sur une plateforme de gestion des données en prestation. L'outil qu'elle propose permet également la dématérialisation des archives pour les migrations lors des changements de prestataires et l'ouverture directe de certaines données.

L'accès à l'information en temps réel par les usagers est également un enjeu primordial. C'est pourquoi certains territoires ont développé des **applications unifiées d'information à destination des citoyens**. En effet, à partir d'entrepôts de données, il est possible d'alimenter une application de type "tableau de bord" permettant au citoyen d'accéder à une grande variété d'informations mises à jour en temps réel. Orange Applications for Business a par exemple développé une application multiservices dédiée aux services de la mobilité et de la vie quotidienne. Disponible sur smartphones et tablettes, elle propose tout un panel de services de proximité facilement personnalisable. Chacun peut configurer son application en fonction de ses besoins (horaires d'arrêts de bus personnalisés, ouverture et bientôt fréquentation en temps réel des piscines de l'agglomération, etc.).



SOURCE : NANTES DANS MA POCHE

Ozwillo



Propose une infrastructure de données plutôt orientée collectivités, mais pas uniquement, visant à partager des connaissances en liant tous types de modèles de données. Sur la plateforme, il est possible de déposer des applications (24 déjà disponibles) pour les rendre communicantes, lier des données, les partager et constituer en quelques clics une plateforme Open Data avec ses API.

OpenDataSoft



Propose une solution en mode SaaS permettant d'agglomérer (via des API) et de rendre compatibles des données venant de canaux très différents (y compris des données non structurées). La startup compte déjà parmi ses clients une trentaine de collectivités et quelques grands groupes sur le secteur privé (dont La Poste, SNCF,

Orange Applications for Business



Propose l'application mobile « Ma Ville dans ma Poche » déployée à Nantes. Nantes dans ma poche offre aujourd'hui 14 micro-services : circulation, agenda, voie publique, TAN, stations bicloo et parking, info écoles, piscines, commerces, activités physiques, cinéma et qualité de l'air et permet d'effectuer des signalements.

La bonne gestion du territoire requiert également une **connaissance des flux comportementaux des usagers, des touristes...** En effet, les collectivités ont besoin de connaître les déplacements des citoyens sur le territoire avant de lancer de grands projets :

- Pour penser l'aménagement de lieux où un grand nombre de personnes sont amenées à se croiser,
- Pour adapter l'offre touristique,
- Pour gérer la signalisation aux carrefours et accorder la priorité aux piétons,
- Et pour bien d'autres analyses de données de flux touchant aux piétons et aux personnes en mobilité.

Cette analyse des flux peut s'appuyer sur plusieurs technologies :

- **Analyse d'images** : la société Placemeter propose une plateforme capable de transformer des contenus issus de vidéos en données de trafic en temps réel (comptage, entrées/sorties...) par des algorithmes de reconnaissance de mouvements (voir encadré ci-contre) ;
- **Analyse des données mobiles** : Orange Applications for Business propose la solution Flux Vision qui offre une analyse de données mobiles pour la mesure de la fréquentation d'une zone et le suivi des déplacements des populations (ex : analyse de la fréquentation d'événements culturels ou sportifs par typologie de population) ;
- **Analyse des données de connexions wifi** (possible lorsqu'une zone est couverte en Wi-Fi public).

Enfin, d'autres outils basés sur la donnée permettent une **analyse fine du tissu économique local pour** connaître avec précision le tissu de startups, analyser certaines filières économiques locales, identifier des écosystèmes d'entreprises (technologies vertes, fintech, start-ups, blockchain, etc.). Des sociétés spécialisées proposent des solutions capables de croiser les données non structurées et les données issues de grandes bases de données nationales, pour établir un portrait assez fidèle et détaillé de la présence des entreprises sur une agglomération, un département, une région. Cette connaissance permet d'identifier clairement des écosystèmes d'entreprises, d'élaborer des stratégies de développement économique ou encore d'assurer une opération plus chirurgicale sur un groupe d'entreprises. C'est ce que propose par exemple l'entreprise française Data Publica, spécialisée dans « l'information sur les entreprises » et le marketing prédictif, qui a par exemple identifié 10 écosystèmes d'entreprises pour la CCI de Paris-IDF (entreprises de l'économie circulaire, de l'industrie 4.0, de l'économie du partage, etc.).

Placemeter



Propose la plateforme « Open urban intelligence platform » spécialisée dans l'analyse de mouvement et de flux de déplacements à partir de contenus vidéo numériques. La start-up travaille actuellement sur un projet de réaménagement de la Place de la Nation. Cette technologie permet en effet de mesurer l'impact "usages" des aménagements de mobiliers urbains déplaçables sur la place (déplacements ponctuels, ou en fonction des saisons).

Data Publica



Propose une base de données comptant 4,5 millions d'entreprises. Celle-ci est constituée à partir d'algorithmes de type machine learning appliqués à des données non structurées (aspirées sur 600 000 sites web d'entreprises en France) et à des bases de données nationales structurées provenant d'Infogreffe, de la DILA, de l'INSEE ou encore de l'INPI.

B. Un contexte législatif et réglementaire plus offensif sur la donnée

Le « droit des données » au sens large a connu une véritable révolution consécutive à l'avènement de la société de l'information et au développement des nouvelles technologies. De nombreux textes, tant communautaires que nationaux encadrent désormais les données, ce qui peut aboutir à rendre parfois le cadre applicable peu lisible.

Dans le cadre de cette étude, le cadre réglementaire présenté n'a pas vocation à être exhaustif mais expose les textes les plus pertinents et les plus récents, tant au niveau communautaire que national.

1. Le cadre juridique communautaire applicable aux données des collectivités

Au niveau communautaire, quatre textes principaux ont influencé le « droit des données » au sens large et inspiré de nombreux textes nationaux.

En premier lieu, la directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données constitue le texte communautaire de référence en matière de protection des données à caractère personnel.

En effet, la directive a mis en place un cadre réglementaire visant à établir un équilibre entre un niveau élevé de protection de la vie privée des personnes d'une part, et la libre circulation des données à caractère personnel au sein de l'Union européenne (ci-après « UE ») d'autre part.

Pour ce faire, la directive a, en son temps, fixé des limites strictes à la collecte et à l'utilisation des données à caractère personnel, applicables aux personnes publiques.

Cette directive a notamment abouti à la création, dans chaque État membre, d'un organisme national indépendant chargé de la supervision de toutes les activités liées au traitement des données à caractère personnel. Ces créations se sont faites sur le modèle de la Commission nationale de l'informatique et des libertés (CNIL) française, créée en 1978 par la loi informatique et libertés.

En deuxième lieu, plusieurs années plus tard, **la directive 2003/98 du 17 novembre 2003 du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public** a établi les premiers fondements de la réutilisation, par les opérateurs économiques, des informations publiques, qualifiées de « ressource essentielle » pour le développement d'activités innovantes.

En revanche, elle n'a pas entendu imposer aux États membres d'autoriser la réutilisation de leurs informations publiques, les laissant libres du choix d'ouvrir ou pas leurs données à la réutilisation.

Dès lors, l'absence de création d'un véritable droit à la réutilisation opposable aux États membres et le faible degré d'encadrement des conditions financières de

réutilisation constituaient de fait, autant de freins à une vraie politique d'ouverture des données publiques la plus large possible en vue de leur valorisation.

Ces constats ont donné lieu à la révision de la directive 2003/98 et à l'adoption, le 26 juin 2013, de la directive 2013/37/CE.

En troisième lieu, **la directive 2013/37/UE modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public, dite « Directive PSI » (« public sector information »)**, a proposé un véritable cadre juridique harmonisé à l'échelle de l'UE, propice au développement de l'ouverture et de la réutilisation des données publiques.

En effet, la directive 2013/37/CE impose dorénavant une obligation générale pesant sur les Etats membres, d'autoriser la réutilisation à des fins commerciales ou non de leurs informations publiques librement accessibles, sauf si celles-ci sont protégées par un droit de propriété intellectuelle au bénéfice d'un tiers.

Cette obligation porte notamment sur les informations publiques des bibliothèques, musées et archives. L'encadrement des tarifs de réutilisation est également prévu par la directive.

Dans la continuité de la directive 2003/98, la possibilité dérogatoire de conclure des accords d'exclusivité pour la réutilisation des données culturelles est néanmoins maintenue lorsqu'elle s'accompagne de leur numérisation au profit de l'organisme public.

Cette directive a été transposée en droit national par la **loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public**. Allant au-delà des obligations issues de la directive, la loi a instauré un principe de gratuité de la réutilisation des informations publiques et restreint les exceptions ouvertes par la directive en matière de tarification et d'accord d'exclusivité (voir ci-après).

Enfin, en quatrième lieu, **le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE** (mise en œuvre effective programmée pour 2018) constitue, en l'état du droit communautaire applicable, l'aboutissement de quatre années de travail et de négociations intenses et marque surtout un tournant majeur dans la régulation des données personnelles.

Parmi les mesures notables, pour la plupart transposées en droit national par la loi pour une République numérique, figurent notamment :

- L'obligation de recueillir un consentement « clair et explicite » avant tout traitement de données personnelles. Il sera interdit de se contenter par exemple d'une politique de vie privée accessible par un lien, ou même de cocher par défaut des cases de recueil du consentement. Celui-ci devra être en opt-in uniquement ;
- L'interdiction des réseaux sociaux ou autres services collecteurs de données pour les enfants, sauf autorisation des parents (article 8). Les Etats membres pourront fixer la limite d'âge entre 13 et 16 ans, selon leur sensibilité ;

- La reconnaissance d'un « droit à l'oubli » (article 17) qui permet à un individu de demander l'effacement des données qui le concernent, y compris chez les sous-traitants ou partenaires, à condition que leur conservation ne soit pas nécessaire pour un motif légitime (recherches historiques, scientifiques, statistiques, santé publique, exécution d'un contrat...), y compris le droit à la liberté d'expression ;
- Le droit à la portabilité des données (article 20) qui offre aux utilisateurs d'un service en ligne la possibilité de prendre leurs données avec eux pour les importer dans un service concurrent, par exemple pour changer de fournisseur de messagerie électronique sans perdre ses contacts ou ses messages ;
- La limitation du profilage par algorithmes (article 21). En principe, aucune décision ne doit pouvoir être prise sur la base d'une détermination purement algorithmique du profil de la personne. Par ailleurs, celui-ci n'est autorisé que si la personne donne son consentement. La portée exacte de l'article reste toutefois à analyser, tant il semble souple.
- Le droit d'être informé en cas de piratage des données (articles 33 et 34) : si une entreprise ou une organisation quelconque est victime d'un piratage de données de ses clients ou de tiers, elle devra immédiatement en informer l'autorité de protection des données (en France, la CNIL), et dans le cas où cette divulgation ne pose pas de problème de sécurité, en informer les principaux concernés.
- Enfin, la possibilité d'infliger des amendes jusqu'à 4 % du chiffre d'affaires mondial d'une entreprise, lorsqu'elle viole le droit à la protection des données. La sanction sera d'autant plus forte que la violation sera grave et consciente.

2. Le cadre juridique national applicable aux données des collectivités

➤ Dispositions générales

Outre la loi pour une République numérique, quatre lois essentielles ont été relevées dans le cadre de la présente étude, tant leur impact est fort sur le « droit des données » et les modalités de la réutilisation des informations du secteur public.

Elles font ici l'objet d'une courte présentation :

En premier lieu, **la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dite « loi CNIL »**, modifiée à de nombreuses reprises, constitue le texte de référence en matière de protection des données personnelles.

Elle définit précisément les principes à respecter à tous les stades de gestion de la donnée, identifiés dans le cadre de la présente étude (collecte, traitement et conservation), en matière de protection des données personnelles.

La loi est applicable dès lors qu'il existe un traitement automatisé ou un fichier manuel, c'est-à-dire un fichier informatique ou un fichier « papier » contenant des informations personnelles relatives à des personnes physiques, à l'exception des « traitements mis en œuvre pour l'exercice d'activités exclusivement personnelles »

tels que par exemple les agendas électroniques, les répertoires d'adresses, les sites internet familiaux en accès restreint.

Sauf désignation expresse par les dispositions législatives ou réglementaires relatives à ce traitement, la loi définit le responsable du traitement de données à caractère personnel comme la personne, l'autorité publique, le service ou l'organisme qui détermine ses finalités et ses moyens.

Elle s'applique donc indifféremment aux personnes publiques ou aux personnes privées.

En application de la loi, un traitement ne peut porter que sur des données à caractère personnel qui satisfont aux conditions suivantes :

Principe de loyauté et de licéité de la collecte des données	Modalités loyales et licites de collecte et de traitement
Principe de finalité	Finalités déterminées, explicites et légitimes
Principe de pertinence et de proportionnalité	Données adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont collectées et de leurs traitements ultérieurs
Principe d'exactitude et de qualité des données	Données exactes, complètes et, si nécessaire, mises à jour ; les mesures appropriées doivent être prises pour que les données inexactes ou incomplètes au regard des finalités pour lesquelles elles sont collectées ou traitées soient effacées ou rectifiées
Principe de durée limitée de conservation des données	Données conservées sous une forme permettant l'identification des personnes concernées pendant une durée qui n'excède pas la durée nécessaire aux finalités pour lesquelles elles sont collectées et traitées

En deuxième lieu, **la loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal, dite « loi CADA »**, reconnaît à toute personne, sauf cas particulier, le droit d'obtenir communication des documents détenus par une administration dans le cadre de sa mission de service public, quels que soient leurs formes ou leurs supports. Les conditions et modalités d'accès aux documents administratifs sont également précisés ainsi que les pouvoirs et modalités de saisine de la Commission d'Accès aux Documents Administratifs (CADA).

Par ailleurs, les informations qui entrent dans le champ de cette loi peuvent être exploitées à d'autres fins que celles de la mission de service public pour les besoins de laquelle elles ont été produites, y compris commerciales. En effet, un vrai droit à la réutilisation des informations publiques est prévu dans la loi.

La loi est désormais codifiée dans le code des relations entre le public et l'administration, entré en vigueur le 1^{er} janvier 2016, qui régit les échanges entre le public et l'administration, les règles de forme et les conditions d'application des actes administratifs et les modalités d'accès aux documents administratifs.

En troisième lieu, **la loi n°79-18 du 3 janvier 1979 sur les archives**, modifiée par la **loi n°2008-696 du 15 juillet 2008**, désormais codifiée dans le *code du patrimoine* (ordonnance n°2004-178 du 20 février 2004) rassemble dans un même texte, archives publiques et archives privées, définies comme l'ensemble des documents, quels que soient leur date, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité. La loi prévoit que les collectivités territoriales sont propriétaire de leurs archives et sont responsables de leur conservation et de leur mise en valeur.

Enfin, en quatrième lieu, **la loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public**, désormais codifiée dans le *code des relations entre le public et l'administration*, transpose en droit interne la directive PSI et pose le principe essentiel selon lequel la réutilisation d'informations publiques est gratuite.

Toutefois, les administrations peuvent néanmoins établir une redevance de réutilisation lorsqu'elles sont tenues de couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service public. Le produit total du montant de cette redevance, évalué sur une période comptable appropriée, ne doit pas dépasser le montant total des coûts liés à la collecte, à la production, à la mise à la disposition du public ou à la diffusion de leurs informations publiques.

Par ailleurs, la réutilisation peut également donner lieu au versement d'une redevance lorsqu'elle porte sur des informations issues des opérations de numérisation des fonds et des collections des bibliothèques, y compris des bibliothèques universitaires, des musées et des archives et, le cas échéant, sur des informations qui y sont associées lorsque ces dernières sont commercialisées conjointement.

LOI NUMERIQUE

PRESENTATION SYNTHETIQUE DES PRINCIPALES DISPOSITIONS DE LA LOI INTERESSANT LE CHAMP DE LA PRESENTE ETUDE

Juridique

L'élaboration de la loi pour une République numérique a donné lieu à de nombreuses concertations et plusieurs autorités ont été saisies pour avis, telles que la CADA, la CNIL mais aussi l'ARCEP et le Conseil d'Etat, lesquelles ont pu exprimer leurs réserves sur certains aspects de la loi.

La loi pour une République numérique contient plus particulièrement deux objectifs intéressant le champ de la présente étude :

- **D'une part, la généralisation de la mise à disposition des données publiques s'articulant autour de trois axes majeurs :**
 - **L'amélioration de l'accès aux documents administratifs** en favorisant notamment les échanges d'informations entre administrations (articles 1, 2, 8, 10 et 11) ;
 - **L'élargissement des obligations de diffusion** (articles 6, 7, 8, 9,10) par le passage d'une logique de demande d'accès à une logique de mise à disposition spontanée par les administrations ;

- **L'ouverture de l'accès à certaines données des titulaires de marchés publics, de marchés de partenariat et de conventions de subventions** (articles 17 et 18).
- **D'autre part, la création d'un service public de la donnée** dont le régime juridique est le suivant :
 - **Nouvelle mission de service public relevant de l'Etat** consistant dans « la mise à disposition des données de références » et à laquelle « concourent » les collectivités territoriales ;
 - **Instauration de la notion de « données de référence »**, qui doivent satisfaire aux conditions suivantes :
 - Elles constituent une référence commune pour nommer ou identifier des produits, des services, des territoires ou des personnes ;
 - Elles sont réutilisées fréquemment par des personnes publiques ou privées autres que l'administration qui les détient ;
 - Leur réutilisation nécessite qu'elles soient mises à disposition avec un niveau élevé de qualité.

En l'absence de précisions dans la loi, il conviendra d'attendre la publication des décrets d'application, lesquels devront à notre sens :

- Préciser les modalités de participation et de coordination des différentes administrations nationales et territoriales ;
- Fixer les critères de qualité que doit respecter la mise à disposition des données de référence ;
- Dresser la liste des données de référence et désigner les administrations responsables de leur production et de leur mise à disposition.

➤ *Dispositions propres aux contrats et marchés publics*

La réforme profonde du droit de la commande publique, menée ces dernières années, a notamment eu pour objectifs de simplifier et de sécuriser le droit de la commande publique et de l'ouvrir davantage aux petites et moyennes entreprises.

Un code unique pour les marchés publics, les concessions de service public, les partenariats public-privé devrait voir le jour en 2016-2017.

Les objectifs visés par cette réforme d'ampleur, transposant les trois directives communautaires du 26 février 2014, consistent notamment à simplifier et à sécuriser le droit de la commande publique en permettant, pour ce qui concerne la présente étude, l'accès aux **données essentielles** des contrats publics, sous format ouvert et librement réutilisable.

Quatre textes essentiels ont déjà été adoptés dans ce cadre :

- L'ordonnance n° 2015-899 du 23 juillet 2015 relative aux marchés publics,
- L'ordonnance n° 2016-65 du 29 janvier 2016 relative aux contrats de concession
- Le décret n° 2016-86 du 1er février 2016 relatif aux contrats de concession

- Le décret n° 2016-360 du 25 mars 2016 relatif aux marchés publics.

L'ordonnance et le décret relatifs aux marchés publics	Obligation mise à la charge de l'acheteur : • <u>de mettre à disposition sur son profil acheteur les données essentielles des marchés publics ;</u> • <u>de conserver les documents relatifs à la passation et à l'exécution des marchés publics.</u>
L'ordonnance et le décret relatifs aux concessions	L'autorité concédante offre, sur son profil d'acheteur et au plus tard le 1er octobre 2018 : • <u>avant le début d'exécution du contrat de concession : un accès libre, direct et complet aux données essentielles du contrat de concession,</u> • <u>chaque année et à l'occasion de chaque modification apportée au contrat de concession : les données relatives au contrat.</u>

Concrètement, ces données « essentielles », tant pour les marchés publics que pour les concessions, comprennent notamment les informations suivantes :

- L'identification de l'acheteur/autorité délégante,
- La nature et l'objet du contrat,
- La procédure de passation utilisée,
- Le lieu principal d'exécution des prestations,
- La durée du contrat et son montant global,
- L'identification du titulaire, la date de conclusion du contrat.

Le 1er octobre 2018, au plus tard, l'acheteur ou l'autorité concédante offrira, sur son profil, "un accès libre, direct et complet aux données essentielles" des marchés publics et des contrats de concessions (à l'exception des informations dont la divulgation serait contraire à l'ordre public).

➤ Dispositions sectorielles

Des dispositions sectorielles relatives au « droit des données » figurent également dans différents textes législatifs et réglementaires nationaux.

Quatre secteurs ont été plus particulièrement étudiés dans le cadre de la présente étude : les transports, la santé, l'électricité-gaz-réseaux de chaleur et enfin le renseignement.

Ces quatre exemples permettent d'appréhender la manière dont sont définies les données communicables dans chacun de ces secteurs.

TRANSPORTS

En matière de droit des données appliqué au secteur des transports, **la loi n° 2015-990 du 6 août 2015 pour la croissance, l'activité et l'égalité des chances économiques** prévoit précisément que les « données des services réguliers de

Données essentielles : conception d'un format en mode collaboratif

Lors d'un Barcamp organisé le 9 mai 2016 par le SGMAP et Breizh Small business act, la communauté des acheteurs et des data scientists a pu apporter sa contribution à la conception d'un format pivot de description des données de marchés publics, en vue notamment de leur ouverture et de leur interopérabilité.

Il permettra d'exploiter les données essentielles publiées, prévues à [l'article 107 du décret n° 2016-360 du 25 mars 2016](#). Pour l'instant, il est soumis à concertation pour enrichissement.

transport public de personnes et des services de mobilité » sont les données numériques relatives aux arrêts, aux tarifs, aux horaires, et à l'accessibilité, notamment.

Par ailleurs, la loi a introduit un nouveau Chapitre V au sein du code des transports intitulé « *l'accès aux données nécessaires à l'information du voyageur* » ayant pour objet d'imposer l'ouverture des données des transports publics.

PRINCIPE	Consécration du principe selon lequel les données des services de la mobilité sont diffusées « <i>librement, immédiatement et gratuitement en vue d'informer les usagers</i> » et de leur fournir le meilleur service public de transport.
DÉROGATION	Les exploitants et AOM sont toutefois réputées remplir leurs obligations dès lors qu'elles sont adhérentes à des « codes de conduite, des protocoles ou des lignes directrices » préalablement établis par elles et rendus publics, pour autant que ces documents établissent les conditions de diffusion et d'actualisation des données et qu'ils ont été homologués par les ministres chargés des transports et du numérique.

SANTE

La loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé précise que les données de santé correspondent aux différentes données issues des systèmes d'information des établissements publics et privés de santé et du système national d'information inter-régimes de l'assurance maladie, notamment.

Par ailleurs, trois éléments relatifs aux données ressortent de la loi, à savoir :

- La définition de principes relatifs à la mise à disposition des données de santé rappelant que les données de santé à caractère personnel recueillies « *à titre obligatoire et destinées aux services ou aux établissements publics de l'Etat ou des collectivités territoriales ou aux organismes de sécurité sociale peuvent faire l'objet de traitements à des fins de recherche, d'étude ou d'évaluation présentant un caractère d'intérêt public* » dans le respect de la loi CNIL ;
- La création d'un système national des données de santé, lequel rassemble et met à disposition les données de santé afin de contribuer, notamment, à l'information sur la santé ainsi qu'à la mise en œuvre des politiques de santé et de protection sociale. La loi précise également que les données du système national des données de santé qui sont mises à la disposition gratuite du public sont traitées pour prendre la forme de statistiques agrégées ou de données individuelles sans identification possible ;
- La création d'un groupement d'intérêt public, dénommé « *Institut national des données de santé* », constitué entre l'Etat, des organismes représentatifs des usagers publics et privés du système de santé et des producteurs de données de santé en charge, notamment, de la qualité des données de santé et des conditions de leur mise à disposition.

Un événement a toutefois marqué la fin du mois de mai et risque de modifier l'édifice prévu par la loi promulguée en janvier. En effet, le Conseil d'Etat a demandé au Ministère de la Santé d'annuler (sous quatre mois) un arrêté qui interdisait jusque-là l'accès aux données de santé (Système national d'information inter-régimes de l'assurance maladie – Sniiram) aux organismes de recherche à but lucratif. Ce qui signifie que tout organisme privé effectuant une recherche pourra accéder à l'une des plus grandes bases de données médico-administrative au monde¹⁵.

Des gardes fous pourront être mis en place pour garantir que les données personnelles seront correctement anonymisées, notamment via le futur Institut national des données de santé (INDS), puis un comité scientifique chargé d'examiner les protocoles scientifiques utilisés avant que la CNIL ne se prononce en dernière instance sur le respect de la vie privée. Mais pour cela il conviendra d'attendre la publication des décrets fixant les prérogatives de ces trois instances.

ENERGIE

La loi n° 2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte (loi TECV) prévoit que les données des secteurs de l'électricité, du gaz et des réseaux de chaleur sont constituées des données issues des systèmes de comptage d'énergie et des données de consommation.

Par ailleurs, la loi a introduit de nouvelles dispositions dans le code de l'énergie ayant pour objet d'imposer l'ouverture des données des gestionnaires du réseau public de transport d'électricité, des réseaux publics de distribution d'électricité, des réseaux du gaz et des réseaux de chaleur dans les conditions suivantes :

PRINCIPE	Les données relatives à la production, au transport, à la distribution, à la consommation d'électricité, de gaz ou de production et de consommation de chaleur sont mises à disposition des personnes publiques par les gestionnaires des réseaux dès lors que les données sont utiles à l'accomplissement des compétences exercées par ces personnes publiques (pour la réalisation des plans climat-air-énergie territoriaux notamment).
DÉROGATION	Chaque gestionnaire de réseau préserve la confidentialité des informations d'ordre économique, commercial, industriel, financier ou technique dont la communication serait de nature à porter atteinte aux règles de concurrence libre et loyale et de non-discrimination.

Des décrets devraient intervenir pour préciser les personnes publiques bénéficiaires des données, la nature des données mises à disposition, la maille territoriale à laquelle les données sont mises à disposition et les modalités de cette mise à disposition.

¹⁵ Cette base fait transiter chaque année 1,2 milliard de feuilles de soins, 500 millions d'actes médicaux et 11 millions de séjours hospitaliers.

RENSEIGNEMENT

La loi n° 2015-912 du 24 juillet 2015 relative au renseignement précise que « les données de connexion » correspondent à une liste de données techniques recueillies auprès des opérateurs de communications électroniques (numéro d'abonnement, connexion à des services de communications électroniques, localisation des équipements terminaux, notamment).

Cette loi a par ailleurs, inséré un nouveau chapitre intitulé « *Des accès administratifs aux données de connexion* » au sein du Code de la sécurité intérieure, au terme duquel les « données de connexion » correspondent aux données recueillies auprès des opérateurs de communications électroniques (données techniques relatives à l'identification des numéros d'abonnement ou de connexion à des services de communications électroniques, au recensement de l'ensemble des numéros d'abonnement ou de connexion d'une personne désignée, à la localisation des équipements terminaux utilisés ainsi qu'aux communications d'un abonné portant sur la liste des numéros appelés et appelants, la durée et la date des communications).

CADRE JURIDIQUE NATIONAL RELATIF AUX DONNEES - SYNTHESE

Juridique

En synthèse, sont éparpillées dans différents textes communautaires et nationaux (codes, lois, ordonnances, décrets) de nombreuses dispositions se rapportant aux données publiques aboutissant :

- **Au cumul de définitions « sectorielles » correspondant généralement à l'inventaire de données existantes et/ou issues des SI des exploitants. Tel est le cas des données suivantes :**
 - Dans la loi santé, les **données de santé** correspondent aux différentes données issues des SI des établissements publics et privés de santé et du système national d'information inter-régimes de l'assurance maladie, notamment ;
 - Dans la loi Macron, les « **données des services réguliers de transport public de personnes et des services de mobilité** » correspondent aux données numériques relatives aux arrêts, aux tarifs, aux horaires, à l'accessibilité, notamment ;
 - Dans la loi TECV, les données des secteurs de l'électricité, du gaz et des réseaux de chaleur semblent cantonnées aux **données issues des systèmes de comptage d'énergie et aux données de consommations** ;
 - Dans la loi renseignement : « **les données de connexion** » correspondent à une liste de données techniques recueillies auprès des opérateurs de communications électroniques.
- **A la multiplication des concepts et des obligations mises à la charge des partenaires de l'administration mais à l'absence de définition claire et précise de certaines notions telles que :**
 - « **les données essentielles** » (ordonnances et décrets concessions et marchés publics) ;
 - « **les données de référence** » (Loi NOTRe).

Viennent s'y ajouter les nouvelles notions introduites par la loi pour une République Numérique :

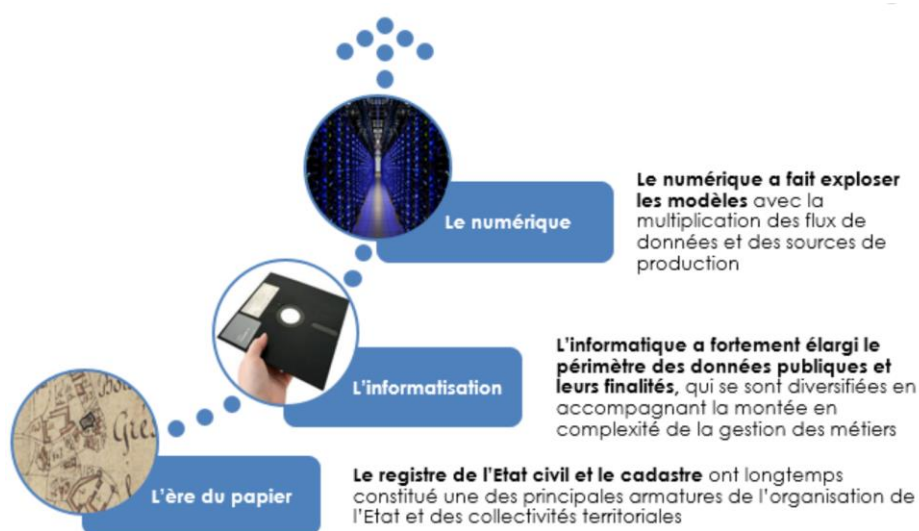
- **La notion de données de références (article 14) qui doit satisfaire les conditions suivantes, à savoir qu'il doit s'agir :**
 - de références communes pour nommer ou identifier des produits, des services, des territoires ou des personnes,
 - réutilisées fréquemment par des personnes publiques ou privées autres que l'administration qui les détient et
 - dont la réutilisation nécessite qu'elles soient mises à disposition avec un niveau élevé de qualité, notamment en termes de précision, de disponibilité ou de fréquence de mise à jour.
- **La notion de données « d'intérêt général » (articles 17 et 18), il s'agit notamment :**
 - soit des « données et bases de données collectées ou produites à l'occasion de l'exploitation du service public » par le délégataire et mises à disposition du délégant,
 - soit des données « essentielles » de la convention de subvention conclue par l'administration avec un partenaire privé.

Ainsi que le relevait la CNIL dans son avis sur le projet de loi pour une République numérique, il devient indispensable d'harmoniser ces différentes notions.

C. Secteur public d'Etat : de l'open data à la gouvernance par la donnée

1. Transformation des modes de production numériques et diversification des usages

➤ Evolution de la gestion des données dans les administrations



Depuis l'invention du papier et des registres, les collectivités ont toujours cherché à classer et à conserver les données concernant les habitants ou la distribution des parcelles sur leur territoire. Aujourd'hui, l'augmentation du nombre de données, et en parallèle, la montée en puissance des ordinateurs, ont installé la donnée au cœur des processus de gestion.

Selon Henri Verdier, Administrateur général des données de l'Etat, la révolution numérique vit son troisième âge :

- D'abord la révolution informatique, dans les années 1980, avec l'augmentation phénoménale de la puissance de calcul des ordinateurs,
- Après la révolution Internet qui permet la mise en réseau des ordinateurs dès le début des années 1990 et l'émergence du web 2.0,
- La révolution de la donnée, que nous traversons actuellement, place la création de valeur du côté des données et non du côté des outils permettant de les manipuler.

➤ Nouveaux modes de production et nouvelles logiques d'action publique

Pour Elisabeth Grosdhomme-Lulin, ancienne haut fonctionnaire et directrice d'une société d'études et de conseil consacrée à la prospective et à l'innovation, le secteur public peut avoir recours au numérique pour mener à bien trois types de finalités :

- La **recherche d'économies** grâce à la simplification des procédures et à la mutualisation de ressources et des compétences ;
- La recherche la performance pour fournir un **service public plus efficace et plus personnalisé** grâce à la collecte et au traitement de nombreuses données personnelles ;
- **L'activation de la citoyenneté** en créant par exemple des outils d'entraide de proximité, de crowdsourcing, de consommation collaborative et de participation citoyenne (ex : budget participatif de la Ville de Paris).

À travers ces trois finalités, on note la place importante de la gestion des données dans l'administration publique et surtout de sa propre révolution. En effet, on assiste à une **diversification sans précédent des sources et des modes de production des services publics** en lien avec l'informatisation des SI (et notamment la messagerie électronique), la numérisation des services publics (ex : nouvelle carte de vie quotidienne), l'Internet des objets (ex : capteurs, wearables) et la coproduction avec les usagers (cf. *Fiche notion Crowdsourcing en annexe*). Cela s'accompagne de **nouvelles logiques pour l'action publique** qui s'appuient désormais sur des politiques « consommatrices » de données et sur la coopération et l'échange de données entre administrations (notamment dans le cadre de programmes comme *"Dites-le nous une fois"* ou *FranceConnect*).

➤ Usages et services innovants autour de la donnée

L'explosion des données produit de nouveaux usages et services : c'est l'un des principaux arguments qui justifie l'investissement sur ces services. Ces nouveaux usages concernent à la fois les problématiques de gestion d'un territoire et de gestion de la relation citoyen :

Les nouveaux usages dans la gestion	Monitoring des services urbains : projet de monitoring environnemental à Nice à partir de données issues de capteurs (ex : capteurs embarqués sur les tramway),
	Data Science et analyse prédictive : • Analyse de la consommation d'électricité sur les bâtiments publics pour nourrir une stratégie d'achat optimisée, • Analyse des comportements de mobilité et prédiction des disponibilités des vélib, • Prédiction des probabilités d'embauche par secteur...
De nouveaux services aux usagers	Suivi temps réel de l'offre de transport et d'autres services urbains : application Optimod'Lyon, Nantes dans ma poche, suivi des consommations énergétiques en temps réel...
	Services personnalisés : plateformes de transport à la demande (MobiSavoie)
	Simplification : programme <i>"Dites-le Nous Une Fois"</i> ...

➤ Partage et réutilisation : un levier de création de valeur

La valeur ajoutée de ces nouveaux usages et services est intrinsèquement liée au partage et à la réutilisation de données. En effet, la circulation des données permet de créer de la valeur selon les cinq axes ci-dessous :

Réduction des coûts de transaction	La réutilisation comme critère de performance	Base de données touristique APIDAE : 1000 producteurs d'information et 1000 projets ré-utilisateurs des données de l'entrepôt 
	Un bénéfice social décuplé	Gratuité du RGE de l'IGN : multiplication par 20 des volumes de données téléchargées, soit un bénéfice social estimé à 114 M€ par an 
Transparence de l'information	Des asymétries d'information aplanies	Mise en ligne des marchés publics : tous les répondants disposent de la même information
Action sur les comportements	Des comportements adaptés à la situation grâce au partage d'information	Compteurs électroniques : la remontée d'information permet aux habitants de modifier leur consommation et d'agir notamment sur les pics de consommation et les fuites d'eau 
Innovation	Une base de connaissances incomparable décuplée si combinée aux données du web	Data Publica : une base de connaissances détaillée des entreprises françaises grâce à l'analyse de 600 000 sites web combinée avec les informations désormais ouvertes (Infogreffe, DILA, INSEE, INPI) 
Collaboration	Des partenariats pour produire, enrichir et améliorer les données	Base adresse nationale (BAN) un chantier national qui associe l'IGN, la Poste et les contributeurs bénévoles de OpenStreetMap 

➤ Des freins qui limitent l'utilisation de ces leviers

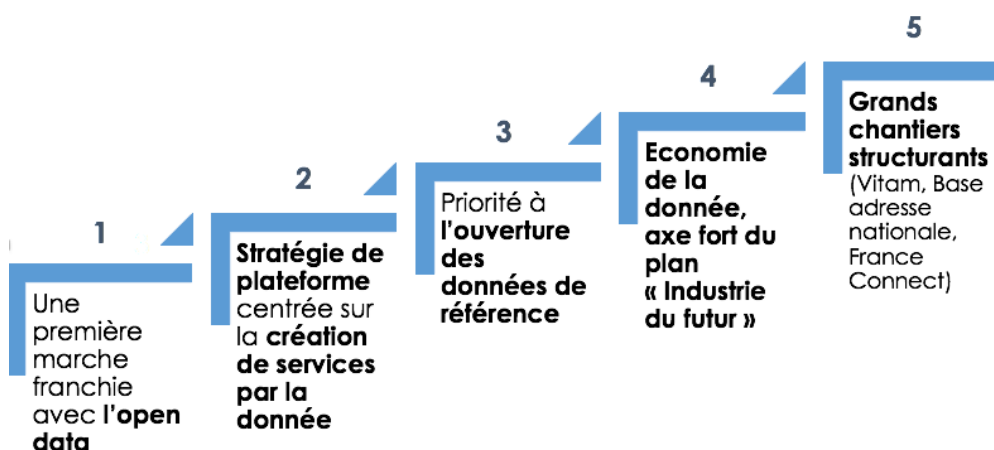
La circulation des données est néanmoins parfois limitée dans le secteur public en général du fait de freins inhérents au déploiement de projets centrés sur la donnée :

- **Manque de culture de la donnée lié à un problème de compétences,**
- **Coût du déploiement des solutions,**
- **Questions sur la sécurité des données et sur la protection de la vie privée.**
- **Manque de transversalité dans l'approche des projets**

Il en va de même au niveau des collectivités locales : l'enquête réalisée par la FNCCR auprès de ses adhérents en janvier 2016 révèle que près de 50 % des répondants ont retenus quatre freins majeurs au déploiement de projets de type big data : le coût d'investissement, le coût d'exploitation, les enjeux de sécurité des données et les enjeux de protection de la vie privée.

2. Un nouvel écosystème public national en gestation

L'Etat est depuis toujours producteur de données pour sa propre gestion et pour répondre aux besoins de la société (ex : INSEE, IGN, INSERM, etc.). Dans les années 1980, l'Etat et les banques sont les premiers à se lancer dans le développement informatique à grande échelle. **L'Etat est aujourd'hui en première ligne sur les questions de gouvernance de la donnée**, à travers les actions mises en œuvre par ses différents services :



➤ Une première marche franchie avec l'open data

Le gouvernement attache une grande importance à l'ouverture et au partage des données publiques. Cette politique est à la fois un axe essentiel de la modernisation de l'action publique et un important levier de stimulation du dynamisme économique et de l'innovation. Cette priorité a été inscrite dans divers documents au fil des années :

- La Charte de déontologie du 17 mai 2012 signée par tous les membres du gouvernement lors du premier Conseil des ministres,
- La feuille de route stratégique adoptée lors du séminaire gouvernemental sur le numérique du 28 février 2013,
- La Charte du G8 pour l'ouverture des données publiques adoptée le 18 juin 2013 par les chefs d'Etat et de gouvernement du G8¹⁶.

Sur le plan opérationnel, le décret du 21 février 2011 crée la mission « Etalab », chargée de la création et de l'animation d'un portail unique interministériel des données publiques pour l'hébergement et la diffusion de ces données. Ainsi, dès 2011, Etalab lance le portail français de l'open data, data.gouv.fr, qui propose à son ouverture environ 330 000 jeux de données (provenant surtout de l'INSEE) et accueillera progressivement des données publiques provenant des administrations, des établissements publics et de certaines collectivités territoriales.

etalab^{gouv.fr}

¹⁶ <http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/charte-g8-ouverture-donnees-publiques-fr.pdf>

Intégré au SGMAP¹⁷ en 2012, Etalab devient un service du Premier ministre dont les objectifs sont précisés dans la feuille de route de 2013 :

- Transparence accrue de l'action publique pour améliorer le fonctionnement démocratique,
- Amélioration du fonctionnement administratif et de l'efficacité de l'action publique,
- Proposition de nouvelles ressources favorables à l'innovation économique et sociale et à la croissance,
- Coproduction avec les citoyens d'innovation sociale (ex : projet Handimap développé à Rennes pour faciliter les déplacements de personnes handicapées moteurs).

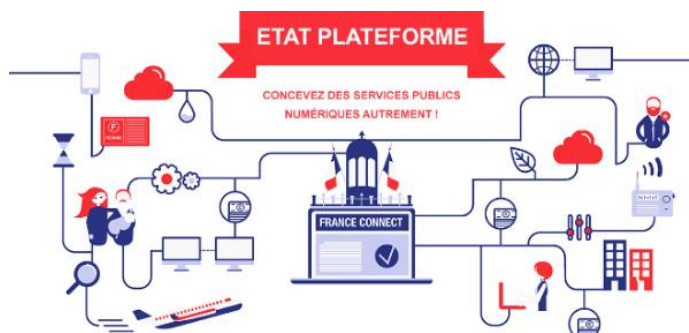
Depuis janvier 2013, Etalab est dirigé par Henri Verdier, devenu depuis Administrateur Général des Données et a fusionné en septembre 2015 avec la DISIC (Direction interministérielle des systèmes d'information et de communication pour former la DINSIC (direction interministérielle du numérique et du système d'information et de communication de l'État).

Etalab mène aujourd'hui les actions suivantes :

Gestion du portail data.gouv.fr	Le portail rassemble et met à disposition des données publiques issues de l'Etat, de ses établissements publics et, si elles le souhaitent, des collectivités territoriales et des personnes chargées d'une mission de service public. 600 organisations au sein de la communauté, 21 000 jeux de données, 1 300 réutilisations.	
Mise à disposition gratuite des données publiques	Conformément au principe général de réutilisation libre, facile et gratuite, en mettant l'accent sur les données à fort impact sociétal et/ou à fort potentiel d'innovation économique et sociale.	
Animation et Incitation à la réutilisation des données publiques	Concours Dataconnexions, qui distinguent des projets innovants de réutilisation de données publiques. 200 startups à fort potentiel identifiées grâce au concours depuis 2012.	

¹⁷ Secrétariat général pour la modernisation de l'action publique, créé par Jean-Marc Ayrault le 31 octobre 2012

➤ *Une stratégie de plateforme centrée sur la création de services par la donnée*



Si le décret du 1^{er} Août 2014, entérine le principe d'un système d'information unique de l'Etat, l'ambition est bien de faire émerger une informatique d'État conforme aux meilleures pratiques actuelles. La stratégie d'«État plateforme», élaborée par la DINSIC et les DSI ministérielles en 2014-2015, pose les fondamentaux d'une nouvelle architecture numérique privilégiant l'interopérabilité des systèmes fondée sur une logique de plateforme et d'API. Cette stratégie de transformation du SI de l'Etat se donne pour objectif de proposer des services centrés sur ses besoins, et non plus issus du découpage des structures administratives.

C'est pourquoi le SGMAP mène actuellement de nombreux projets basés sur des processus agiles et qui partent des besoins des utilisateurs pour créer des services originaux et construire de nouveaux écosystèmes. En effet, l'objectif est bien de faire émerger de nouveaux services publics numériques, résultant d'une meilleure circulation des données entre les administrations, et placés sous le contrôle des usagers.

Simplification des démarches administratives	FranceConnect, fédération d'identités ouverte à toutes les administrations.	
	Dites-le nous une fois, pour faire en sorte que les entreprises ne fournissent plus qu'une fois les justificatifs à l'administration publique. Cela suppose des systèmes complexes d'échanges entre administrations avec le consentement de l'utilisateur.	
Stratégie d'«APIfication»	Développement d'API et d'interfaces d'accès aux données : l'Etat interconnecte ses grandes bases de données qui sont des entrepôts à part entière.	

Quelle vision de l'Etat plateforme et de ses futurs développements ?

Synthèse de l'interview d'Henri Verdier, Administrateur général des données
12 avril 2016

Pour mener à bien « la transformation numérique de l'Etat », la DINSIC a fait le choix d'exploiter les « stratégies de plateformes » qui consistent à « **mettre à disposition des ressources aisément réutilisables, via des interfaces de programmation (API) pour produire de nouveaux services.** Les API fournissent l'ensemble des instructions pour qu'une application puisse se "brancher" facilement sur une base de données ou une autre brique logicielle. » Cette organisation permet « d'organiser et de stimuler la capacité d'innovation des tiers (autres administrations, société civile) et éventuellement de mettre en place des régulations douces ».

Les API développées par la DINSIC répondent chaque fois à un « problème concret » : proposer de nouvelles applications pour les chauffeurs de taxi face à la concurrence des VTC (ex : API Le Taxi), unifier les bases d'adresses (la BAN), apporter de la visibilité sur le marché de l'emploi local (ex : La Bonne Boite), simplifier les réponses à appel d'offre et les échanges entre administrations (ex : MPS), etc.

La question des données privées est bien souvent un enjeu majeur dans le développement des API. En effet, certains transferts de données ne peuvent se faire qu'avec le consentement de l'utilisateur. « Cette question du consentement est centrale dans un monde où nous entrons, dans lequel les usagers demanderont de plus en plus de services intégrés et efficaces. La construction des API en cours, qui place le consentement au cœur de l'architecture (« privacy by design »), fonde des principes extrêmement sains, définis avec la

➤ *Priorité à l'ouverture des données de référence*

Depuis début 2014, **l'Etat poursuit une stratégie d'ouverture de bases de données nationales**, référentiels incontournables pour les acteurs du territoire, qu'ils soient publics ou privés.

Ainsi, plusieurs ministères et agence nationales, grands producteurs de données publiques, ont mis à disposition des bases de données comme l'indique le tableau ci-dessous :

Ministère de l'intérieur	La liste des maires au 17 juin 2014
Ministère de la justice	Les condamnations (statistiques à partir du casier judiciaire nationale)
Ministère de l'économie, de l'industrie et du numérique	Le prix des carburants en France depuis 2007
ADEME	Les émissions de CO2 et de polluants des véhicules commercialisés en France
Ministère des affaires sociales et de la santé	La base de données publiques du médicament
Ministère des affaires étrangères	L'aide publique au développement de la France

DILA	Les bases de données économiques liées aux marchés publics et aux annonces officielles (BOAMP et BALO depuis 2005 et BODACC depuis 2008)
INPI	Les données de propriété industrielle (marques, brevets et dessins)
RNCS	Les données Infogreffe

Du côté de l'INSEE, il ne manquait que le fichier SIRENE (Système informatique pour le répertoire des entreprises et des établissements), pour offrir une vision complète des données sur les entreprises françaises (numéro Siren ou Siret, localisation, effectifs, secteur d'activité, date de création, etc.). La base sera ouverte le 1er janvier 2017. Les entreprises innovantes du marché de l'information seront ainsi en mesure de fournir de nouveaux services à valeur ajoutée aux entreprises et aux territoires, à l'image des services de marketing prédictifs déjà fournis par Data Publica.

De plus, la mise à disposition gratuite du répertoire met fin à un système dans lequel les organisations publiques devaient acheter ces données auprès de l'INSEE, ce qui devrait favoriser l'usage de cet outil de référence et simplifier son utilisation.

D'autres **référentiels stratégiques ont été ouverts gratuitement aux collectivités et à leurs prestataires**. Ainsi, depuis le 3 janvier 2011, le référentiel à grande échelle (RGE) de l'IGN est utilisable sans contrepartie pour l'exercice de missions de service public ne revêtant pas un caractère industriel ou commercial.

Le SGMAP et la DINSIC développent par ailleurs divers projets pour faciliter les échanges de données confidentielles entre administrations (particuliers et entreprises), avec le consentement de l'utilisateur. **Divers programmes reposant sur des API ont en effet été mis en place**. Contrairement au modèle papier, ils permettent de ne rapatrier que les données strictement nécessaires au traitement de la procédure engagée. Plusieurs services sont développés sur ce principe par le SGMAP et la DINSIC :

- **Marché Public Simplifié (MPS)** : ce dispositif du programme « Dites-le-nous une fois » permet aux entreprises de candidater à un marché public avec son seul numéro de SIRET. La place de marché récupère les informations de l'entreprise via une plateforme gérée par le SGMAP ;
- **FranceConnect, officiellement lancé fin juin** est un dispositif de fédération d'identité au niveau national qui permet aux internautes de s'identifier et de s'authentifier auprès d'une administration à partir de leurs comptes existants : Sécurité Sociale, impôts...
- **Téléprocédure simplifiée** est un générateur encore expérimental de démarches en ligne pour les administrations. Les administrations fabriquent leur formulaires électroniques en utilisant des champs simples et quelques autres dotés d'API connectées à des bases de données. Avec cet outil les champs "adresse" proposé à l'utilisateur pourra être vérifié et validé via la base adresse nationale et éventuellement corrigé en temps réel sur les erreurs commises lors de la saisie.

UN MARCHÉ PUBLIC SIMPLIFIÉ, ÇA SE PASSE COMMENT ?



➤ *Economie de la donnée, axe fort du plan « industrie du futur »*

Depuis quelques années, le gouvernement place la donnée au cœur des enjeux de développement économique. L'économie de la donnée devient même un axe stratégique du plan « Industrie du futur » de la Nouvelle France Industrielle et l'un des leviers mentionnés par Emmanuel Macron, Ministre de l'Economie, de l'Industrie et du Numérique, dans la présentation de la feuille de route du projet de loi NOE.



La Direction Générale des Entreprises (DGE) poursuit également un objectif d'amélioration de la gestion et de la valorisation des données dans les entreprises afin de mieux structurer le secteur sur ces nouvelles approches de gouvernance.



Elle mène avec d'autres partenaires, plusieurs actions destinées à dynamiser ce secteur par des initiatives emblématiques à valeur d'exemple.

Création d'un observatoire des usages du big data	Observatoire porté par la Direction générale des entreprises et Tech in France dont le lancement est prévu fin 2016, avec une analyse de maturité par secteur impacté	
Mise en place d'une plateforme de données agricoles	Cette plateforme nationale rassemblera les données du secteur privé, comme celles des fermes, des coopératives pour proposer un écosystème, créer des innovations de ruptures à l'instar d'API-Agro plateforme de référence agronomique ¹⁸ .	
Développement du datatourisme	Grâce au déploiement d'une plateforme nationale de données touristiques ouverte aux acteurs privés, de type APIDAE	
Projet de place de marché de services sur le cloud computing	Le projet s'inspire du modèle "g-smart" de la Grande Bretagne et vise à proposer des services et applications cloud aux administrations locales et nationales, via le cloud. Objectif mettre en place un "Proof of concept" d'ici la fin 2016	
Renforcement de l'animation	Le challenge big data fait suite à un AMI qui a permis de sélectionner des animateurs (Capdigital, Numa, Tuba).	

¹⁸ La finalité du projet est de construire une plateforme informatique favorisant l'interopérabilité des références agronomiques des partenaires du projet (INRA, Agro-Transfert R&T et la plupart des instituts techniques agricoles).

➤ De grands chantiers structurants

L'Etat a lancé plusieurs chantiers stratégiques qui feront partie du socle numérique nécessaire à la gestion des données. Ces chantiers ne sont pas à la seule destination de l'Etat et devraient s'adresser à l'ensemble du secteur public.

Sur la partie archivage, par exemple, l'Etat souhaite proposer une solution logicielle d'archivage électronique unique à travers le programme **VITAM**, piloté par le SGMAP et lancé par les ministères des Affaires Etrangères, de la Culture et de la Défense, pour développer un socle d'archivage électronique réutilisable par les administrations de l'Etat. En effet l'objectif d'ici mi-2019 est de travailler sur la conception, la réalisation et la maintenance mutualisée d'une solution logicielle de backoffice, sous licence libre, réutilisable dans différents contextes d'archivage. Pour y parvenir les actions suivantes sont actuellement menées dans les trois ministères :

- Définir une stratégie de collecte, de traitement et de conservation des documents (y compris les messageries électroniques), adaptée aux différents contextes des partenaires du projet,
- Identifier les outils et fonctionnalités techniques nécessaires pour un traitement technique de ces documents, en amont de leur versement dans la solution Vitam,
- Définir un modèle de métadonnées pour les documents.

L'outil pourrait à terme être mis à disposition des collectivités. En effet, l'archivage électronique est un enjeu pour les collectivités et la loi est en train d'évoluer avec l'assouplissement des conditions de mutualisation de l'archivage électronique.

VITAM pourrait par ailleurs devenir un outil big data pour l'Etat et les administrations centrales.

La création d'une **Base Adresse Nationale (BAN)**, ayant pour but de référencer l'intégralité des adresses du territoire français, est également l'un des chantiers structurants portés par la DINSIC : la base de données contient actuellement la position géographique de plus de 25 millions d'adresses. Elle est constituée en collaboration par :

- Des acteurs nationaux tels que l'IGN et La Poste,
- Des acteurs locaux tels que les collectivités, les communes, les SDIS,
- Des citoyens par exemple à travers le projet OpenStreetMap et l'association OpenStreetMap France.

Le projet est co-gouverné par l'Administrateur Général des Données et le Conseil National de l'Information Géographique.

➤ Un cadre structurant défini par la doctrine de la CNIL sur le Big Data

Les nouvelles technologies de big data engendrent un recueil massif des données qui semblent aller à l'encontre des principes posés en matière de protection des données personnelles, notamment par la CNIL.



De par la nature massive et indéterminée de la collecte des données, le « big data » remettrait notamment en cause les principes de pertinence et le respect de la finalité de la collecte des données qui imposent que la collecte de données soit strictement nécessaire à la poursuite de finalités déterminées, ce qui suppose notamment que les usages aient été définis préalablement¹⁹.

La CNIL ne retient pas une approche très stricte du principe de finalité. La loi Informatique et Libertés prévoit notamment la possibilité de présomption de compatibilité en cas de traitement ultérieur des données à des fins statistiques ou sous réserve que ce traitement ne soit pas utilisé pour cibler un individu.

La CNIL veille également au respect de ce principe en vérifiant que le traitement des données, notamment algorithmique, ne conduise pas à la prise de décisions discriminatoires vis-à-vis des individus.

Des possibilités de réutilisation massive des données sont également permises par le biais de l'anonymisation.

La notion de données à caractère personnel rend nécessaire l'anonymisation de celles-ci à des fins de réutilisation. La CNIL adopte une approche flexible en la matière : le seuil retenu pour l'anonymisation des données dépendant du type de données concernées.

Ainsi, en ce qui concerne les données de consommation énergétique par exemple, la maille retenue est celle de l'immeuble de dix foyers au moins. Des règles plus strictes sont imposées en matière de traitements des données détenues par les opérateurs de télécommunications.

Cette approche a également vocation à évoluer en accord avec les évolutions des technologies de chiffrement des données permettant de « dépersonnaliser » les données personnelles.

D. Collectivités : début de structuration sur la gestion des données

1. Etat des lieux général

La moitié des décideurs des collectivités locales s'appuieraient désormais sur des outils de reporting (46%) et sur des tableaux de bords décisionnels (51%), même si ces solutions sont le plus souvent sous-utilisées en interne, confirme une enquête Markess de septembre 2015. D'ici à 2017, 23 % des décideurs interrogés dans le cadre de

¹⁹ L'analyse du cadre juridique du traitement des données personnelles, des obligations des collectivités territoriales et des algorithmes est développée dans au sein des annexes du présent rapport.

cette étude souhaitent investir dans des tableaux de bords décisionnels et dans des solutions d'analyse de la qualité des données.

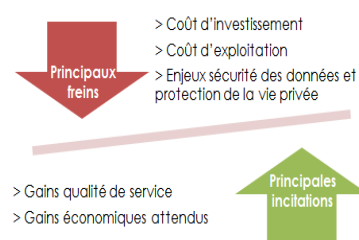
Sur un autre registre qui touche également l'aide à la décision, les dispositifs numériques connectés (de type capteurs) permettant de collecter des données, vont également progresser, puisque, cette fois 75% des répondants à une enquête flash, réalisée en janvier 2016, par la FNCCR²⁰, mettent en place ou prévoient de mettre en place des dispositifs numériques connectés.

La place des data sciences et du big data demeure plus mesurée, puisque 8% des répondants à l'enquête FNCCR déclarent utiliser ou avoir pour projet d'utiliser ces technologies. Selon Markess, le décollage ne se produirait pas avant 2017 et sur les modes opératoires suivants :

- Mise en corrélation de données (39%),
- Analyse de données non structurées (35%),
- Analyse de données structurées (30%),
- Analyse en temps réel (26%),
- Visualisation de données (23%).

Dans la hiérarchie des freins au développement des data-sciences, les répondants placent les difficultés de mise en œuvre d'une gouvernance de la donnée largement devant les questions d'appropriation des technologies de big data. La complexité du système d'information des collectivités et tout particulièrement celui des communes en est une des principales causes. Il n'est pas rare de voir les plus grandes villes gérer des portefeuilles supérieurs à 250 voire 300 applications. Cette diversité complique et ralentit la transition des systèmes d'information mais elle ne l'arrête pas. Les collectivités territoriales commencent en effet à s'organiser pour conduire cette mini-révolution par la donnée. Aujourd'hui quatre leviers de transformation sont esquissés dans les territoires :

- **Le mouvement open data**, né à la fin des années 2000 demeure un des principaux catalyseurs de la prise de conscience (progressive) par les collectivités territoriales de la valeur ajoutée de leurs données. Son développement reste à ce jour toujours limité.
- **Les projets de gouvernance de la donnée**, font une apparition encore timide, presque exclusivement portés par les métropoles. Mais en recentrant les systèmes d'information sur la réutilisation, le contrôle et la diffusion des données, ils vont constituer l'élément central dans la transformation numérique du territoire.
- **Les data sciences**, apportent un nouveau potentiel d'innovation dans la mise en œuvre des politiques publiques locales, dans la gestion urbaine, dans la meilleure connaissance des usagers et de leurs besoins, mais leur usage reste circonscrit aux grandes agglomérations.
- **Les infrastructures de données publiques**, qui incluent réseaux de collecte et plateformes de données mutualisées vont cibler l'intégration des entités publiques ne disposant pas de solutions autonomes dans l'accomplissement



20 Enquête réalisée auprès des adhérents de la FNCCR.

de leurs obligations légales. Elles demeurent également embryonnaires sauf dans les Syndicats informatiques et les autres structures de mutualisation informatique (SMI) qui existent dans environ 1/3 des territoires et doivent être développées ailleurs.

2. Open data : à la recherche d'une consolidation au niveau local

L'open data progresse dans les collectivités, mais dans des proportions modestes. Le nombre de portails publics locaux en France est estimé à une soixantaine. Parmi ceux-ci, une bonne moitié provient des villes et on dénombre également une vingtaine de plateformes mutualisées accueillant d'autres collectivités. C'est peu au regard de la fragmentation territoriale : l'ensemble représenterait près de 300 communes, soit à peine 10% des 3000 communes de plus de 3500 habitants qui auront bientôt par voie législative, obligation à ouvrir leurs données. L'entrée en vigueur de la loi pour une République numérique devrait, par voie de conséquence, fortement amplifier cette présence, d'où l'intérêt de dresser un état des lieux, pour mesurer la capacité des collectivités à passer la vitesse supérieure.

➤ *La marque de fabrique métropolitaine*

Dans les métropoles et les très grandes villes, l'open data devient partie intégrante du système d'information avec le remplacement progressif des portails par des plateformes open data. Même si l'idéologie et le marketing public ont parfois masqué les véritables enjeux, l'ouverture des données apparaît désormais comme une action à forte valeur ajoutée. Plus seulement dans l'optique de création de nouveaux services par des startup avisées – ce qui n'a pas toujours répondu aux attentes – mais plutôt au regard du potentiel de réutilisation et dans une optique par défaut.

Dans les métropoles, la fusion progressive des systèmes d'information résultant de la unification des DSI (ville-centre et EPCI), va conduire à une généralisation de plateformes mutualisées. Le Grand Lyon est déjà dans cette configuration (sans DSI unifiée) et Bordeaux, après avoir essuyé un premier échec, devrait réunir avant la fin 2016 sur la même plateforme, les données de la Métropole, la ville centre et huit autres communes du territoire.

Si l'open data reste le principal levier de transformation des politiques publiques autour de la donnée, quelques nuances apparaissent dans les stratégies locales entre des agglomérations qui comme Rennes Métropole, considèrent toujours cette approche comme la clé de voûte de leur politique et d'autres qui la repositionnent comme un élément "parmi d'autres" dans une stratégie plus ambitieuse de gouvernance de la donnée. Les nuances s'expliquent généralement par l'impact réel obtenu sur le terrain.

Rennes tire un bilan positif de ses cinq années d'activisme. L'open data demeure le levier privilégié de sensibilisation du public et des entreprises : il contribue à amplifier la réutilisation et à générer de nouveaux usages. Historiquement les politiques locales ont donné lieu à la création d'une cinquantaine d'applications - sur smartphone

OpenData France, association des collectivités



L'association réunit une quarantaine d'adhérents (villes, EPCI, départements, régions) et travaille sur divers sujets relatifs à la standardisation :

Marchés publics : insertion de clauses "open data" que les villes insèrent dans leurs cahiers des charges et contrats, puisque à défaut de lois ce dernier s'impose.

Normalisation : des groupes de travail se réunissent dans le but de construire des modèles de données. Cinq domaines : budgets, subventions, état-civil, établissements publics et délibérations.

Catalogue : l'association prévoit de s'investir rapidement sur les catalogues de méta-données.

Licences : l'association réécrit les licences ODBL avec Paris et Toulouse.

principalement – dédiées principalement à la mobilité, et les créations se poursuivent aujourd'hui. Sur un volet plus politique, la transparence de l'action publique reste également une forte exigence. Rennes a été une des premières villes à ouvrir ses données budgétaires et à les associer à des outils de data visualisation pour en accroître la lisibilité et l'accessibilité²¹. D'autres initiatives plus pédagogiques réalisées à partir des données Insee, dressent le portrait des habitants du territoire²². Toujours dans le registre de l'innovation, la métropole prépare un **tableau de bord ouvert des politiques publiques** dont la mise à jour régulière nécessitera un important travail de collecte et la participation des citoyens.

Plus nuancée dans son approche, Bordeaux Métropole reste mobilisée sur l'open data mais tire un bilan mitigé de son impact en termes de création directe d'emploi. L'agglomération et la ville se consacrent désormais à de nouveaux chantiers destinés à couvrir une gamme étendue de besoins : en s'appuyant par exemple sur le parc applicatif pour alimenter en données des applications tierces ou en mettant l'accent sur les données de référence que les quelques 5000 agents territoriaux devraient pouvoir réutiliser. L'open data n'a pas donné les résultats attendus notamment sur le volet de la création d'applications par la voie privée. Le futur portail open data – en cohérence avec une nouvelle politique de la donnée – restera ouvert à tous. Chacun pourra venir y chercher des données, les exploiter et éventuellement en déposer en retour. D'autres accès seront créés pour les partenaires – étudiants, associations, industriels – dans le cadre de projets spécifiques.

Si les grandes villes comme Nice, Lyon, Rennes, Nantes, Paris, Strasbourg s'inscrivent dans ce mouvement avec les nuances déjà évoquées plus haut, les choix retenus par les villes intermédiaires et les petites villes restent inégaux.

En cette période de difficultés budgétaires, la diffusion de jeux de données en mode hébergé semble faire plus d'adeptes. Il existe de nombreuses formules d'hébergement au niveau national avec Etalab qui développe une politique d'accueil proactive et au niveau local : des offres régionales (Open PACA), départementales (Datalocale en Gironde, Seine-Maritime et Eure) ou encore portées par des syndicats de mutualisation (Manche numérique).

²¹ Cf. : <http://dataviz.rennesmetropole.fr/budget/budget-rennes-metropole-2016/> - 25 fév. 2016

²² Cf. <http://dataviz.rennesmetropole.fr/quisommesnous/index-fr.php>

Licences open data : à considérer aussi comme levier de négociation avec les GAFA



La politique d'ouverture des données repose sur la gratuité, ce qui n'est pas incompatible avec le système des licences utilisées. Quelques villes ont amorcé une réflexion sur le sujet au sein de l'association "OpenData France" afin de définir une position plus ouverte notamment en direction des géants de l'Internet. En mettant ses données sous licence ODBL Paris dit avoir suscité une forme d'inquiétude de la part des GAFA (Google, Apple, Facebook, Amazon). Cette licence met la donnée publique à disposition de tous, tout en garantissant qu'elle ne pourra pas être accaparée par un acteur spécifique. Des mécanismes engageant en effet les réutilisateurs ayant enrichi la donnée initiale à la republier sous les mêmes conditions de licence. Ce cercle vertueux permet de créer des communs pour lesquels chacun apporte sa pierre mais qui risquent de contaminer de manière virale des données susceptibles d'être commercialisées. L'usage de ce type de licence a surtout pour vocation dans certains cas bien précis (Temps réel, etc...) de donner aux collectivités locales des marges de manœuvre de négociation, une monnaie d'échange, avec ces grands acteurs

➤ *De nouvelles exigences avec la réglementation*

L'engagement du législateur dans une politique volontariste d'ouverture des données publiques, modifie le cadre d'action des collectivités territoriales. La loi NOTRe a mis en place une obligation de transparence des données des collectivités territoriales de plus de 3500 habitants et de leurs établissements publics à fiscalité propre et de rendre accessibles en ligne les informations qu'ils détiennent lorsque ces informations se rapportent à leur territoire et sont disponibles sous format électronique.

Cette obligation, qui devrait toucher près de 3000 communes et plusieurs centaines d'EPCI, n'a pas encore d'effet contraignant. Mais la loi pour une République numérique précise les obligations qui seront à la charge des collectivités. Le gouvernement estime en effet nécessaire de passer de l'incitation à **l'obligation pour les administrations de mettre à disposition les données qu'elles détiennent**. C'est pourquoi il se propose d'introduire davantage de continuité entre les phases de communication, de publication et de réutilisation des données en généralisant la mise à disposition par les administrations des données qu'elles détiennent. Dynamique qui va bien dans le sens d'une publication par défaut des données administratives. Les collectivités territoriales vont devoir s'y préparer sans doute collectivement pour les plus petites et dans un cadre si possible utile et opérationnel, afin de favoriser une véritable réutilisation.

➤ *Les premières mobilisations supra communales*

Certains territoires ont commencé à prendre les devants. La Région PACA a été pionnière en créant le portail Open PACA destiné à accueillir l'ensemble des acteurs de la région. Aujourd'hui une cinquantaine de partenaires cohabitent sur la plateforme dont une vingtaine de communes et EPCI (Aix, Marseille, Avignon, Métropole Aix-Marseille Provence ainsi que des structures intercommunales de taille intermédiaire comme le pays Dignois dans les Alpes de Haute Provence.

Open PACA s'inscrit dans un projet de refonte original de son portail tout en étudiant avec le CRIGE-PACA - centre de ressources régional en géomatique – l'ouverture d'un guichet commun unique de dépôt de données à l'échelle régionale. Des plateformes de nouvelle génération s'appuieront sur des entrepôts de données ouverts, pour faciliter le regroupement des acteurs autour de référentiels communs et thématiques et passer d'un système de distribution de données froides, chronophage sur la publication et la maintenance, à une logique de flux donnant accès à des données « chaudes » incluant même les données de flux temps réel. Le CRIGE et Open PACA proposeront leurs propres interfaces de programmation ainsi que celle d'Etalab en cours d'expérimentation dans quelques collectivités.

Martigues : open data en mode partagé



L'ex-communauté d'agglomération du pays de Martigues qui a depuis rejoint la Métropole Aix Marseille Provence s'est engagée dans une démarche d'ouverture de ses données publiques en sollicitant la Région Provence-Alpes-Côte d'Azur pour intégrer le portail partenarial régional Open PACA.

Aujourd'hui la ville et le pays proposent respectivement 106 et 50 jeux de données sur le portail régional qui sont également répliquées sur leur site web, cette fois en marque blanche. La mairie organise régulièrement des cartoparties pour enrichir avec les bénévoles les fonds cartographiques ainsi que des rencontres de quartier pour sensibiliser le public (open lab) à d'éventuelles campagnes de production de données.

L'agglomération est candidate à l'expérimentation de la future plateforme régionale. Le changement de plateforme est très attendu compte tenu de la pesanteur et de la complexité d'utilisation du site Open PACA dans sa forme actuelle. Et dans la perspective de l'arrivée d'un guichet régional unique de dépôt des données, la collectivité a d'ores et déjà mis à l'étude la publication de nouveaux jeux de données, mais cette fois géographiques.

L'objectif est de proposer un espace mutualisé préfigurant la constitution d'une infrastructure de données publiques ouverte à tous les acteurs notamment aux petites collectivités (plus de 3500 habitants et moins) ne disposant pas des moyens

de s'organiser. La Région espère de la sorte, réduire la fragmentation de la production, de la diffusion et de la réutilisation des données publiques.

L'Etat en région. Open Data Lab²³ est une autre initiative, portée cette fois par la Préfecture de région Languedoc-Roussillon-Midi-Pyrénées.. Cette expérience en vraie grandeur, financée dans le cadre du PIA (320 000 € sur 18 mois), poursuit plusieurs objectifs :

- Mettre en place l'open data dans les services déconcentrés de l'Etat, structurer l'ouverture des données filière par filière (eau, pesticides, tourisme...) en étudiant les besoins de réutilisation et l'offre existante et potentielle ;
- Proposer un accompagnement aux collectivités pour mettre en place l'open data.

Sur ce dernier point la Préfecture a réalisé avec l'aide de quelques communes pilotes un kit collectivités comprenant une brochure de sensibilisation, une check list des prérequis pour lancer un projet et un guide de lancement de l'open data. L'idée est de convaincre les collectivités que l'open data c'est simple et ça ne coûte rien puisqu'elles peuvent installer gratuitement leurs jeux de données sur data.gouv.fr, le site national, tout en valorisant leur démarche sur leur propre site puisque l'utilisation de certains scripts permet de réafficher le catalogue de jeux de données en marque blanche avec un moteur de recherche, une cartographie simplifiée et une présentation du catalogue d'API. On notera pour finir qu'il s'agit d'une initiative interne à la Préfecture, les collectivités territoriales de la région n'y sont pas associées.

3. Gouvernance de la donnée : un changement de paradigme métropolitain

➤ *Nouvelles stratégies urbaines et logiques de plateformes*

Au-delà de l'open data, certaines métropoles et grandes agglomérations sont engagées dans la transformation de leur système d'information pour le recentrer sur la donnée. Leur taille, l'étendue des compétences et les changements apportés par la révolution numérique, les exposent en effet à de nouvelles formes de "déluges" d'information dont elles connaissent la valeur et qu'elles cherchent à maîtriser pour en exploiter les richesses.

Mais la trajectoire risque de s'avérer longue et complexe. Les promoteurs de cette nouvelle gouvernance de la donnée, vont devoir en effet composer avec l'intégration des systèmes d'information hétérogènes de leurs communes membres, accélérer la dématérialisation, s'approprier la gestion de l'Internet des objets et des systèmes de suivi temps réel, et plus globalement intégrer les projets de smart city, sans oublier les réseaux sociaux, désormais "clés" pour une meilleure connaissance des "pulsations" de la ville et de ses habitants...

²³ www.opendatalab.fr

La bonne exploitation de ces ensembles de flux, impose des réorientations profondes dans les architectures et les modes de gestion des systèmes d'information pour assurer le découplage et la circulation de l'information, permettre de rassembler les données de référence du territoire, souvent dispersées et assurer le tout dans un cadre sécurisé pour les données.

La majorité des Métropoles comme Paris, Lyon, Marseille, Aix en Provence, Nice, Bordeaux, Nantes, Rennes, Toulouse sont déjà engagées dans un processus de transformation, ce qui confirme son importance. Toutefois, les chantiers demeurent largement expérimentaux. Dans les villes et agglomérations de taille plus limitée, la gouvernance de la donnée reste encore à l'état de concept, d'hypothèse de travail avec quelques exceptions comme l'agglomération de Chartres Métropole. Engagée dans un programme de smart city elle ré-urbanise son système d'information pour l'adapter aux transformations déjà engagées. Ce mouvement devrait progressivement s'étendre à des cités ou à des unités territoriales de plus petite taille, à condition de trouver des solutions économiques et probablement fortement mutualisées pour mieux répartir les coûts.

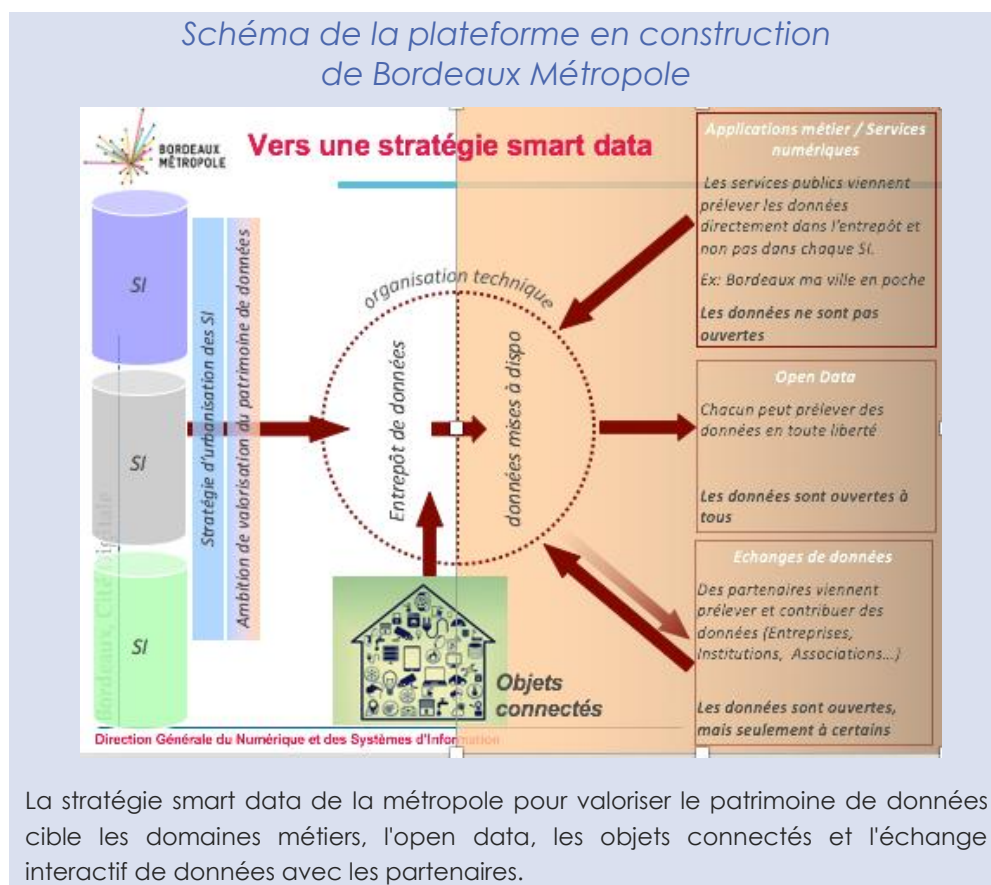
➤ *Les transformations du système d'information*

Les données sont aujourd'hui en majorité associées au logiciel qui les gère, ce qui rend leur extraction délicate. Lorsque la question se pose sur 50, 100, voire 150 applications métiers différentes, les problématiques de réutilisation, d'échange de données ou d'interopérabilité deviennent compliquées voire même impossibles à traiter pour des raisons techniques et budgétaires. Or ce cloisonnement est de plus en plus pénalisant dans les environnements fortement numérisés. L'application "Bordeaux ma ville en poche" qui offre à tout habitant la capacité d'acheter différentes prestations municipales - tickets piscine, accès à la bibliothèque, règlement du stationnement - et de signaler les dégradations constatées sur la voie publique est connectée à six applications métiers différentes ce qui dans une architecture traditionnelle a provoqué une forte dégradation des performances du système de requête. Pour y remédier un serveur relais faisant office de cache a été installé, pour permettre à l'utilisateur de retrouver un confort d'utilisation. A terme, dans une architecture de gouvernance des données, la solution consistera à verser les données actives des applications dans un entrepôt de données. L'application mobile ne communiquera plus directement avec les applications métiers mais uniquement avec l'entrepôt, via des webservices ou des Interfaces de programmation spécifiques. Le système sera plus compact, plus performant et capable d'effectuer lui-même automatiquement les mises à jour sur le compte client de l'utilisateur.

L'exemple bordelais n'est pas fortuit et les besoins de faire autrement se multiplient, ce qui a conduit la plupart des métropoles à s'engager dans des chantiers de transformation. Concrètement cela consiste d'abord à collecter des données existantes sur un territoire en provenance des différentes sources : des systèmes d'information métiers de la collectivité ; des objets connectés (caméras, horodateurs pour les parkings, smartphones et autres mobiles, capteurs environnementaux, systèmes de pilotage des feux tricolores aux intersections...) ; des partenaires, fournisseurs, sociétés titulaires d'une délégation de service public (ERDF, opérateurs de transports...).

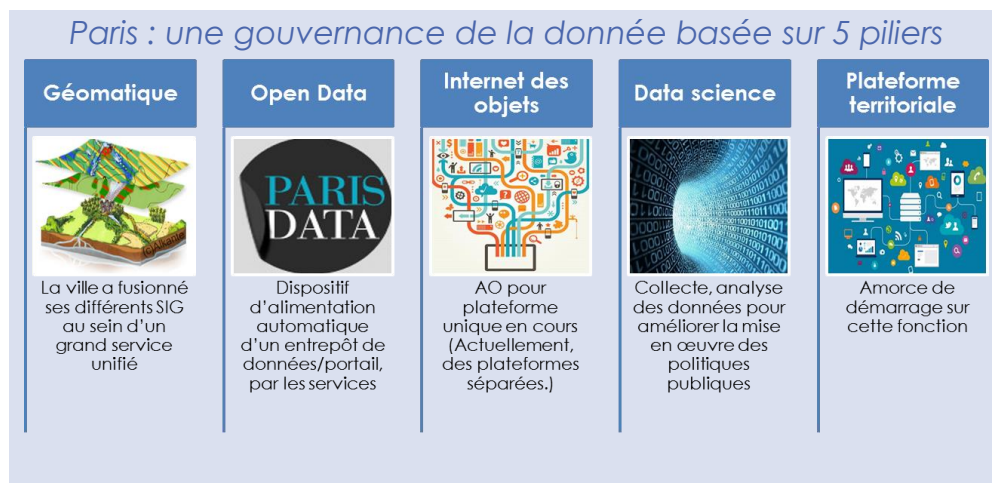
Puis cela se poursuit avec la construction brique par brique de plateformes intégrant divers composants : des entrepôts de données plus ou moins spécialisés (données alphanumériques, SIG, big data, Internet des objets, open data) associés à des logiciels ayant pour vocation de filtrer, convertir et extraire certaines données, sans oublier en bout de chaîne les outils de reporting (superviseur et hyperviseur) et d'analyse utilisés pour exploiter les données recueillies (Business intelligence, big data).

La plateforme a ainsi vocation à stocker toutes les données métiers et d'une manière générale les données utiles à de fins de réutilisation mais elle peut aussi se concevoir en réseau, en connexion avec d'autres systèmes. L'entrepôt est plus qu'un réservoir. Pour rendre les données disponibles à l'ensemble des agents de la collectivité et ouvrir certaines vers l'extérieur, celles-ci font l'objet de conversions dans des formats homogènes, elles sont indexées et accessibles via des services de consultation, de visualisation et le cas échéant d'enrichissement avec l'apport d'informations complémentaires.



Le déploiement de telles infrastructures est amorcé, bien qu'il n'existe pas encore à proprement parler de plateforme généraliste traitant des principaux métiers de la ville. La diversité des métiers et des sources d'information à traiter, fait de ce chantier une construction progressive et personnalisée. Chaque projet étant déterminé au démarrage par le contexte local et les priorités fixées par la collectivité.

- **Nice** s'appuie sur un entrepôt de données intégrant plusieurs composants dont une plateforme open data et une infrastructure dédiée à l'Internet des objets, et semble progresser sur les données environnementales (pollution et risques naturels) ;
- **Lyon** développe une approche généraliste à partir d'une plateforme ouverte et quelques spécialités axées sur les données de transports et de l'énergie débouchant sur des applications big data ;
- **Bordeaux** ne dispose pas encore de plateforme, mais sa feuille de route prévoit l'ouverture d'un nouveau portail open data en 2016 et d'un entrepôt de données généraliste en 2017 ;
- **Rennes**, collectivité pionnière de l'open data, développe ses projets sur les volets citoyens et décisionnels et axe sa démarche de gouvernance dans la perspective de création d'une régie locale de la donnée ;
- **Paris**, qui dispose déjà d'une plateforme de données ouvertes étendue (notamment sur le partage de toutes les données de mobilité), vient d'achever l'unification de ses systèmes d'information géographiques, et mise maintenant sur déploiement des réseaux intelligents et la création d'un entrepôt basé sur l'Internet des objets et le développement en interne des data sciences tout en se disant en retard sur la constitution d'une plateforme territoriale de données partagée ;
- **Chartres** met en place un centre de supervision urbain (CSU) de nouvelle génération, dans une architecture de gouvernance de la donnée.



➤ *Des solutions performantes, ouvertes et qui se diversifient*

Les applications de gouvernance des données sont de plus en plus variées et surtout elles couvrent le plus souvent des champs nouveaux, inédits, voire expérimentaux, mais dans des domaines qui restent pour l'instant assez concentrés sur quatre axes :

- **Les transports et l'énergie** sont des points de focalisation forts ayant déjà fait l'objet d'études collectives²⁴ (Voir encadré ci-dessous). Côté transports et mobilité, plusieurs projets d'optimisation des déplacements en cours (Optimod et Opticités à Lyon ; Gertrude à Bordeaux) reposent sur le développement d'applications d'analyse prédictive du trafic et ne pourraient atteindre le même niveau de performance sans plateforme d'agrégation des données. Côté énergie, les expérimentations de smart grids de quartier ou de gestion intelligente des bâtiments et d'analyse de leurs données de consommation énergétique, via des capteurs, sont notamment en cours dans des villes comme Nice, Rennes et Paris.
- **Les réseaux urbains** - eau, gaz, électricité, ramassage des ordures, qualité de l'air, circulation, stationnement - deviennent plus intelligents et se placent au premier rang des domaines impactés par le déploiement de l'Internet des objets. Nice expérimente un système d'hypervision multi-métiers permettant le pilotage de grandes fonctions environnementales, dans le cadre d'une expérience de monitoring urbain. De son côté Chartres met en place une centrale de surveillance associant vidéo surveillance, pilotage des carrefours et remontée d'incidents et d'informations sur le trafic.
- **Les applications mobiles de e-services** sont promises à des développements significatifs dans plusieurs directions : paiement et accès à des services locaux (piscine, stationnement, médiathèque) à Bordeaux ou encore services de vie quotidienne à travers la fourniture d'informations en temps réel sur une gamme étendue de domaines (horaires dynamiques et personnalisés à l'arrêt de bus, niveau de fréquentation des équipements, événementiels) etc... La faisabilité de ces applications qui mobilisent un nombre croissant de systèmes d'information métiers repose essentiellement sur le développement de plateformes internes regroupant toutes les données y compris les données de flux.
- **La dématérialisation administrative** avec l'application progressive aux collectivités locales du principe qu'une administration ne peut demander à l'utilisateur une information dont elle dispose déjà (programme "dites-le nous une fois"), va impacter les systèmes d'information locaux. Le lancement de FranceConnect et de quelques applications illustratives comme "marché public simplifié" et "demandes de subventions simplifiées" donne un signal fort sur les évolutions en cours. Autant de services qui commencent à être réutilisés aujourd'hui par les collectivités territoriales (Alpes-Maritimes, Hautes-Alpes...), et qui reposeront de plus en plus sur des logiques de gouvernance de la donnée.

²⁴ A signaler sur la mobilité, l'existence d'un groupe de travail national assez avancé sur les formats de données, animé par l'AFIMB (l'Agence française pour l'information multimodale et la billettique) et auquel participent l'Etat et les collectivités territoriales.

Valorisation des données : un groupe de travail métropolitain animé par la Caisse des Dépôts

La Caisse des Dépôts anime depuis trois ans un groupe de travail sur la valorisation des données réunissant huit métropoles françaises (Bordeaux, Lyon, Lille, Montpellier, Rennes, Nantes, Toulouse, Paris). Objectifs : produire une réflexion sur la gestion des données, échanger des bonnes pratiques, assurer une veille sur les chantiers nationaux et sur les modèles de gouvernance des villes à l'international. Créé à l'occasion du lancement d'une étude sur les données de transport (2013), le groupe poursuit ses travaux sur la base de réunions trimestrielles à l'occasion desquelles de grands acteurs sont invités (Ex : rencontre avec Google sur l'intégration des données de mobilité dans Google Maps et Google Transit). Les chantiers actuels portent sur les objets connectés et sur les données de l'énergie. Ce dernier axe de travail a donné lieu à la mise en œuvre d'une nouvelle étude initiée en début d'année par la Caisse des Dépôts et cinq métropoles et qui sera finalisée fin 2016.

➤ *Les datasciences au service de la ville et des territoires*

Le développement des data sciences au service de la smart city, de la ville et des territoires en général est aussi expérimental que prometteur. Les premières illustrations relevées confirment le potentiel innovateur de cette nouvelle discipline dans la **gestion des collectivités** (consommation énergétique des bâtiments), dans la **régulation** (pilotage de la gestion des feux tricolores à la sortie d'un grand événement populaire) ou encore en matière prédiction ("ce que seront les conditions de circulation dans une heure" afin d'anticiper sur les comportements). Ces fonctions font largement appel au big data et à l'algorithmique et commencent à être déclinées sur certains segments d'activité dans les villes. Les exemples relevés au cours des entretiens apportent quelques illustrations concrètes qui constituent des indicateurs de tendance.

Paris, Lyon, et Nice font figure de postes avancés. Le projet Datacity lancé par Paris il y a quelques mois avec le NUMA, proposait une dizaine de défis urbains, sur la mobilité, l'environnement, la gestion des déchets, l'énergie, les bâtiments et l'espace public afin de déterminer comment résoudre des problèmes complexes par l'usage de la data. Il conduit à la sélection de cinq projets accompagnés et financés par la ville : qui traitent aussi bien du repérage des logements insalubres dans l'habitat diffus, à partir de modèles d'identification, que de la détection des locations temporaires illégales en accompagnement du contrôle que la ville met en place. La ville de Lyon, comme indiqué précédemment, a beaucoup investi sur les déplacements et la prévision de trafic. Avec Optimod'Lyon, la métropole collecte, centralise, et traite l'ensemble des données de la mobilité urbaine sur une plateforme unique dans le but de créer des services innovants qui faciliteront les déplacements et la vie des usagers. Par ailleurs, elle a créé avec les grands opérateurs de services urbains, le Tuba, un laboratoire collaboratif. Ouvert aux porteurs de projets et aux entreprises, il sert de lieu d'expérimentation, d'échange, et de partage de connaissances, notamment sur l'ouverture des données et le big data. Il réunit des startups et des chercheurs du monde académique et peut mobiliser la plateforme métropolitaine pour des expérimentations. La ville est par

ailleurs engagée dans plusieurs programmes européens tels que BloTop²⁵ sur les objets connectés ou Smarter Together²⁶ sur la maîtrise de l'énergie. Pour élargir son champ d'action la métropole vient de lancer une nouvelle phase exploratoire en chargeant un stagiaire data scientist d'évaluer l'impact organisationnel du big data en tant qu'infrastructure. L'enjeu est l'extension du domaine d'expérimentation aux données non structurées et aux données temps réel. De son côté la Métropole Nice Côte d'Azur (MNCA) mène de front plusieurs programmes dont un sur la gestion des risques. Une des applications les plus immédiates est celle de la surveillance de la montée des eaux sur le réseau hydrographique local avec la mise en place de caméras numériques associées à un algorithme permettant d'anticiper les risques éventuels (à deux heures). Par ailleurs, les démonstrateurs de monitoring urbain (déchets, pollution, trafic), ou encore les projets réalisés dans le cadre du programme RéFLeXE²⁷ dédié à la sécurité du réseau électrique, via la gestion de l'effacement et de **Cityopt**²⁸ programme dédié à l'optimisation énergétique dans les foyers (suivi temps réel des données de consommation via le compteur Linky) constituent d'autres illustrations des chantiers en cours.

On pourrait encore citer bien d'autres manières d'utiliser les data sciences dans le domaine de l'analyse prédictive (risque incendie dans les immeubles, insécurité, vols) ou de l'analyse comportementale. Sur ce point, près d'une cinquantaine de collectivités et d'offices de tourisme ont déjà contractualisé avec la filiale d'Orange qui propose Flux Vision pour produire de l'analyse comportementale des touristes sur un territoire défini. D'autres – et parfois les mêmes – commencent à exploiter leur propres données réseaux (Wifi public et vidéo surveillance) pour effectuer ce type d'analyse, ce qui leur donne plus d'autonomie dans le choix des traitements à effectuer. Les villes de Paris, Strasbourg, Aix en Provence, Chartres, Bordeaux portent actuellement des projets de cette nature (Voir encadré ci-dessous).

25 <http://www.economie.grandlyon.com/projets-europeens/biotop-programme-pour-faciliter-la-creation-de-nouveaux-services-autour-des-objets-connectes-intelligents-35.html>

26 <http://www.lyon-confluence.fr/fr/actualites/horizon-2020.html>

27 <http://www.presse.ademe.fr/2015/06/projet-reflexe-valorisation-de-la-flexibilite-electrique-au-service-des-smart-grids.html>

28 <http://www.nicecotedazur.org/environnement/l-%C3%A9nergie-et-le-climat/cityopt>. L'objectif est de diminuer la demande en électricité lors des pics de consommations en planifiant différemment l'usage des appareils électriques au cours de la journée (c'est le principe de la « Demand-Reponse »).

Big data et aménagement urbain

Voici quelques illustrations complémentaires des domaines dans lesquels l'algorithmique appliquée à l'action publique peut rendre de réels services :

- Constitution d'îlots mixtes (bureaux et logements) susceptibles de mutualiser leur consommation énergétique afin de mieux équilibrer les variations en journée (Paris) ;
- Analyse des flux de circulation (véhicules et piétons) à partir de capteurs vidéo pour monitorer l'usage de l'espace public et réinventer les places en y installant du mobilier urbain déplaçable. L'analyse des flux permet de mieux comprendre les modes d'appropriation des lieux réaménagés dans le cadre d'un urbanisme tactique (Paris) ;
- Analyse de la fréquentation et des mouvements de foule à partir de capteurs de mouvement posés sur les vitrines des magasins (Chartres) ;
- Exploitation des données de Wifi public (réseau de bornes densifié) pour mieux comprendre l'utilisation de l'espace public par les habitants et les touristes à travers la réalisation de cartographies des flux (Strasbourg).

L'autre versant décisionnel s'appuie sur des technologies plus éprouvées de "Business intelligence" mais relativement peu répandues dans les collectivités territoriales en dehors du champ de la gestion interne. Sur le volet applicatif elles sont en voie de renouvellement en offrant une ergonomie plus intuitive. Quelques collectivités les utilisent notamment pour la réalisation de tableaux de bord de pilotage de l'activité de leurs propres directions ou pour diffuser de l'information à des tiers. Rennes Métropole met par exemple en place un tableau de bord à destination des citoyens présentant des chiffres sur six domaines de l'action publique (mobilité et transports, pilotage de l'action publique, Environnement et énergie, développement économique, aménagement et habitat, culture). Sur ce domaine de l'information le Casa (Centre for advanced spatial analysis de l'Université de Londres) publie un tableau de bord dynamique à l'adresse de chaque grande agglomération britannique. Il est composé d'indicateurs environnementaux et d'indicateurs se rapportant à l'activité urbaine (trafic, incidents éventuels sur les lignes de métro, nombre de bus en service et de rames de métro en circulation...) et pourrait facilement être transposé et enrichi par les villes elles-mêmes.


 SOURCE : CITY DASHBOARD - LONDON²⁹

➤ *Relations fournisseurs et délégataires des collectivités : un début de clarification sur la propriété et l'usage des données*

Dans ce contexte de transformation et même de changement de paradigme, les questions relatives à la propriété des données dans l'environnement de travail des collectivités focalisent l'attention. Surtout envers les nombreux professionnels et concessionnaires de services publics en charge de la gestion des services urbains. Il serait en effet vain de mettre en œuvre des politiques ambitieuses de gouvernance de la donnée et de n'accéder qu'à une partie des actifs informationnels de la collectivité.

C'est pourtant encore fréquemment le cas pour diverses raisons qui tiennent la plupart du temps à un manque d'appétence ou de volonté des collectivités et au refus de certains concessionnaires :

- Le formalisme juridique à respecter dans les contrats sur la propriété des données n'est pas toujours assuré par les collectivités, que ce soit dans les contrats de délégation de services publics ou dans les contrats de plus courte durée avec les autres fournisseurs.
- Les technologies propriétaires utilisées par certains prestataires font obstacle à un accès direct ou à la récupération des données en fin de contrat.
- La maîtrise des données suppose de la part des administrations locales, des compétences techniques et engendre souvent des dépenses qu'elles ne sont pas toutes prêtes à assurer, y compris aujourd'hui.
- Certains délégataires ne satisfont pas aux demandes de transmission de l'autorité délégante, invoquant leurs obligations en matière de protection des données personnelles ou le secret des affaires.

Mais la posture des collectivités évolue. La plupart des grandes villes renforcent aujourd'hui leur arsenal juridique, ce qui n'exclut pas les surprises. Ainsi le formalisme à respecter dans les contrats est traité avec plus d'attention. Certaines villes, comme

²⁹ <http://citydashboard.org/london/>

Bordeaux, ont élaboré des modèles de textes et de procédures à des fins de diffusion systématique dans les services. Des blocs de texte sont à insérer dans les cahiers des charges de tous les prestataires publics ou privés, de manière à établir un droit de propriété et de distribution indiscutable sur la donnée. Pour les contrats importants (Type concessions de service public) les dispositions sont négociées en amont lors des phases de discussions préalables. Ce positionnement systématique commence à porter ses fruits.

Ce travail est désormais relayé par l'association OpenData France qui réunit les bonnes pratiques et les traduit en clauses types à insérer dans les marchés publics. Celles-ci prévoient même la prise en compte de l'open data.

Mais la mise en œuvre d'un droit d'usage sur les données pendant la durée d'exploitation du contrat n'est pas automatique et doit être formulée avec précision.

La ville de Bordeaux a obtenu à l'occasion du renouvellement de sa DSP transport, que l'entrepôt de données du délégataire ainsi que ses composants soient installés dans les locaux de la collectivité (symboliquement, pour affirmer son droit de propriété, mais aussi pour maîtriser plus facilement les équipements nécessaires). De quoi obtenir une totale transparence avec le système d'information du délégataire et une réversibilité des données sans douleur pour l'autorité délégante, en cas de changement de délégataire.

Seul inconvénient, le fait que la donnée soit totalement réutilisable et disponible à l'intérieur du système d'information auquel elle est rattachée, génère des coûts de réplication, de documentation et de développement de web service ou d'API pour la consommer. En outre la négociation suppose de la part de la collectivité une capacité de modélisation des informations et des précisions sur les nombreux indicateurs qu'elle souhaite recevoir de manière automatique de la part de ses fournisseurs, ce qui suppose qu'elle soit elle-même entrée dans une logique de pilotage par la donnée. Ce travail de formalisation, encore relativement inédit, représente une charge financière et une mobilisation des services de la collectivité qui explique sans doute la rareté des initiatives de ce genre. Beaucoup de collectivités estiment plus facile et plus immédiat en termes de résultats de négocier des services à la charge du titulaire du contrat plutôt que le rapatriement des données brutes. La maîtrise des données représente un prix à payer, même si le délégataire prend à sa charge l'installation d'une plateforme ou la transmission de données il y aura un coût supplémentaire pour la collectivité.

Le changement visible de posture des villes et agglomérations et la montée en puissance des nouvelles formes de gouvernance de la donnée devraient toutefois faire évoluer cette vision encore trop peu partagée.

➤ *Administrateur général des données et réorganisation interne*

Les nouvelles politiques de la donnée impactent l'organisation interne des grandes métropoles et celle de leurs systèmes d'information.

Les transformations en cours sont incarnées par la création de la nouvelle fonction d'administrateur général des données. Déjà effective à Lyon (mi 2015) et à Paris (début 2016), cette fonction est fortement marquée par sa dimension transversale. A

Lyon le titulaire anime une communauté autour de la donnée dans une approche territoriale et non exclusivement centrée sur la collectivité. A Paris, la fonction est à la fois axée sur la pédagogie - sensibiliser et accroître le niveau d'acculturation des services -, sur la coordination des actions et la mutualisation des moyens. Dans les deux cas il y a volonté de décroisonner l'organisation et de la rendre plus communicante et interactive. Cette volonté se traduit également dans la constitution de véritables équipes, du moins pour les collectivités les plus avancées. Le Grand Lyon s'appuie déjà sur une équipe chargée de la diffusion des données et du déploiement des géoservices, composée d'un noyau de cinq personnes en back office sur la plateforme de données et chargées de la faire évoluer en fonction des demandes formulées par les usagers. La métropole s'appuie par ailleurs sur des experts en interne (bases de données, architectes des Systèmes d'information et urbanistes), sur des prestataires extérieurs et sur un réseau de référents métiers au sein des directions et sous-directions concernées (eau, urbanisme, mobilité...). Tous ces acteurs se réunissent mensuellement pour suivre l'évolution du patrimoine de données et coordonner les initiatives notamment afin de déterminer les conditions d'ouverture des données (tous publics ou publics restreints). L'ensemble représente au total une vingtaine d'agents en interne. De son côté, Paris s'appuie sur un réseau d'innovateurs internes notamment pour promouvoir les data sciences. Afin d'en démontrer le potentiel, l'administrateur des données anime un groupe de travail en interne réunissant cinq ou six fois dans l'année directeurs et élus, pour assurer la coordination des équipes et des projets.

Le positionnement hiérarchique de l'administrateur des données est généralement assez élevé dans l'organigramme de la collectivité. Dans la vision (non définitive) des responsables d'une des métropoles interrogées, l'administrateur (bientôt nommé), sera chargé de mobiliser l'administration sur l'application des règles et des bonnes pratiques autour de la donnée. Il sera suffisamment proche des élus pour inscrire la diffusion de la donnée dans l'agenda politique et dans une position hiérarchique lui donnant la capacité de mobiliser les acteurs du numérique en interne. Dans une position charnière entre les élus et les acteurs du numérique il se posera en conseiller et facilitateur de la stratégie numérique décidée et arbitrée par les élus et comme le coordinateur des équipes de production et du service des études dans la perspective de la création d'un "service de valorisation des données". Ce service aurait pour mission de poursuivre le travail de sensibilisation en interne.

4. Infrastructures de données publiques territoriales

➤ Mise en réseau et plateformes de mutualisation

A côté des grandes agglomérations, le bloc communal (Communes + EPCI) ne dispose pour l'instant, ni des mêmes capacités d'action ni des moyens pour les financer. Alors qu'il est confronté à des défis numériques à la fois dans son organisation interne et sur le développement territorial : accroître sa présence sur le web à l'échelle du territoire afin d'améliorer l'accessibilité aux services locaux, simplifier les démarches administratives au bénéfice des citoyens tout en les rendant plus efficaces ou croiser et réutiliser facilement certaines données produites sur le territoire (afin de se comparer avec d'autres ou pour optimiser sa gestion), ...

D'autres besoins de regroupements et d'agrégation de données existent et devraient même s'étendre au fil du déploiement des réseaux très haut débit. Ils se matérialisent déjà sur des activités à forte valeur ajoutée économique, comme le tourisme ou l'énergie :

- **Le tourisme** devient un secteur consommateur de données. Dans les stations de sports d'hiver et de loisirs d'été en montagne, le long des zones côtières ou encore dans les territoires chargés d'histoire, les données touristiques - patrimoine, événementiel, hébergements, restauration, modes de déplacements et de consommation des touristes - deviennent essentielles. L'existence d'APIDAE, une des premières initiatives du secteur illustre cette tendance. Ce service collaboratif porté initialement par les offices de tourisme et le comité régional du tourisme de la région Rhône Alpes est une base d'information en ligne alimentée par ses membres et qui a pour enjeu leur réutilisation massive. Aujourd'hui cet entrepôt de données rassemble une communauté de 900 structures dont 450 contributeurs (offices de tourisme), 75 diffuseurs publics et privés et une quarantaine d'agences web et intégrateurs. Ce service rayonne sur les régions Rhône-Alpes Auvergne, Provence Alpes Côte d'Azur (bientôt sur l'Île de France) et alimente déjà plus de 1000 projets réutilisateurs des données sur le Web. La Direction générale des entreprises, étudie les moyens d'industrialiser le modèle (Voir C. [Secteur public d'Etat : de l'open data à la gouvernance par la donnée](#) ~~Secteur public d'Etat : de l'open data à la gouvernance par la donnée~~)
- **Le secteur de l'énergie**, à travers le développement des énergies renouvelables et des politiques de performance énergétique, va devenir un consommateur de données massives y compris en milieu rural. Il touche déjà le déploiement de l'électromobilité et des bornes de recharge électriques (IRVE), les premières expériences de microgrids villageois, ou encore les opérations de rénovation thermiques du parc de logements des bailleurs sociaux et des bâtiments publics, et fait monter la courbe des besoins en systèmes d'hypervision et en analyse des données.

Quelques initiatives de coopération ont vu le jour au cours des derniers mois. Dans la Drôme la plateforme Ozwillio a ouvert la voie aux plateformes de services centrées

Mis en forme : Police :Italique

sur la donnée³⁰, quelques syndicats informatiques ont engagé des chantiers de gouvernance de la donnée pour leurs membres (SICTIAM, Manche Numérique) et les syndicats d'énergie développent des politiques ambitieuses de développement des énergies renouvelables (Loire, Calvados, Haute-Savoie, Nièvre, Vienne, Vendée, Dordogne, Manche, Morbihan, etc...) qui les porteront à mettre aussi en place des réseaux intelligents associés à de nouveaux systèmes de gestion des données.

Mais pour éviter une grande fracture les chantiers à venir sont conséquents et induisent la mise en réseau des collectivités territoriales.

Les structures de mutualisation informatiques (SMI) des opérateurs techniques de la donnée...

On dénombre aujourd'hui une soixantaine de SMI réparties sur le territoire et sous des formes juridiques variables : syndicats mixtes de mutualisation³¹ - les plus nombreux -, centres de gestion³², associations³³, agences techniques départementales³⁴, DSI de conseils départementaux³⁵... etc.

Ces structures proposent des solutions métiers, des outils collaboratifs, des applications de gestion de la relation citoyen et soutiennent leurs adhérents dans la mise en œuvre des grands chantiers de dématérialisation portés par l'Etat.

Une dizaine d'entre-elles proposent déjà leurs solutions sur des plateformes de services informatiques – SICTIAM (Alpes-Maritimes), Alpi (Landes), e-Mégalis (Bretagne), e-Bourgogne, conseil départemental de l'Aube, Agence territoriale d'ingénierie publique (Bas Rhin), Soluris (Charente Maritime), e-Collectivités Vendée, Syndicat mixte des systèmes d'information (Seine St-Denis) etc.

Cette organisation en plateforme, va leur permettre d'offrir à leurs collectivités adhérentes de nouvelles prestations autour de la donnée : intégration de FranceConnect, exploitation des API nationales proposées par l'Etat, gestion des procédures d'échanges dématérialisés entre applications, entrepôts de données mutualisés....

Dans ce contexte et compte-tenu de leur présence répartie sur le territoire, ces structures pourraient jouer un rôle plus déterminant dans la mise en place d'infrastructures de données publiques et sur l'accompagnement des collectivités pour leur mise en conformité avec les obligations actuelles et futures en matière d'ouverture des données.

³⁰ Depuis quelques mois, Ozwillio a fait le choix de devenir un acteur économique et va bientôt se transformer en société coopérative d'intérêt collectif.

³¹ Syndicats intercommunaux : SICTIAM, Gironde Numérique, Manche Numérique, Alpi40...

³² Centre départementaux de Gestion : Côtes d'Armor, Isère, Haute-Loire, Nord...

³³ Associations : ADM54, Pôle Numérique 26,

³⁴ Agences techniques départementales : Allier et Dordogne

³⁵ Conseils départementaux : Aube, Hautes-Alpes

E. Freins et facteurs clés de succès identifiés

1. Des enjeux masqués par l'open data

Aujourd'hui, l'open data reste le principal levier du changement mais tend à masquer les autres enjeux en particulier celui de gouvernance de la donnée. La position de l'open data s'explique en partie par les effets de visibilité politique et la « bonne image » véhiculée par le concept (transparence vis-à-vis des citoyens). À l'inverse, la gouvernance de la donnée bien que stratégique, reste technique, portée en interne par les collectivités et peu visible pour les électeurs.

Néanmoins les collectivités progressent en maturité et déplacent progressivement le curseur sur cette gouvernance de la donnée.

L'open data n'est pas à jeter aux orties, bien au contraire les politiques mises en œuvre restent toujours des catalyseurs de prise conscience de la valeur des données, et un vecteur de sensibilisation des élus. Mais elles ne doivent pas occulter tout le reste, les données à récupérer auprès des délégataires pour la bonne gestion du service, les données à partager entre services, les données en provenance des réseaux de capteurs...

Le principe de l'ouverture "par défaut" des données publiques qui sera appliqué prochainement devrait accélérer cette transformation. Mais l'open data par défaut n'est pas l'ouverture de toutes les données et ne couvre pas tous les champs d'action de la collectivité.

Aussi il convient de mettre en avant la gouvernance des données plutôt que l'open data qui reste certes un élément clé mais en tant que sous-ensemble de la gouvernance.

2. Moyens et engagements budgétaires limités

Les moyens humains et financiers dédiés à la gouvernance de la donnée au sein des collectivités restent limités au regard des enjeux, y compris au sein des collectivités les plus avancées. À titre de comparaison, la Métropole de Nice et le Grand Lyon consacrent moins de cinq ETP à cette gouvernance alors qu'il n'est pas rare de voir les grandes villes américaines mobiliser une vingtaine d'agents.

Mais cette position évolue rapidement. Nice s'appuie sur des personnels répartis dans les services tandis que le Grand Lyon s'est doté d'un poste d'administrateur général des données chargé d'animer un réseau de « référents donnée » par service.

Les petites collectivités, même celles qui ont conscience des enjeux, sont souvent démunies pour faire face aux compétences et aux budgets requis. C'est pourquoi l'on observe des tentatives d'organisation autour de plateformes agrégatives régionales (ex : Open PACA) ou départementales (structures de mutualisation informatique). Quelques centres de ressources géomatiques comme le GIP Aquitain ATGeRi ont fait évoluer leur activité vers la gestion de la donnée et mobilisent pour cela des équipes de taille croissante. Dans le cas précité, une vingtaine de personnes assurent la gestion de près de 1000 conventions de collecte de données.

3. Le taux de mutualisation devient déterminant

Le niveau d'avancement de la mutualisation des fonctions support est un déterminant de la gouvernance des données à l'échelle d'un EPCI. La mutualisation de la DSI entre la métropole et la ville-centre est un réel facilitateur pour amorcer des regroupements de gestion et optimiser les centres de coûts. Certains territoires ont depuis élargi leur périmètre de mutualisation à d'autres communes, à l'enseigne de Bordeaux Métropole qui a intégré les SI de huit communes supplémentaires et dispose désormais d'une DSI de 235 agents. Dans les territoires à dominante péri-urbaine et rurale, la mutualisation constituera sans doute un élément incontournable. La présence d'une soixantaine de structures de mutualisation réparties sur le territoire devrait constituer un levier de développement efficace dans le cadre de projets de plateformes de service.

4. Un retour d'expériences limité

Les initiatives de big data restent encore circonscrites à quelques grandes métropoles, comme Paris, Lyon ou Nice. Et les retours d'expérience ne permettent pas de tirer des conclusions sur les bonnes approches, les pratiques et les technologies, etc. Ce qui peut conduire les collectivités à différer un passage à l'acte ou simplement à ne pas voir un potentiel d'usage. La promotion par l'exemple et par les expérimentations en vraie grandeur vont devenir capitale pour assurer le développement des data sciences dans les territoires.

5. Faible acculturation

En dehors de quelques grandes agglomérations et des structures de mutualisation informatiques dotées de plateformes de services, les collectivités restent le plus souvent en retard sur les sujets liés à la donnée. Cela est en grande partie dû à une absence de prise de conscience des enjeux. Le travail de sensibilisation des élus et des agents doit être mené, condition sine qua non pour faire émerger, à terme, des projets et surtout des actions collectives.

6. Situation de gouvernance implicite

Comme au niveau national, dans la grande majorité des collectivités territoriales, il n'y a pas encore de stratégie explicite de gestion des données publiques, mais plutôt une réflexion pour chaque projet. Ce sont les règles définissant les relations entre services ou les choix d'architecture pour le système d'information qui esquissent, de fait, une gouvernance par défaut.

Cette absence de formalisation peut s'expliquer au regard de la variété des données à prendre en considération et de l'appétence mesurée des élus à traiter le sujet, notamment en lui donnant une traduction budgétaire significative.

Seules quelques métropoles et agglomérations sont mobilisées. . Quelques autres s'approprient à le faire en anticipant sur la prochaine obligation d'ouverture des données publiques.

SICTIAM : un opérateur de mutualisation pour les collectivités

Créé en 1989, ce syndicat mixte ouvert compte plus de 250 collectivités adhérentes réparties sur sept départements. Il offre une gamme complète de solutions applicatives ainsi que des prestations d'accompagnement : solutions informatiques dédiées à la gestion communale ; Plateformes de dématérialisation ; Centrales d'achat ; Aménagement numérique du territoire et des usages. Le SICTIAM est engagé dans une démarche de gouvernance par la donnée qui pourrait déboucher sur de nouvelles offres de

7. Redevance : facturation des services

La loi Valter³⁶ consacre le principe de gratuité de l'utilisation et de la réutilisation des informations publiques, ce qui est un principe sain si l'on considère l'équation économique de la réutilisation.

Reste à savoir comment la réglementation établira la frontière entre données de base et données enrichies (donc service donnant lieu à rémunération).

Paris a par exemple pour projet d'ouvrir en open data le plan 3D de la ville, mais certains professionnels attendent de cette base un niveau de granularité beaucoup plus fin et donc plus coûteux aussi à obtenir.

Un système "freemium" permettrait à la ville de proposer un service de base réutilisable gratuitement et un service "premium", payant aux entreprises qui le souhaitent.

La question est de savoir si la formule premium est bien un service donnant lieu à rémunération supplémentaire dans le cadre des dérogations ou hors dérogation.

Dans ce cas, la collectivité devra-t-elle attendre la publication d'un décret fixant le montant de la redevance ou sera-t-elle libre de fixer librement ses prix.

Les collectivités qui souhaitent développer des services à valeur ajoutée à partir de leurs données vont désormais être plus souvent confrontées à ces questions de modèle économique et de redevance.

LA PROBLEMATIQUE DE LA REDEVANCE

La loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public (dite « loi Valter ») résulte de la transposition de la directive 2013/37/UE du Parlement européen et du Conseil adoptée le 26 juin 2013.

Cette loi consacre en droit français et après plusieurs années de débats, le principe de gratuité de l'utilisation et de la réutilisation des informations publiques tout en prévoyant deux dérogations, dont une à vocation générale permettant aux collectivités d'établir une redevance de réutilisation « lorsqu'elles sont tenues de couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service public. ». Cette précision signifie que les redevances ne pourront donc plus prendre en compte, comme cela était le cas avant l'entrée en vigueur de la loi VALTER, la rémunération raisonnable des investissements.

L'article 5 désormais inséré au sein du nouvel article L. 324-1 précité du code des relations entre le public et l'administration prévoit que « Le produit total du montant de cette redevance [...] ne dépasse pas le montant total des coûts liés à la collecte, à la production, à la mise à la disposition du public ou à la diffusion de leurs informations publiques. »

³⁶ LOI n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public
<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031701525&dateTexte=20160914>

Il est regrettable, que, pour des raisons liées au retard du gouvernement français dans la transposition de la Directive susvisée, le sujet des conditions financières de réutilisation des informations publiques n'ait pas été traité dans le cadre de la loi pour une République Numérique.

Comme cela a été relevé par les rapporteurs du projet de loi (Sénat et Assemblée Nationale), alors que le rapport Trojette incitait l'Etat à engager une réflexion sur les « modèles économiques » qui pourraient être envisagés pour les administrations, ni la loi VALTER, ni la loi pour une République numérique n'en font mention.

Ainsi, en l'absence de précisions du législateur, il appartiendra donc aux collectivités territoriales de rechercher des modèles économiques stimulant l'innovation.

Quid de la question des modèles économiques et de la gratuité des données au niveau local ?

Synthèse de l'interview d'Henri Verdier, Administrateur général des données, publiée par Localtis (31 mai 2016).

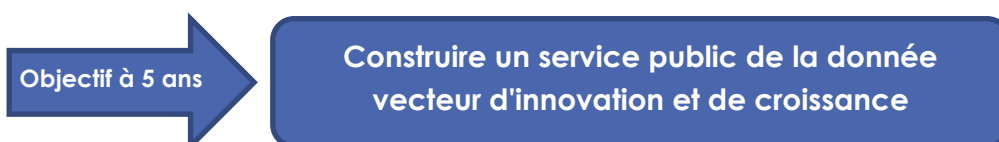
Selon Henri Verdier, si la loi pour une République numérique a notamment pour ambition « d'éviter que les administrations ne se vendent des données entre elles, notamment au niveau de l'Etat », l'impact de cette loi sur les collectivités est encore incertain, d'autant plus que « très peu de collectivités vendent aujourd'hui des données, alors que quasiment toutes en utilisent, de plus en plus quotidiennement ». Pour l'administrateur général des données, au-delà des questions de redevance, l'enjeu est celui de la création de valeur rendue possible par la réutilisation des données : « la gratuité des données n'est pas un jeu à somme nulle, où on ne ferait que compenser des transferts d'argent. Les données jouent aussi un rôle de référentiel, créant de la valeur grâce à l'innovation mais aussi à la rationalisation des stratégies des acteurs. Nous allons collectivement avoir besoin de développer notre "intelligence de la valeur de la donnée", afin de déterminer de mieux en mieux lesquelles doivent être ouvertes à tous, uniquement aux entreprises ou aux chercheurs, lesquelles doivent être exploitées en propre ou mises à disposition des autres administrations. La question d'un aller-retour permanent entre le meilleur usage de la donnée et sa meilleure diffusion est bien posée ici. »

III. Vers une nouvelle gouvernance de la donnée

Poursuivre le développement des chantiers de gouvernance de la donnée dans les villes et développer une infrastructure numérique de données performante, accessible aux entités publiques les moins dotées en ressources requiert du temps, des moyens humains et financiers et va mobiliser toutes les forces en présence.

Aussi pour prendre la mesure du programme d'actions et du chantier à réaliser, une projection de l'édifice à construire à cinq ans a été effectuée afin de fixer un objectif et un cap à suivre. Cet exercice théorique permet de définir une première trajectoire et des objectifs d'organisation.

A. Recouvrer plus de souveraineté sur la donnée pour dynamiser l'économie



Fixer une échéance de cinq années pour construire "un service public de la donnée à l'échelle du pays" est un pari ambitieux, réalisable et nécessaire pour aligner la France sur les transformations de la société numérique.

Le projet repose sur la création d'un service public de la donnée national, décliné localement, capable d'améliorer significativement la gestion des administrations, la qualité des services et la relation citoyenne et de fournir aux entreprises des données susceptibles de produire de la valeur ajoutée.

Cette ambition repose sur l'orchestration d'initiatives visant à assurer l'industrialisation des dispositifs, la maîtrise des données et un regain d'agilité dans les administrations. Déjà amorcées au niveau national et engagées localement elles portent principalement sur quelques actions structurantes :

- **Etendre la culture de réutilisation dans les services publics et les entreprises** sur l'ensemble du territoire, à travers la généralisation de la création et de la diffusion de bases de données de référence et d'intérêt général.
- **Généraliser des API et des web services** pour assurer l'interopérabilité des systèmes d'information, faciliter les échanges de données et la circulation des informations en mode dématérialisé. L'ouverture de FranceConnect, de chantiers de simplification comme "Dites-le nous une fois", de l'archivage électronique devraient provoquer un effet catalyseur et être relayés localement par des plateformes de dématérialisation (à l'exemple de celles qui existent déjà pour certaines procédures administratives).
- **Automatiser des processus métiers**, afin d'accroître, la performance administrative et, côté entreprises, faciliter les regroupements sur des

plateformes de commercialisation (afin de mieux maîtriser la désintermédiation au niveau local)

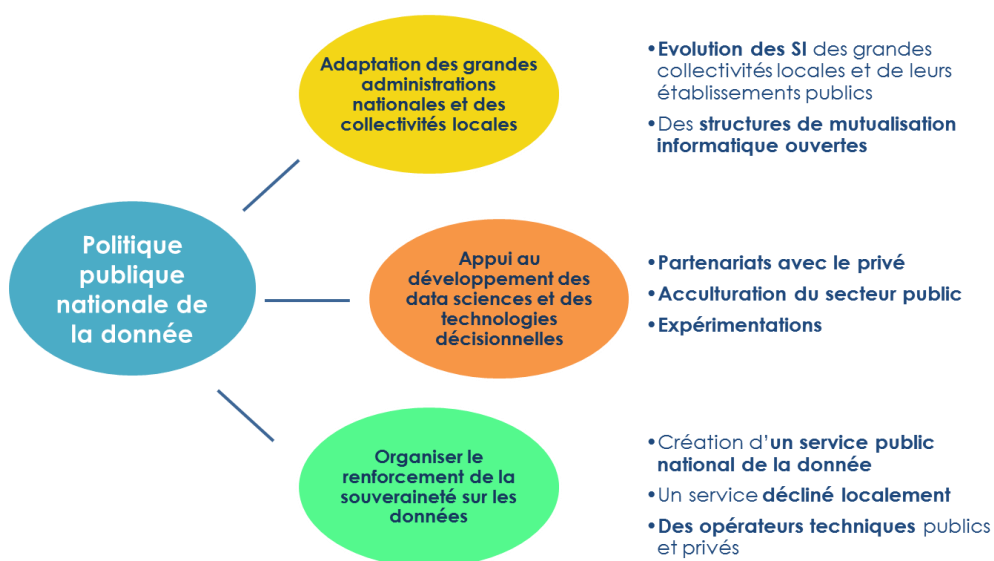
- **Etendre l'open data et le shared data** pour amplifier la réutilisation et la création de nouveaux services, tableaux de bords statistiques et autres outils d'aide à la décision...
- **Intégrer le big bang des data sciences.** Ce secteur naissant, dans le champ de l'action publique, devrait jouer un rôle primordial dans les services urbains, transports, la gestion de l'énergie, la santé, l'éducation, l'administration avec des retombées en termes de création de richesse qui apparaissent déjà prometteuses
- **Organiser une gouvernance partagée des données** entre l'Etat et les collectivités territoriales. Ce point est essentiel même si l'Etat retient plutôt le terme de "mission de service public" plutôt que de service public.

B. Politique à mettre en œuvre pour atteindre les objectifs à 5 ans

1. Une politique publique nationale de la donnée équilibrée

La mise en œuvre d'une politique nationale de la donnée, efficiente et partagée, repose sur la mobilisation des échelons nationaux et locaux pour opérer et orchestrer le renforcement de la souveraineté les données et en exploiter la valeur ajoutée. Elle se fonderait sur trois principes :

- Faire basculer les systèmes d'informations des administrations et des entreprises dans la **gouvernance des données**.
- Associer à ce chantier fondamental des actions d'appui et d'accompagnement au développement des **data sciences** (business intelligence, big data, data visualisation, monitoring...).
- Installer **un service public national de la donnée, décliné localement** (infrastructures, formation, opérateurs techniques), pour animer orchestrer le déploiement de cette nouvelle gouvernance, et particulièrement à l'adresse des territoires ou des administrations ne disposant que de moyens limités.



L'élaboration d'un programme national d'action pour une nouvelle gouvernance de la donnée pourrait comprendre des dispositions communes (et coordonnées) au local et au national, reprenant les grandes lignes de l'action amorcée par l'Etat et enrichies par les collectivités territoriales et les entreprises.

Ces dernières seraient associées à l'élaboration du programme de **construction de l'infrastructure de données publiques à l'échelle du pays** et mobilisées sur sa mise en œuvre : accompagnement des territoires ou des entreprises, des filières, expérimentations de projets innovants, campagnes d'information et de sensibilisation, développement de modèles industrialisés...

Dans les développements qui suivent on distinguera les acteurs disposant d'une autonomie leur permettant une prise en charge et une mise en œuvre directe des projets et ceux qui seront conduits à faire appel à des solutions mutualisées (villes moyennes, petites communes...).

2. Adaptation des grandes administrations nationales, des collectivités locales et de leurs établissements publics

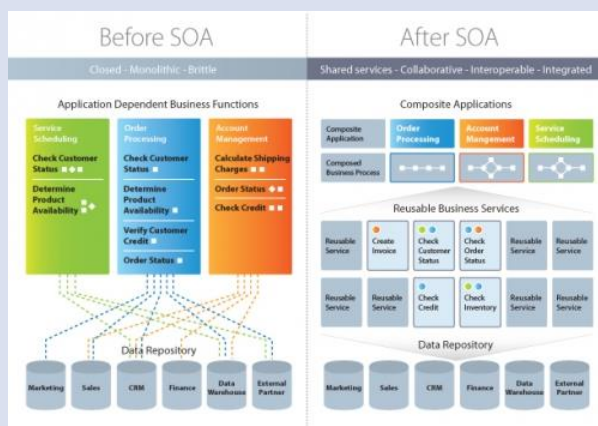
Les métropoles et agglomérations, qui disposent d'une autonomie informatique, vont poursuivre la transformation de leur système d'information en vue de l'établissement d'une gouvernance de la donnée. Le chantier passera par l'intégration graduelle des systèmes d'information des autres collectivités et par la conquête ou la reconquête du droit de propriété sur leurs données avec la rédaction de nouveaux contrats.

➤ Evolution des SI des grandes collectivités locales et de leurs établissements publics

Le recentrage des systèmes d'information autour de la donnée, va faciliter la circulation de l'information.

Nous avons vu que ces démarches, encore en construction dans les collectivités, conduisaient à centraliser et consolider l'ensemble des informations en un ou quelques lieux (entrepôts) afin de faciliter les échanges inter-applicatifs, l'alimentation en données de nouvelles applications ou pour assurer des fonctions de reporting et d'analyse. L'urbanisation des données, facilite le travail dans des environnements hétérogènes complexes, facilite les échanges de données en temps réel ou le traitement en masse, automatise les modifications apportées aux référentiels, sur l'ensemble des bases concernées (création de nouveaux arrêts de bus, lampadaires d'une rue nouvelle, etc...) et renforce l'efficacité générale. En outre elle permet de réorganiser sans avoir à modifier l'ensemble du système d'information.

Urbanisation du SI pour une meilleure adéquation avec la stratégie de la collectivité



SOURCE : TRIDENS

Un système d'information urbanisé (à droite sur la figure) peut absorber toutes sortes d'évolutions dans l'organisation : qu'elles soient liées à des changements de stratégie ou encore à l'apparition de nouveaux besoins ou contraintes. Elle apporte une capacité à visualiser et à représenter les évolutions en cours tout en optimisant les ressources (applications, données, infrastructures, etc.) grâce au découplage et à la simplification. Le système d'information devient plus réactif, flexible et agile.

Ainsi, lorsqu'un service public, comme celui de l'eau est géré à l'issue d'un regroupement intercommunal par trois délégataires dotés de modèles de données différents, l'entrepôt et les outils qui lui sont associés vont permettre d'harmoniser la gestion, en proposant que chaque espace privé converge vers cet espace public. Chacun conservera ses protocoles et ses procédures, tout en offrant des données normalisées dans l'entrepôt.

Le renforcement de la mutualisation est une des conséquences des mouvements de réorganisation autour de la donnée. Associées aux technologies de virtualisation dans le cloud, les logiques de gouvernance des données y trouvent de nouveaux débouchés. La mutualisation s'imposait déjà avec les regroupements de collectivités, elle va s'amplifier sous le coup de la récupération des données métiers, des données géographiques, des données issues de l'Internet des objets ou de sources non structurées venant par exemple des réseaux sociaux.

Extraction des données de référence. Cet autre élément clé de l'urbanisation se rapporte à des objets génériques - mobilier urbain, arrêts de bus, écoles, etc. – associés à des processus métiers. Dans le modèle classique, les données sont dispersées dans de nombreuses applications : la base élèves s'appuiera sur sa propre base "écoles", la gestion des bureaux de vote, les transports et la restauration scolaires, les centres de loisirs font de même... Alors que dans un système d'information urbanisé, les données sont réunies dans un référentiel unique qui les qualifie, précise leur identité et leurs modalités d'usages. Chaque modification du référentiel donne lieu à des mises à jour automatiques de toutes les bases concernées. Ce principe permet également de corriger une grande quantité de données de mauvaise qualité : auparavant plusieurs versions d'un même référentiel souvent inégalement mises à jour, coexistaient. En limitant l'intervention humaine, les données de référence en base unique partagée, deviennent une source d'économies.

Exploitation des bases et outils extérieurs. Les collectivités seront aussi conduites à exploiter des bases de données et des outils génériques réalisés à l'extérieur et notamment en provenance de l'Etat (API classiques, API big data, API entreprise, services de fédération d'identité comme FranceConnect).

La bonne intégration de ces outils dans les applications de gestion comme dans les projets de services des collectivités locales, va devenir primordiale. Elles vont devoir apprendre à exploiter au mieux les capacités nouvelles induites par ces moyens.

Ouverture des données publiques. L'ouverture "par défaut", va inciter les villes et les territoires à réorganiser la gestion de leurs données géographiques et alphanumériques. Cette montée en charge sera plus facile à réaliser dans le cadre de politiques de gouvernance de la donnée et notamment d'urbanisation des systèmes d'information. Les collectivités pourront de cette manière automatiser leurs processus de versement et s'appuyer sur des plateformes associant des outils performants d'accès aux données (géocodeurs, API).

Certaines collectivités seront alors en position de jouer un rôle fédérateur, de tiers de confiance au sein de leurs propres services et auprès des grands producteurs de données présents localement (concessionnaires et entreprises publiques notamment).

Restent à régler les questions de partage de la valeur générée par le mécanisme de l'ouverture, de protection du secret des affaires et de protection des données personnelles. Mais les nouvelles dispositions permettent de fonder de sérieux espoirs.

La Bonne boîte



Sur la base d'algorithmes, cette application restitue aux candidats une liste d'entreprises ayant une forte probabilité d'embaucher dans les 6 mois à venir.

Ce ciblage est obtenu par l'analyse d'une base de données contenant plusieurs millions de déclarations d'embauche émanant de toutes les entreprises françaises (historique de 4 à 5 ans). Ce "module" fonctionne à partir d'une API et peut être réutilisé dans de nombreuses applications.

<http://labonneboite.pole-emploi.fr/>

L'ouverture des données par défaut pourrait bien accélérer la convergence avec les autres approches " data " et " smart city " des villes.

Le développement de la "business intelligence" (BI) et l'exploitation des données numériques issues du big data vont constituer de grands chantiers stratégiques pour la ville. La première approche adresse la problématique du reporting et l'accès aux données à des fins d'analyse et de pilotage, la seconde vise à renforcer cette maîtrise en y ajoutant celle des risques, des flux et de la gestion des actifs sur le territoire.

L'axe Business intelligence est le plus rapidement exploitable, étant donné l'état des compétences et des savoir-faire actuels. Les produits classiques restent de puissants outils d'analyse présentant l'avantage d'être déjà utilisés dans les services. Quant aux nouvelles applications du marché, elles introduisent plus d'ergonomie et d'agilité. Elles permettent notamment de créer des outils d'analyse à la volée, d'explorer les données en profondeur à partir de palettes de visualisation, de révéler des connexions cachées, d'effectuer des "recherches intelligentes" le tout en produisant des tableaux de bord interactifs ou des rapports analytiques dynamiques. Ces technologies proposent des outils sur lesquels les collectivités vont devoir investir mais principalement en élargissant le champ des usages.

Le développement des data sciences appliquées au territoire est en phase d'initialisation. Quelques projets sont lancés, le plus souvent soutenus par de grands prestataires, mais l'appropriation directe des outils de big data n'est pas encore effective. Les grandes collectivités vont mobiliser les data sciences, mais graduellement, et en faisant pour l'instant appel à des prestataires extérieurs pour lancer des expérimentations. L'irruption de l'Internet des objets offre potentiellement l'opportunité d'une mise en place de plateformes dédiées au big data (projet à Nice) et au développement du monitoring urbain. Les postes de commandement multifonctions seront conduits à s'appuyer sur de puissants systèmes de calcul pour accompagner la prise de décision dans les domaines de la prévention et de la gestion de crise. Deux dimensions stratégiques en milieu urbain qui adressent aussi bien le trafic que les risques naturels.

La sécurité des systèmes d'information fait partie des éléments à renforcer dans les dispositifs de gouvernance de la donnée pour assurer la disponibilité des services, garantir l'intégrité des données et prémunir les systèmes des risques de cyber attaque. En effet les collectivités territoriales n'ont pas encore suffisamment pris conscience de la nécessité de sécuriser leurs systèmes d'information alors même qu'elles sont des cibles potentielles.

Les formations pour sensibiliser les agents sont encore rares, les systèmes d'information ne sont pas toujours conformes au référentiel général de sécurité et le recours au chiffrement des données est quasi inexistant (tunnel sécurisé et crypté pour accéder au système d'information de l'extérieur), quant aux audits de sécurité et tests d'intrusion pour caractériser les vulnérabilités et évaluer les risques ils restent peu fréquents. Or les collectivités sont potentiellement de cibles d'attaques qui pourraient se traduire par des pertes de données sensibles.

Les villes et agglomérations qui projettent d'assurer des missions de tiers de confiance de la donnée sur leur territoire seront également tenues de mettre en place de

véritables programmes de sécurité et à prendre des engagements vis à vis de leurs mandants, comme s'apprête à le faire La Poste dans une charte qui prévoit de donner aux usagers le contrôle de leurs données, de protéger les données et de rendre compte de leurs traitements et de garantir aussi les collaborations de confiance entre partenaires.

Le fait de centraliser l'information dans des entrepôts de données, informations qui peuvent se révéler sensibles, imposera aux grandes agglomérations le recrutement de responsables de la sécurité des systèmes d'information (RSSI). Le développement de la gouvernance des données peut être une occasion de remettre à plat ces questions afin de dresser des barrières de défense plus solides.

CYBERDEFENSE : NOUVEL ENJEU DE SECURITE NATIONALE

Juridique

Afin de profiter des progrès des technologies de l'information et de la communication, les infrastructures critiques – entendues comme toutes les infrastructures (y compris les services), considérées comme essentielles au fonctionnement de l'économie et de la société – reposent de plus en plus sur des systèmes d'information complexes largement connectés à des réseaux publics.

Ces interconnexions rendent les infrastructures interdépendantes et les exposent à l'ensemble des vulnérabilités des systèmes informatiques.

Dans ce cadre, le programme européen PEPIC a vocation à fixer un cadre commun, au niveau de l'Union européenne, pour la protection des infrastructures critiques en Europe afin de s'assurer que tous les États membres offrent des niveaux de protection suffisants de ces infrastructures.

Au niveau national, si un haut degré dans la diffusion et l'usage des systèmes d'informations a été atteint, l'Etat français n'a pas encore accordé suffisamment d'importance à la sécurité de ces systèmes.

Or, face à l'espionnage informatique, la problématique de la sécurité des systèmes d'information des entreprises - et notamment de celles des secteurs jugés stratégiques - représente un enjeu majeur.

Par ailleurs, en la matière, les collectivités territoriales n'ont pas encore suffisamment pris conscience de la nécessité de sécuriser leurs systèmes d'informations, alors même qu'elles sont bien évidemment des cibles potentielles.

Si des dispositifs de protection ont été mis en place par l'ANSSI, permettant aux collectivités territoriales de sécuriser leurs systèmes d'information, celles-ci peinent à atteindre un niveau minimal de protection de leurs données, et sont, de fait, mal préparées aux futures révolutions apportées par le guichet unique et la ville connectée. Une vigilance renforcée et des moyens supplémentaires devraient être mis en œuvre pour améliorer le cadre d'ensemble des dispositifs.

➤ Des structures de mutualisation informatique ouvertes, pour les territoires

Dans les petites villes et territoires ruraux, comme indiqué dans le chapitre précédent, les conditions d'organisation et d'exploitation diffèrent mais les besoins de gouvernance de la donnée existent dans plusieurs domaines comme le tourisme, l'énergie et la gestion administrative.

Pour atteindre une taille critique la mutualisation semble être, au stade actuel, la solution la plus prometteuse pour ces territoires. Les actions pourront s'orienter sur quatre modalités principales :

- **Mutualisation des financements** dans le cadre de regroupements autour d'opérateurs spécialisés (syndicats mixtes de mutualisation, conseils départementaux, centres de gestion ou groupements d'intérêt public) ou via des groupements de commande ou des ententes entre collectivités ;
- **Mise en commun de compétences** pour réunir les capacités en ingénierie de projets autour des systèmes d'information et de la donnée ;
- **Partage d'outils numériques** via des plateformes de services et des API permettant aux entités publiques d'accéder à des services de qualité ;
- **Regroupement de moyens d'animation, de formation, et de conduite du changement** seront également cruciaux pour développer une culture de la donnée.

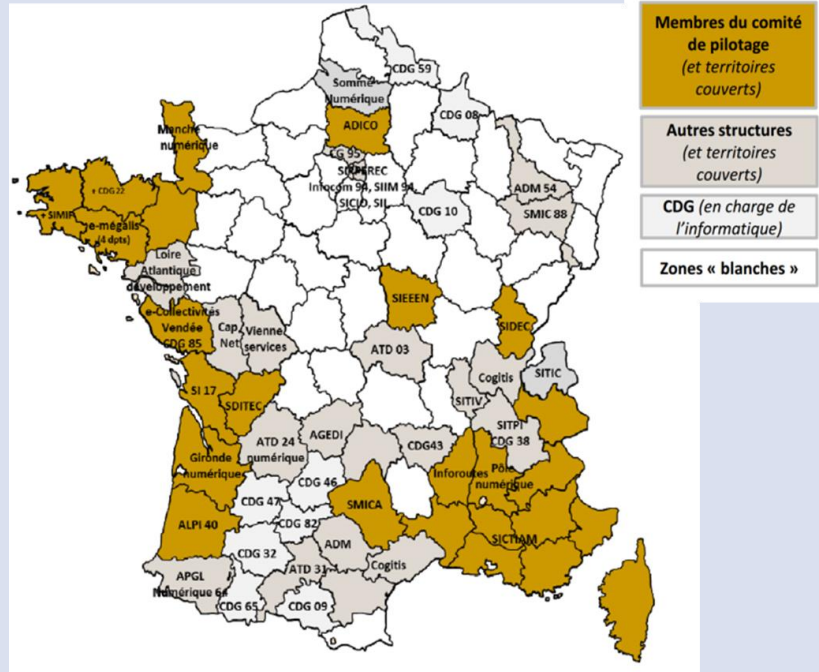
Dans ce cadre, la maille départementale reste la bonne échelle : elle donne de la visibilité sans être trop éloignée de la proximité et surtout elle présente déjà des structures et des opérateurs capables de construire des infrastructures de données publiques performantes et mutualisées :

Les départements sont en position d'accompagner la structuration des communes et des EPCI de leur territoire, dans le cadre de leurs missions de solidarité territoriale – renforcée depuis la loi NOTRe- pour autant qu'ils en aient la volonté.

Les structures de mutualisation informatiques (syndicats mixtes informatiques, centres de gestion, etc.) jouent déjà ce rôle fédérateur sur la gestion des collectivités locales. Ils semblent par conséquent bien placés pour se positionner sur la gestion des données, en appui direct du bloc communal.

De telles structures, pourraient assurer des missions de mise en conformité des communes autour de la dématérialisation, préparer la mise en place de plateformes d'e-services à l'échelle départementale réutilisées par les communes en marque blanche, assurer une veille sur les nouveaux outils publics et s'ouvrir au développement des data sciences. La mutualisation serait sans doute aussi la voie la plus prometteuse pour développer des outils marketing dans les zones touristiques, ou pour l'appui au développement des circuits courts agricoles à partir de plateformes agrégatives.

Départements couverts par une structure de mutualisation informatique (SMI)



SOURCE : FNCCR-INFHOTEP 2015

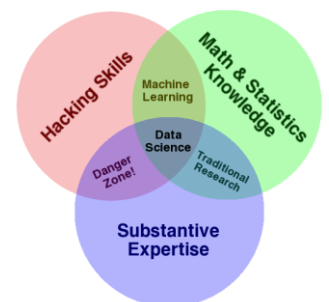
Près de 60% du territoire français sont couverts par une structure de mutualisation informatique.

3. Appui au développement des data sciences et des technologies décisionnelles

Les data sciences constituent un levier à très forte valeur ajoutée et recèlent un réel potentiel d'amélioration. Aussi, l'enrichissement des politiques publiques reposera sur la croissance du marché des data sciences et sur leur contribution au développement et à l'aménagement des territoires.

➤ *Acculturation du secteur public*

Les data sciences et l'exploitation du big data en général constituent un champ prometteur à défricher dans le secteur public. Pour obtenir des résultats probants, une phase d'acculturation des acteurs locaux constitue un préalable. Elle repose sur un travail d'information fondé sur les réalisations à valeur d'exemple afin d'illustrer le potentiel du big data. La sensibilisation passera par l'organisation de campagnes d'action et de séminaires ciblant les élus et le réseau de correspondants locaux ainsi que par la création d'outils de communication dédiés : site web, MOOC, applications de e-learning, maquettes, etc.



Exemple : Datalab de la Région
Pays de la Loire

Pour aller plus loin vers la mise à disposition d'une capacité de maîtrise d'ouvrage sur les projets, une task force orientée data sciences pourrait être constituée à l'échelle régionale. Elle serait chargée de mobiliser et de mettre à contribution les entités publiques et privées (labos de recherche, start-up, sociétés informatiques) sur la conduite de projets en data sciences. La recherche de compétences dans ce domaine apparaît primordiale, ne serait-ce pour la rédaction de cahiers des charges et pour engager la réalisation de POC³⁷ ou de projets collaboratifs (énergie, tourisme, administration, mobilité, santé).

➤ *Partenariats avec le privé*

Pour stimuler le marché du big data et fournir aux entreprises l'occasion de développer des solutions adaptées aux besoins des collectivités, ces dernières peuvent choisir d'établir des partenariats avec des acteurs privés sous plusieurs formes :

- Un système de licences ou de redevances pour l'accès aux données et aux services,
- Le partage de données dans le cadre d'expérimentations,
- La commande ou le co-développement de nouveaux outils au service de l'aménagement urbain et du développement économique, prenant appui sur les data sciences.

Un rapprochement du secteur public avec le monde universitaire serait également le moyen de stimuler la recherche dans le domaine des data sciences à travers le montage de projets de recherche et leur participation aux expérimentations, l'accueil de doctorants et la création de data labs dédiés.

4. Renforcer la souveraineté sur les données

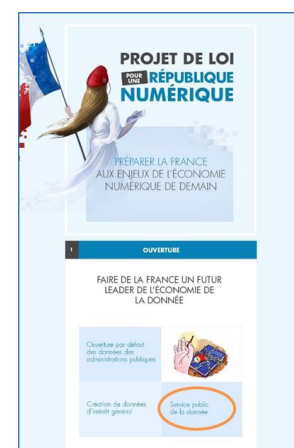
Pour porter la politique publique nationale de la donnée, l'article 14 du projet de loi pour une République numérique instaurait initialement la création d'un service public national de la donnée. Dans la loi du 7 octobre 2016, le Parlement a fait disparaître le service public et le remplace par le terme plus vague de **"Mission de service public relevant de l'Etat"**.

Pour décliner ce service localement, les territoires devront s'organiser ensemble.

Teralab



L'institut Mines Télécom opère une plateforme Big Data dédiée à la recherche, à l'innovation et à l'enseignement. Ouverte en avril 2014 elle héberge à ce jour plus de 20 projets dans les domaines diversifiés : mobilité, assurance, logistique, énergie, e-commerce, administration. Cette plateforme pourrait servir d'infrastructure pour l'expérimentation de projets territoriaux. De plus, il existe déjà des outils qui pourraient probablement se rapprocher des territoires pour des expérimentations.



³⁷ « Proof Of Concept », en français « démonstrateur »

LOI POUR UNE REPUBLIQUE NUMERIQUE : DE LA NOTION DE SERVICE PUBLIC A CELLE DE "MISSION DE SERVICE PUBLIC "

Juridique

L'article 14 de la loi pour une République numérique évolue dans sa formulation de la gestion des données publiques :

Il substitue à **la notion de service public de la donnée** (formulée dans le projet de loi initial) celle de **"mission de service public"** relevant de l'Etat.

Mission consistant en **« la mise à disposition des données de références »**.

Les collectivités territoriales seront directement impliquées dans le processus. En effet, la définition de l'Etat est à prendre au sens large comme le précise la loi :

« Toutes les administrations mentionnées au premier alinéa [soit l'Etat, **les collectivités territoriales et leurs groupements** ainsi que les personnes morales de droit privé ou de droit public concourant à l'exercice de mission de service public] **concourent à cette mission. »**

« La notion de service public de la donnée reconnaît l'importance de garantir l'accès à ces données au niveau de qualité requis par les innovateurs, et le besoin de s'organiser pour le faire »

Henri Verdier

➤ *Création d'un service public national de la donnée*

Dans le cadre des discussions sur le projet de loi pour une République numérique, l'Etat a réfléchi aux missions qui pourraient incomber à un futur service public de la donnée. Il faut reconnaître que le positionnement demeurerait imprécis.

Désormais, le texte de la loi du 7 octobre 2016 ne mentionne plus explicitement une telle création mais indique que "la mise à disposition de données de références en vue de faciliter leur réutilisation" constitue une mission de service public relevant de l'Etat. Toutes les administrations – Etat, collectivités territoriales et autres personnes de droit public- concourent à cette mission.

Les données de référence, sont des données publiques qui "constituent une référence commune pour nommer ou identifier des produits, des services, des territoires ou des personnes". Ce sont des données qui sont "réutilisées fréquemment par des personnes publiques et privées autres que l'administration qui les détient" et qui par conséquent nécessitent "qu'elles soient mises à disposition avec un niveau élevé de qualité".

C'est en principe un décret en conseil d'Etat qui précisera les modalités de participation et de coordination des différentes administrations ainsi que les critères de qualité qu'elles devront respecter. Il dressera également la liste des données de référence et désignera les administrations responsables de leur production et de leur mise à disposition. L'Etat aura donc la charge de définir et d'identifier les données de référence, d'harmoniser le cadre et les conditions de leur production, de fixer des normes de qualité et des modalités de diffusion.

Le Secrétariat général pour la modernisation de l'action publique (SGMAP) sera très vraisemblablement chargé d'assurer cette mission de service public ce qui pose la

question du rôle des collectivités territoriales qui en raison de leur position et de leurs missions figureront parmi les principaux fournisseurs de données de référence. Comment seront-elles associées et comment la charge du financement sera-t-elle répartie ? Quelques amendements de parlementaires ont bien proposé que les collectivités territoriales partagent avec l'Etat la gouvernance de cette mission, mais ils ont tous été rejetés, le gouvernement renvoyant le débat aux discussions et négociations préparatoires prévues dans le cadre de la rédaction du futur décret.

➤ *Des missions locales de service public pour la gestion des données d'intérêt général à la maille régionale*

Si la question des données de référence est cruciale en termes de valeur ajoutée et de service, elle ne constitue, pour les collectivités territoriales, qu'une facette de la gouvernance de la donnée.

Au-delà se posent des questions d'organisation des entités publiques locales impliquées dans cette gouvernance, d'élargissement de l'ouverture des données publiques et des données des entreprises (lorsqu'elles sont d'intérêt général), de prise en compte des inégalités territoriales notamment par rapport à la capacité des collectivités à financer et mettre en œuvre les moyens de production et de diffusion nécessaires à cette ouverture et d'appropriation des data sciences pour le développement des territoires.

Tous ces sujets induisent de nouvelles formes de collaboration, d'échanges, de mutualisation de projets et donc de coordination dans la mise en œuvre d'infrastructures de données publiques locales. Une coordination et des moyens communs qui pourraient être pilotés et gérés par un service local de la donnée. Cette entité serait légère, flexible, à géométrie variable, au service du secteur public et aussi des acteurs économiques locaux.

Elle interviendrait plutôt à l'échelle régionale pour garantir la cohérence des actions et disposer de leviers d'action suffisants mais serait tenue à un devoir de neutralité et à des règles de confidentialité équivalentes à celles de tiers de confiance, afin d'accompagner si nécessaire une filière économique dans sa stratégie de redéploiement.



Une mission locale de service public à géométrie variable

L'hétérogénéité des approches et des contextes territoriaux, induit une grande flexibilité dans la structuration et l'organisation de cette mission. Néanmoins, elle rassemblerait si possible quatre familles d'acteurs, dans une configuration à déterminer :

- Les Régions et les Départements,
- Les métropoles et les grandes agglomérations,
- Les opérateurs techniques présents sur le territoire, qui accompagnent notamment les plus petites collectivités (syndicats, structures de mutualisation sur le numérique),
- Les représentants des acteurs économiques locaux.

Le service en lui-même se situerait plutôt à la maille régionale, mais la Région (tout comme les Départements), bien que largement impliquée, n'en serait pas le pilote. Une structure ad hoc pourrait endosser ce rôle ou bien être créée spécifiquement.

La maille régionale semble aujourd'hui l'échelle pertinente pour installer cette mission locale de service public de la donnée. Elle offre une visibilité étendue surtout après le redécoupage du territoire national en 13 unités régionales ; à ce niveau s'exerce déjà une coordination sur les données géographiques à travers les CRIGE et



certaines structures de mutualisation publiques (ex : GIP ATGeRi). La loi NOTRe a d'ailleurs officialisé cette coordination en la confiant à la région avec la charge de constituer une plateforme de services numériques qu'elle anime, pour en favoriser l'accès et la réutilisation. La proximité et l'implication des instances régionales en tant que chef de file de la stratégie d'aménagement du territoire et du développement économique, serait également des atouts pour la maille régionale sachant que s'y développera progressivement une culture de la collaboration et de la concertation avec l'ensemble des autorités publiques locales³⁸.

Missions

La structure de gouvernance de la donnée à la maille régionale devra mener 4 missions principales :

- Développer la gouvernance de la donnée à l'échelle du territoire en assurant l'équilibre entre territoires urbains et territoires à dominante rurale,
- Sensibiliser et accompagner les acteurs publics et privés,
- Dynamiser le développement du territoire,
- Coopérer avec les autres niveaux de collectivités territoriales et avec le privé.

Mission 1 : Gouvernance de la donnée

- **Création d'un schéma des données d'intérêt général** sur le territoire, d'un magasin d'API et d'un annuaire de métadonnées : déclinaison au niveau régional des outils susceptibles de valoriser le territoire régional ;
- **Mise en place d'une infrastructure de données publiques**, incluant la création d'une plateforme et d'un volet open data ;
- **Constitution de modèles de données métiers ;**
- **Veille et partage de bonnes pratiques sur l'utilisation et la réutilisation des données de référence et les API** diffusées par l'Etat et les autres niveaux territoriaux.

Mission 2 : Sensibilisation, accompagnement

- **Mobilisation des acteurs et des territoires :**
 - Initiation, diffusion de bonnes pratiques et de retours d'expériences dans les organisations privées et publiques,
 - Développement de nouvelles filières de formation,
 - Promotion de l'innovation,
- **Accompagnement du monde économique** à travers trois axes :
 - Veille économique et entrepôts de données partagés,
 - Organisation des filières économiques qui en expriment le besoin,
 - Appels à projets sur des actions mobilisant les data sciences.

Mission 3 : Dynamiser le développement du territoire

Pour dynamiser à la fois le développement du territoire et la gouvernance de la donnée, la mission locale pourrait lancer la **commande de plusieurs services structurants** tels que :

- Tableau de bord économique régional dynamique (partenariat INSEE),
- Applications mobiles de données dynamiques permettant aux utilisateurs de créer leur tableau de bord de vie quotidienne (transports, événementiels...)

³⁸ La suppression de la clause de compétences générale au niveau de la région impliquera de fait l'obtention d'un consensus pour le montage de certains projets par la concertation, le dialogue et la collaboration

- Tableau de bord e-gov de l'utilisateur, guide interactif du touriste ou du voyageur en déplacement.

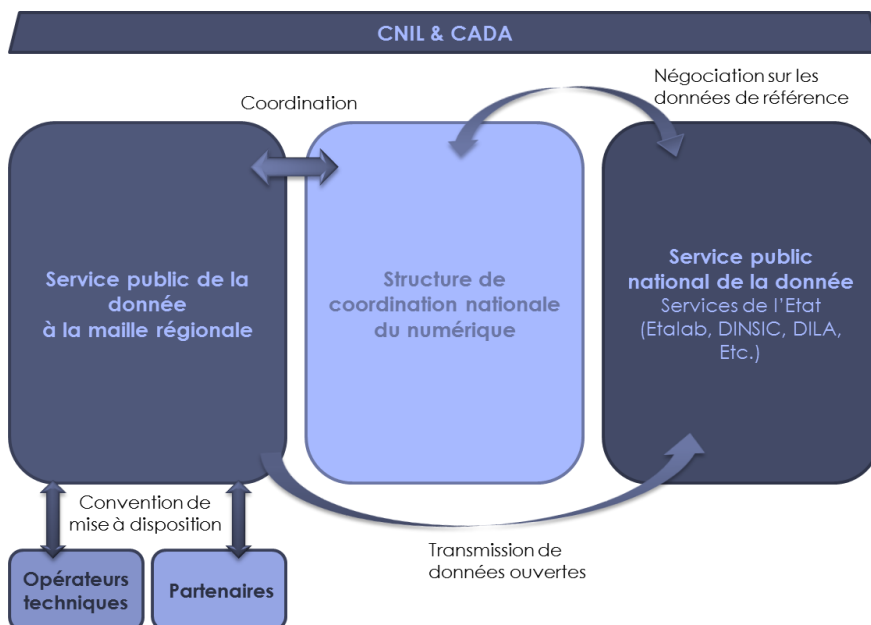
Dans cette même logique, la Mission pourrait s'engager sur **le chantier des data sciences** en s'associant avec les French Tech locales, les laboratoires de recherche, les universités et les écoles d'ingénieurs, en développant des services d'intelligence économique et en montant des dossiers de demande de financements européens sur des projets d'intérêt régional.

Mission 4 : Coopération avec les autres collectivités et le privé

La structure à la maille régionale aura aussi pour mission de s'assurer de la bonne participation des acteurs du territoire à cette mission locale de service public de la donnée. Elle devra donc coopérer avec différents acteurs :

- **Avec les métropoles et agglomérations du territoire régional**, notamment dans le cadre des échanges de bonnes pratiques et de savoir-faire et dans l'exploration des domaines d'application des data sciences,
- **Avec les autres structures régionales** : une coopération interrégionale entre les 13 entités pourra être organisée au sein d'une structure de coordination nationale sur le numérique. Celle-ci permettrait le travail entre acteurs à la maille régionale, sans intervention de l'Etat et irait au-delà de la question des données pour inclure tous les aspects numériques (télécom et applications) et pourrait également inclure les métropoles.
- **Avec le niveau national** : les structures portant les services publics locaux de la donnée, pourront prendre part dans l'instance de concertation du futur service national de la donnée, à travers la structure de coordination nationale sur le numérique par exemple.
- **Avec le privé et la recherche** à travers des partenariats avec
 - Les entreprises,
 - Les laboratoires de recherche et le monde universitaire,
 - Les associations (médiation numérique, consommateurs, etc.).

Structure de coordination nationale sur le numérique
Mission 1 : Infrastructures télécom
Mission 2 : Services, applications
Mission 3 : Données



Organisation et structuration des missions locales

Afin d'assurer la gouvernance de la donnée sur le territoire régional, le service se dotera d'une **instance de pilotage**, chargée du déploiement d'une infrastructure de données publiques sur le territoire, de la valorisation des données publiques, de la diffusion des nouveaux usages de la donnée et de définir de nouvelles stratégies d'action.

La structure pourra être pilotée par un **Chief Data Officer**, en charge des missions d'animation et de coordination avec les acteurs publics et privés du territoire pour construire une gouvernance de la donnée. Il participera également au lancement de chantiers structurants pour dynamiser l'écosystème local.

La structure devra par ailleurs acquérir des compétences en maîtrise d'ouvrage sur la donnée. En effet, la transformation des organisations publiques, la mise en place de nouveaux écosystèmes d'information et de nouvelles architectures suppose une maîtrise à la fois des méthodes et des moyens nécessaires à la mise en œuvre de ces ressources. Pour ce faire, le service public local pourra s'appuyer sur des compétences et des opérateurs extérieurs pour expérimenter, développer de nouveaux projets à valeur d'exemple et accompagner les acteurs régionaux dans le développement d'une gouvernance de la donnée.

La forme juridique de cette structure sera à géométrie variable pour s'adapter au contexte et au jeu des acteurs local. Elle devra a minima présenter les caractéristiques suivantes :

- Etre la plus **ouverte** possible afin d'y inclure l'Etat, les collectivités, les entreprises et les structures de mutualisation informatiques, etc.
- Offrir des garanties de neutralité et de confidentialité dans son fonctionnement pour créer la confiance auprès des entreprises et des citoyens.

Plusieurs formes de structures juridiques pourraient répondre à ces critères comme les groupements d'intérêt public, les GIE, les associations, sachant que le choix sera défini en fonction des caractéristiques du territoire : forces en présence, existence d'entités susceptibles de porter le service, présence des entreprises dans la gouvernance de la structure, solidité juridique et formalisme administratif inhérent à chaque statut, délais de création, mise en place ou non d'une structure de préfiguration..

ZOOM SUR LE STATUT DES GIP

Juridique

Au regard, tant des enjeux que des comptes rendus d'entretiens ayant eu lieu dans le cadre de la présente étude, il existe aujourd'hui une très grande hétérogénéité dans les processus de gestion de la donnée par les collectivités territoriales sur le territoire national.

Si la tendance paraît s'orienter vers un pilotage à l'échelon « régional » de la gestion de la donnée, des interrogations persistent sur les modalités juridiques les plus adaptées pour confier la gestion et le traitement de la donnée à un guichet unique institué à un niveau régional.

En effet, il n'existe pas un scénario général adapté à la diversité des situations des collectivités ainsi qu'à la complexité des sujets se rapportant à la gestion de la donnée territoriale.

Dans ce cadre régional et au vu des échanges ayant eu lieu dans le cadre de la présente étude, le Groupement d'Intérêt Public (GIP) nous paraît être la structure d'intervention et de coopération la plus adaptée pour le portage de la future plateforme applicative.

En effet, le recours au GIP nous semble présenter de nombreux avantages :

- Il apparaît comme la structure d'intervention et de coopération la plus souple, pérenne et compatible avec les problématiques liées à l'impossibilité de contraindre les collectivités territoriales à l'externalisation de la gestion de leurs données ;
- Il permettrait d'accueillir de nombreux partenaires pour la mise en œuvre d'un véritable big data territorial ;
- Les membres du GIP peuvent partager les coûts d'achat et de développement des services mutualisés ;
- Les collectivités membres du GIP peuvent recevoir des services du GIP sans le mettre en concurrence.

Le recours aux structures existantes ou aux plateformes d'ores et déjà en place au niveau local ne devra néanmoins pas être négligé.

Membres potentiels et gouvernance

Les entités suivantes pourraient prendre part à la gouvernance de la structure en charge de cette mission locale de la donnée :

- L'Etat à travers ses services déconcentrés et ses préfetures,
- Les collectivités territoriales : la Région, les Départements, les EPCI (et notamment les métropoles),
- Les établissements publics (hôpitaux, universités, SDIS, etc.) qui gèrent souvent une grande quantité de données,
- Le CRIGE et les infrastructures de données géographiques,
- Les citoyens ou leurs représentants,
- Les opérateurs techniques.

Sur le plan pratique, la gouvernance pourrait fonctionner avec un ou plusieurs collèges selon le contexte local. Le groupement d'intérêt public e-bourgogne est par exemple constitué de 13 collèges disposant chacun d'une voix à l'exception du collège rassemblant les membres fondateurs qui disposent chacun d'une voix.

5. Opérateurs techniques

Les opérateurs techniques recouvrent l'ensemble des structures locales dotées d'une capacité en ingénierie de projets et services numériques. Elles peuvent disposer d'une plateforme de services et de moyens d'assistance et d'accompagnement des collectivités territoriales dans la mise en place d'une gouvernance par la donnée à l'échelle régionale incluant de nombreuses entités publiques et, le cas échéant, privées.

Ils représentent une diversité d'acteurs tant publics que privés :

- Les opérateurs publics qui peuvent comprendre les DSI mutualisées des collectivités, les services informatiques des départements, les syndicats informatiques de mutualisation, les centres de gestion, les agences techniques départementales ainsi que des structures particulières susceptibles de revêtir une forme associative.
- Les opérateurs privés qui comprennent notamment des éditeurs engagés dans la constitution d'infrastructures de partage de données
- Les opérateurs publics-privés, rassemblés au sein d'associations de droit privé, ou de sociétés coopératives d'intérêt collectif (SCIC), comme l'association Ozwillo se prépare à le devenir.

Pour assurer la mise en œuvre opérationnelle des projets portés par les missions locales de service public de la donnée ces structures seront en mesure de jouer un rôle de maîtrise d'ouvrage ou d'assistance à maîtrise d'ouvrage dans le pilotage de la transition numérique, la mise en réseau des acteurs et le développement d'une infrastructure de données publiques

IV. Recommandations

Juridiques	Préciser et poursuivre la dynamique de la loi pour une République Numérique
	Affirmer la responsabilité des collectivités en termes de maîtrise des données au sein des contrats
Opérationnelles	Couverture nationale de structures de mutualisation informatiques
	Affirmation de la création d'une Infrastructure de données publiques locale
	Mobilisation sur le développement des data sciences comme enjeu d'aménagement et de dynamisation des politiques publiques locales
Coordination et groupes de travail	Coordination nationale de tous les acteurs de la data sur les modèles de références
	Observatoire des données et des structures de gouvernance des données publiques
	Articulation avec le niveau européen
	Participer à la réflexion sur la protection des infrastructures sensibles

A. Poursuivre la dynamique de la loi pour une République numérique

Le texte étant désormais adopté, l'attention devra se porter sur la préparation des décrets d'application qui viendront préciser cette loi ainsi que ceux qui se rapportent aux lois précédemment votées (loi de transition énergétique notamment).

Ainsi, cinq points de vigilance sont à suivre :

La notion de service public de la donnée. Elle a sensiblement évolué depuis le projet de loi, puisque cette innovation a été effacée pour passer au rang de "**mission de service public**" dans la loi du 7 octobre 2016.

Toutefois certaines dispositions devront être clarifiées afin d'instituer une gouvernance adéquate :

- Les modalités de choix des bases de données susceptibles d'être qualifiées en bases de données de référence et leurs modalités de gestion ;
- La liste des données de référence et les administrations responsables de leur production et de leur mise à disposition ;
- Les critères de qualité à respecter pour les données de référence mises à disposition

- Les modalités de participation et de coordination des administrations afin que les données de référence constituent bien une compétence partagée entre l'Etat et les collectivités territoriales.

Les valeurs équivoques associées à la mise à disposition des « données de référence ». Le législateur les a définies comme étant une "référence commune pour nommer identifier des produits, des services, des territoires ou des personnes", réutilisées fréquemment, ce qui nécessite un « niveau élevé de qualité ». Toutes ces notions donnent une idée générale de ce que doit être une donnée de référence, mais il conviendra de veiller aux éventuelles précisions qui seraient apportées ultérieurement par le pouvoir réglementaire.

Le rôle des collectivités territoriales au regard de leur capacité à s'administrer librement. L'attention sera portée sur les données dont elles ont la propriété, notamment sur les thématiques « Transports » et « Energie ». A cet égard la vigilance devra également cibler les échanges en cours relatifs aux modalités de rédaction des décrets d'application liés aux autres lois constitutives d'une gouvernance des données territoriales et notamment :

- La loi pour la croissance, l'activité et l'égalité des chances économiques (dite « **Loi Macron** ») (Loi n° 2015-990 du 6 août 2015) ;
- La loi de transition énergétique pour la croissance verte, dite « **Loi TECV** » (Loi n°2015-992 du 17 août 2015).

La capacité de s'opposer à la récupération en direct des données dont les collectivités territoriales et leurs établissements publics seraient propriétaires, sans leur consentement. Les collectivités locales restent en effet opposées à la transmission automatique et directe à l'Etat des données recueillies par leurs concessionnaires.

La détermination des textes législatifs qui resteront à adopter après ceux qui auront été validés dans le cadre de la loi pour une République numérique.

En tout état de cause, et au vu de la diversité des « chantiers législatifs » ouverts et à ouvrir sur la construction d'une gouvernance des données, un travail de mise en cohérence des différentes notions et définitions figurant dans des textes épars, va s'imposer. Ce travail de clarification pourrait s'inscrire dans le cadre d'un projet de loi de modernisation et de simplification du droit au regard de l'enjeu que représente cette gouvernance.

B. Affirmer, dans les contrats, la maîtrise des collectivités sur leurs données

Il n'est pas toujours aisé pour les collectivités territoriales d'obtenir, de la part de leurs partenaires privés, l'ensemble des informations relatives aux services qu'elles délèguent alors que tout contrat public donne aujourd'hui lieu à la production d'un volume croissant de données. Aussi, la collectivité doit pouvoir disposer de ces données pour ses propres besoins :

- Afin d'assurer leur mise à disposition dans le cadre des obligations qui sont à sa charge au titre de l'open data ;
- Afin d'assurer efficacement le suivi, l'évolution et le renouvellement de ses contrats de délégation de service public.

En conséquence Il appartient aux collectivités territoriales de décider du degré de contrôle qu'elles souhaitent exercer sur la gestion des données des services publics délégués.

Un tel sujet implique nécessairement une maîtrise du dispositif de protection des données personnelles, lequel se préoccupe peu des relations susceptibles d'exister entre une personne publique et son cocontractant.

LA RESPONSABILITE ADMINISTRATIVE ET PENALE LIEE A LA PROTECTION DES DONNEES PERSONNELLES

Juridique

Le respect, par les collectivités locales, des règles de protection des données à caractère personnel est un facteur de transparence et de confiance à l'égard des usagers. A cet égard, les conditions de licéité des traitements de données à caractère personnel peuvent être résumées autour de 5 notions clefs :

- La finalité ;
- La pertinence ;
- La conservation ;
- Le respect des droits des personnes ;
- La sécurité.

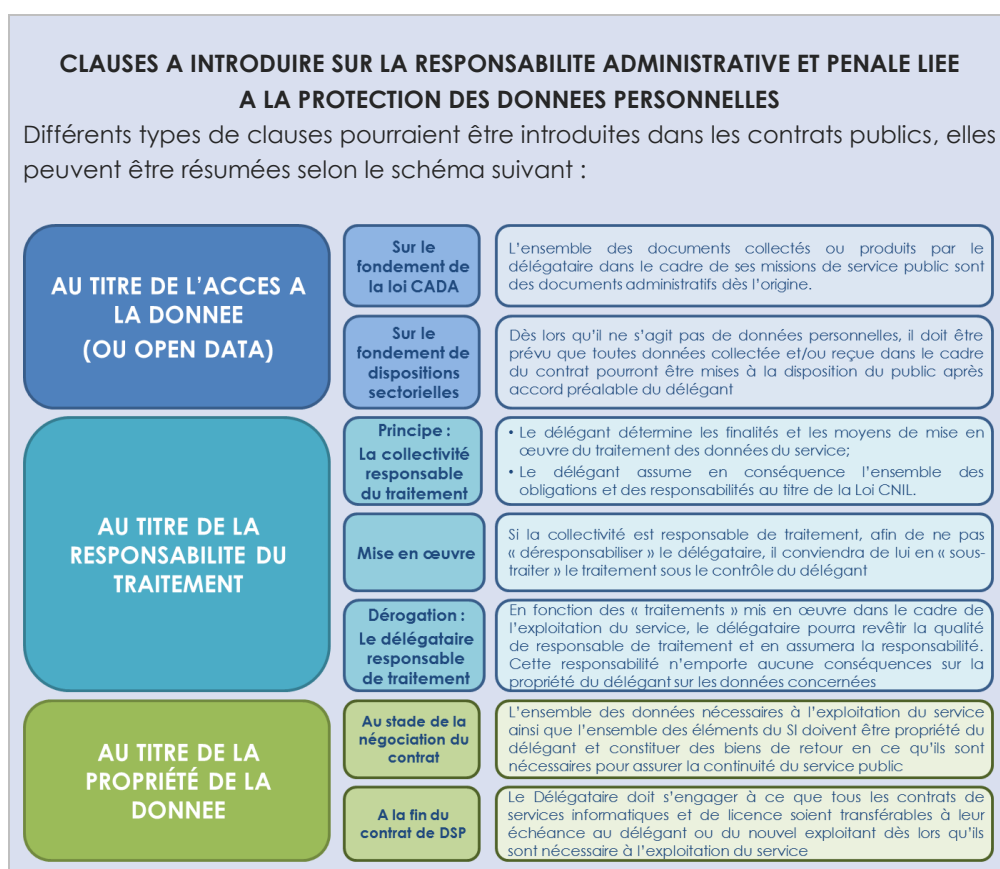
La protection des données personnelles fait l'objet d'une double protection dans la loi CNIL :

- Administrative : la CNIL, dès lors qu'elle est compétente, dispose d'un pouvoir de sanction en cas de réutilisation contraire à une disposition de la loi CNIL. La loi CNIL renvoie enfin, aux termes de son article 50, pour les sanctions pénales, aux dispositions des articles 226-16 à 226-24 du code pénal, relatifs aux atteintes aux droits de la personne résultant des fichiers ou des traitements informatiques ;
- Pénale : toute atteinte aux droits de la personne résultant des fichiers ou des traitements informatiques est réprimée pénalement aux termes des articles 226-16 et suivants du code pénal.

Aussi pour que les collectivités territoriales soient parfaitement en adéquation avec le contrôle qu'elles souhaitent exercer sur la gestion des données des services publics délégués, ils doivent assurer pleinement leur responsabilité dans le traitement de la donnée, prendre la responsabilité de traitement et ne pas se contenter de la sous-traiter au délégataire comme c'est souvent le cas.

Dès lors que la collectivité délégante déterminerait les finalités et les moyens de mise en œuvre du traitement des données du service, elle serait considérée comme responsable du traitement et assurerait à ce titre l'ensemble des obligations prescrites par la Loi Informatique et Libertés ;

En tout état de cause il reviendrait au concessionnaire, en qualité de "sous-traitant", au sens des dispositions de la loi CNIL, d'assurer la confidentialité et la sécurité des données du service conformément aux engagements inscrits dans le contrat pour la couverture des risques résiduels. En sa qualité de sous-traitant, le concessionnaire ne pourrait alors agir que sur instruction du Délégrant.



Dans la perspective d'une prise de responsabilité plus étendue des collectivités en tant que collecteur et producteur des données un certain nombre de dispositions générales devront en tout état de cause être prises :

1) Harmoniser la rédaction des contrats et la prise de responsabilité pénale. **L'objectif serait de pouvoir proposer rapidement aux collectivités territoriales des clauses types** en raison de la responsabilité civile et pénale qu'elles endosseraient en tant que collecteur et producteur des données.

2) S'interroger sur le lien contractuel qui pourrait être institué dès avant la conclusion du contrat entre le concessionnaire et le concédant (avant la constitution du cahier

des charges) et sur le niveau de responsabilité qui pourrait être envisagé entre personne publique et personne privée.

3) Affirmer que le système d'information et les données sont des biens de retour pour la collectivité et surtout sont sous la maîtrise de la collectivité.

4) S'assurer de la faisabilité technique de l'exercice du droit de propriété des collectivités territoriales et de la gestion des données liées aux contrats.

5) Assurer une garantie de confidentialité forte pour susciter la confiance des citoyens. Pour cela un personnel assermenté assurera cette responsabilité et ne pourra être soumis à autorité hiérarchique ou politique.

6) Réaliser un livre blanc permanent intégrant l'évolution des pratiques et du savoir-faire porté par la future structure nationale de coordination sur le numérique.

C. Préparer la création de missions locales de service public de la donnée

Les enquêtes et les entretiens effectués dans le cadre de cette étude, confirment la grande hétérogénéité des processus de gestion de la donnée portés par les collectivités territoriales.

Les acteurs sont ainsi en attente d'un cadre de gouvernance neutre et partagé, prenant en compte cet existant afin de mieux l'organiser du point de vue de la collecte, de la gestion et de l'exploitation des données.

Dans cette perspective, la maille régionale apparaît comme l'échelle pertinente pour l'installation d'une structure d'animation et de coordination capable d'assurer le lien entre les besoins de proximité et les exigences de cohérence à l'échelle du pays.

La pérennité d'un projet régional reposera sur la confiance et notamment sur le respect des règles d'utilisation et de réutilisation des données publiques. La définition d'un cadre de gouvernance neutre et partagé, associant territoires, services de l'Etat, devrait apporter cette sécurité ainsi que les moyens de travailler en bonne intelligence.

A cet égard le groupement d'intérêt public (GIP) semble la structure appropriée pour assurer le portage d'une plateforme de services et accueillir un large éventail de partenaires pour la mise en œuvre du big data territorial. Mais d'autres formes pourront être retenues (GIE, association...) en fonction des caractéristiques du territoire. Par souci de cohérence et de coordination aux différentes échelles, chaque instance régionale participera également à :

- Une structure de coopération interrégionale réunissant les treize régions sur le numérique.
- Une instance nationale de concertation avec l'Etat, lieu de prise de codécision sur les grandes orientations et sur les projets de développement

Le pendant de la mission locale est la mise en réseau des acteurs pour assurer la collecte des données et aussi favoriser les échanges et les coopérations.

D. Créer un maillage national de structures de mutualisation informatique

Mutualiser les moyens et les services pour atteindre une taille critique semble être au stade actuel une solution prometteuse pour les territoires. La maille départementale représente l'échelle où s'exerce la solidarité - et se renforce même à ce niveau depuis le vote de la loi NOTRe -, notamment pour appuyer le bloc communal. Le département est en mesure de favoriser cette mutualisation sur le numérique mais il s'appuie le plus souvent sur des opérateurs techniques (Agences, syndicats) capables d'assurer des missions de maîtrise d'ouvrage et de développer des plateformes de services sur le numérique.

Dans cette approche, les structures de mutualisation informatique (Syndicats informatiques, centres de gestion) jouent le plus souvent un rôle fédérateur à travers le conseil, la fourniture de services et l'accompagnement du bloc communal dans l'usage de l'informatique et du numérique.

Certains doivent renforcer leurs compétences et se professionnaliser, mais leur généralisation sur l'ensemble du territoire :

- Apporterait un appui aux EPCI, aux régions et à l'Etat dans le développement des socles accélérateurs sur le numérique, basés sur le cloud, la virtualisation dont le pays a besoin pour entrer pleinement dans la révolution numérique.
- Faciliterait le déploiement d'infrastructures de données nécessaires aux chantiers de big data et de ville et territoires intelligents
- Accélérerait la mutualisation des solutions applicatives métiers dans le cadre de la mise en place de plateformes partagées en marque blanche garantissant aux collectivités une mutualisation enrichissante et valorisante plutôt que confiscatoire.

Il convient donc de favoriser le renforcement de ces opérateurs de l'action publique sur le numérique, non seulement sur la donnée, mais sur l'ensemble de la sphère numérique. L'ampleur des chantiers à mener le justifie pleinement aujourd'hui.

E. Mettre les data sciences au service des politiques publiques locales

La taille des collectivités territoriales demeure un obstacle récurrent pour l'appropriation des technologies génératrices de progrès et de performance. Un constat qui s'applique pleinement aux data sciences et aux outils d'analyse de données des organisations (Business intelligence). Pourtant les besoins émergent ou vont émerger dans les domaines de l'énergie (smart grid), des transports, du tourisme ou encore des risques naturels.

Si les métropoles et les grandes agglomérations commencent à s'engager dans les data sciences pour accroître leur capacité d'action et de développement, les territoires sont aussi concernés et devraient pouvoir bénéficier des mêmes technologies.

Cette évolution passe par la construction d'une stratégie d'action commune qui pourrait être portée par la mission locale de service public de la donnée sur les points suivants :

- Poursuivre et étendre le champ des actions de R&D,
- Lancer des expérimentations en vraie grandeur dans les domaines décisionnels, prédictifs, préventifs et autour de l'amélioration de la gestion,
- Partager, valoriser et diffuser les initiatives et les bonnes pratiques réalisées sur le territoire, au niveau transrégional, national et international.
- **Consécration d'un droit** d'accès aux règles définissant le traitement algorithmique ainsi qu'aux principales caractéristiques de sa mise en œuvre pour tout usager faisant l'objet d'une décision individuelle. Ce nouveau dispositif prévu dans la loi pour une République numérique (voir encadré ci-dessous) s'étendrait donc à toutes les personnes, physiques ou morales, ainsi qu'à toutes les décisions administratives individuelles prises sur le fondement d'un traitement algorithmique.

ALGORITHMIQUE

Juridique

Comme le relevait le Conseil d'Etat dans son Etude annuelle 2014 consacrée au numérique et aux droits fondamentaux, il devient aujourd'hui essentiel de prendre la mesure du rôle d'intermédiation joué par les algorithmes et de concevoir leur encadrement juridique et leur utilisation.

Or, aujourd'hui de nombreuses décisions individuelles prises par l'administration résultent de traitements algorithmiques, c'est-à-dire d'outils soumettant les données entrées par l'administration à une suite d'opérations ou d'instructions permettant d'aboutir à un résultat.

L'article 4 de la loi pour une République numérique introduit un nouvel article L. 311-3-1 au sein du code des relations entre le public et l'administration intitulé « *Droit d'accès aux règles et caractéristiques de l'algorithme intervenu dans la prise d'une décision individuelle* ».

Cet article a vocation à créer, au bénéfice de l'utilisateur qui a fait l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, un droit d'accès aux règles définissant le traitement algorithmique ainsi qu'aux principales caractéristiques de sa mise en œuvre.

F. Structurer nationalement tous les acteurs de la data

La structuration nationale des missions locales de service public de la donnée leur donnera plus d'influence et de visibilité. Elle pourra se matérialiser par la création d'une coordination nationale intégrée dans un ensemble plus large : le réseau national d'échange sur le numérique que les collectivités territoriales pourraient créer. Ce réseau prendrait ainsi appui sur les trois piliers du numérique : les télécoms autour du développement des réseaux très haut débit, les services et solutions applicatives orientés mutualisation et la gestion de la donnée.

Cette coordination des acteurs de la data sera chargée de construire la stratégie inter-régionale de gestion des données publiques (sans intervention de l'Etat) et de monter des actions et projets communs susceptibles de mobiliser des financements publics nationaux et européens.

Elle renforcera l'efficacité d'ensemble de l'édifice régional et local et constituera un acteur crédible et légitime dans la négociation des politiques publiques nationales.

G. Créer un observatoire des données publiques

Pour assurer un suivi efficace de l'évolution des infrastructures de données publiques locales, la création d'un observatoire s'impose. Chargé d'évaluer les politiques publiques locales, il prendra place au sein d'un observatoire national du numérique que les collectivités territoriales souhaitent créer pour suivre également l'évolution de la couverture territoriale en très haut débit et les usages du numérique.

Ce volet consacré à la donnée aura vocation à produire des études et des analyses et notamment les éléments suivants :

- Tableau de bord de suivi du développement des politiques locales.
- Etat des lieux de l'avancement des projets par rapport aux trajectoires de développement établies.
- Recensement et diffusion des initiatives innovantes et des bonnes pratiques.
- Evaluation des économies réalisées et démonstration par l'analyse des coûts.

Cette fonction d'observation offre un triple intérêt :

- Comme instrument de valorisation des démarches et de leur impact économique,
- Comme vecteur d'accélération des projets structurants à la maille régionale
- Comme outil de négociation auprès des instances nationales et européennes.

H. Renforcer la coopération au niveau européen

Un rapprochement vers les institutions européennes sur la thématique des données publiques devient nécessaire notamment pour une meilleure connaissance des politiques et programmes en préparation et aussi pour développer une stratégie d'influence. Ce rapprochement pourrait cibler notamment :

- La mise en place d'une stratégie d'influence à travers la participation à des groupes de travail sur la normalisation (European Innovation Platform) et aux groupes de structuration des appels à projets européens (Smart city and communities) ;
- Le lancement des programmes d'expérimentation basés sur la constitution de consortiums ;
- La veille pour suivre la préparation des règlements ;
- Les échanges de bonnes pratiques et de savoir-faire.

I. Participer à la réflexion sur la protection des infrastructures sensibles

La cyberdéfense devient un enjeu de sécurité nationale auquel les collectivités sont désormais directement confrontées. Elles se doivent de donner l'exemple notamment en s'inspirant de la législation sur les opérateurs d'infrastructures vitales, identifiés par l'Etat comme ayant des activités indispensables, donc plus fortement exposés aux risques que les autres.

L'exposition aux risques (terrorisme) n'est pas négligeable, notamment sur le fonctionnement de certains réseaux d'utilités au niveau local. Aussi une gouvernance par la donnée peut conduire à la mise en œuvre de dispositifs de sécurisation efficaces.

Pour sensibiliser les collectivités territoriales qui n'ont pas encore suffisamment pris conscience de la nécessité de sécuriser leurs systèmes d'information et mettre en œuvre concrètement en place des politiques locales plus performantes.

L'instance nationale sur le numérique à créer constituera un groupe de travail chargé de proposer des solutions de renforcement de la sécurité par la donnée, pour certaines installations sensibles dans les collectivités territoriales, notamment dans les domaines de la gestion de l'eau ou encore dans le cadre de la gestion des postes de commandement (circulation, vidéosurveillance, mobilité) mis en place par les villes et dont le rôle est de plus en plus stratégique notamment en cas de gestion de crise.

V. Annexes

A.	Fiches notions	92
B.	Résultats de l'enquête flash	109
C.	Fiches collectivités	114
D.	Fiches entreprises	125
E.	Analyses juridiques	129
F.	Personnes et entités interrogées	158

A. Fiches notions

1. Données de référence

Définitions

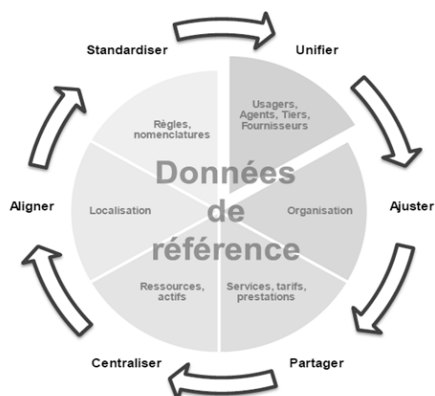
Informatique

Ce sont des **données**

- **partagées** par l'ensemble des processus qui soutiennent l'activité courante d'une organisation et ses prises de décision,
- **stables** (définition stable dans le temps et caractérisée par une période/ de validité),
- **consultées fréquemment**.

Juridique (projet de loi République numérique)

Sont qualifiées de données de référence **celles qui font l'objet ou sont susceptibles de faire l'objet d'une utilisation fréquente par un grand nombre d'acteurs tant publics que privés et dont la qualité, en termes notamment de précision, de fréquence de mise à jour ou d'accessibilité, est essentielle pour ces utilisations.**



Exemple 1 : Facturation des titres de transport

- La liste des titres achetés sont des données opérationnelles : elle varie à chaque commande
- La description du trajet ainsi que l'identité du client sont des données de référence : elles varient peu avec le temps et seront utilisées par diverses applications (gestion des places disponibles, gestion des flottes de véhicules. etc.)



Les données de références (Client, Produit,...) alimentent les processus.

Les données de référence sont le carburant des processus.

Les processus métiers produisent de nouvelles données (Contrat, commande, offre,...) à partir des données de références.

2. Données de référence au sens de la loi pour une République numérique

Définition

Article 14 de la loi pour une République numérique : Il s'agit des informations publiques mentionnées à l'article L. 321-1 qui satisfont aux conditions suivantes :

- Elles constituent une référence commune pour nommer ou identifier des produits, des services, des territoires ou des personnes ;
- Elles sont réutilisées fréquemment par des personnes publiques ou privées autres que l'administration qui les détient ;
- Leur réutilisation nécessite qu'elles soient mises à disposition avec un niveau élevé de qualité.



La notion est directement liée à celle de service public de la donnée

Aux termes de l'article 14 de la loi « *la mise à disposition des données de référence en vue de faciliter leur réutilisation constitue une mission de service public relevant de l'État. Toutes les administrations mentionnées au premier alinéa de l'article L. 300-2 concourent à cette mission.* »

En l'absence de davantage de précisions au sein de la loi, il conviendra d'attendre la publication du décret d'application de ces dispositions qui devra :

- préciser « les modalités de participation et de coordination des différentes administrations » ;
- fixer les critères de qualité que doit respecter la mise à disposition des données de référence ;
- dresser la liste des données de référence et désigner les administrations responsables de leur production et de leur mise à disposition.

3. Données d'intérêt général au sens de la loi pour une République Numérique

Définition

Articles 17 et 18 de la loi pour une République numérique. Il s'agit :

- Des données et bases de données collectées ou produites à l'occasion de l'exécution marchés publics et aux marchés de partenariat dès lors que le titulaire concourt à l'exécution d'une mission de service public et qui sont **indispensables** à l'exécution de ladite mission;
- Et des données « essentielles » de la convention de subvention conclue par l'administration avec un partenaire privé.



La notion de « donnée essentielle » des concessions

L'article 53 de l'Ordonnance n° 2016-65 du 29 janvier 2016 relative aux contrats de concession et l'article 34 de son décret d'application n° 2016-86 du 1er février 2016 relatif aux contrats de concession prévoient que l'autorité déléguante publie, sur son profil d'acheteur et au plus tard le 1er octobre 2018, les « données essentielles du contrat de concession » avant le début d'exécution du contrat de concession, puis chaque année (s'agissant des données financières relatives à l'exécution du contrat de concession) et à chaque modification apportée au contrat de concession.

4. Métadonnées

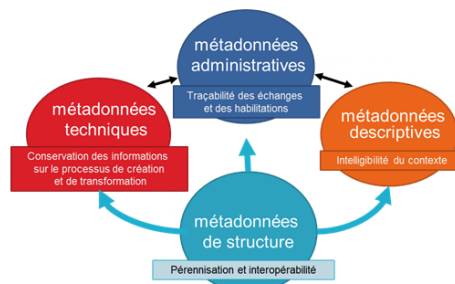
Définition

Ce sont **les données qui décrivent d'autres données**. Elles sont **porteuses d'information sur le contexte, le sens et la finalité de la ressource informationnelle** portée par la donnée brute.

Cet ensemble structuré de données permet notamment de retrouver les ressources stockées.

Les métadonnées doivent non seulement permettre de retrouver des données stockées et de les analyser mais aussi **garantir l'agilité de l'organisme**.

Illustration : différents types de métadonnées

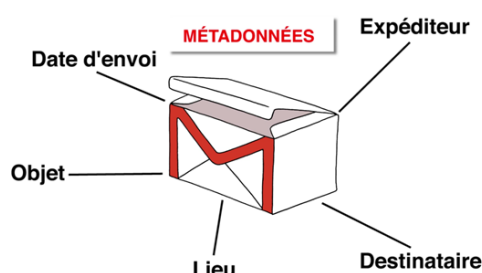


Une notion essentielle à l'interopérabilité

Les données doivent pouvoir être exploitées à tout moment dans le temps, quels que soient les changements subis par l'organisme. Ceci nécessite que les métadonnées soient dès le départ pensées dans une logique qui en permet l'identification par des tiers.

C'est pourquoi des standards de métadonnées métiers et des modèles interopérables sont apparus.

Exemple métadonnées d'un email



5. Données structurées et non structurées

Définition

Les données structurées sont celles dont l'ensemble des valeurs possibles est déterminé et connu à l'avance.

Par exemple, dans une base de données rassemblant les résultats d'une enquête d'opinion, l'âge ou la catégorie socio-professionnelle des individus interrogés sont des données structurées.

Par opposition, les données non structurées sont des données représentées ou stockées sans format prédéfini.

Une problématique à l'origine du Big Data

La problématique de l'exploitation et de l'analyse des données non structurées est à l'origine de l'essor du Big Data. La nature non structurée de certaines données font qu'elles ne peuvent pas être prises en compte par les bases relationnelles classiques. En marketing, les technologies Big Data permettent par exemple d'analyser les données sonores issues d'enregistrements téléphoniques en conjonction avec des données de type plus traditionnel tels que les historiques d'achat.

Data non-structurée

Est produite par toutes sortes d'acteurs

- Entreprise / Marques
- Consommateurs
- Prospects
- Internauts
- Salariés

Existe sous toutes sortes de formes

- Phrases longues et complexes dans une langue soutenue
- Phrases courtes avec des fautes d'orthographe et/ou de grammaire
- Suite désordonnée de concepts énoncés les uns à la suite des autres
 - en français
 - en anglais
 - en italien
 - en russe
 - en mandarin...

Correspond à toutes sortes de supports

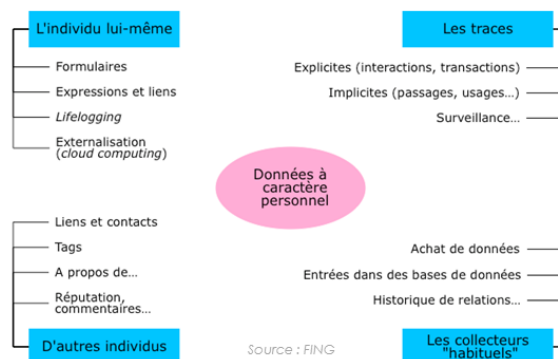
- Pages de sites internet / extranet
- Articles de blog
- Posts et réponses sur les réseaux sociaux (Facebook, Twitter...)
- E-newsletter / Webzine
- Communiqués de presse
- Livres blancs / brochures produits / brochures savoir-faire
- Documents numérisés online ou offline
- Verbatims d'enquêtes et d'études
- Réclamations recueillies par une Marque
- Réponses ouvertes à des questionnaires
- Données textuelles du CRM
- Plateforme collaborative interne
- Contenus online et offline

6. Données personnelles

Définition

Article 2 de la Loi CNIL : « toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne ».

Les sources de données à caractère personnel



Sous-définition 1 La notion de traitement de données à caractère personnel

CNIL : « toute **opération** ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé, et notamment la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, ainsi que le verrouillage, l'effacement ou la destruction. »

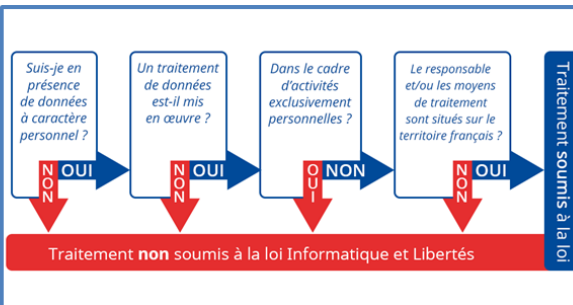
Sous-définition 2 Personne concernée par un traitement de données à caractère personnel

Article 2 de la Loi CNIL : « personne à laquelle se rapportent les données qui font l'objet du traitement. ».

7. Responsable de traitement

Définition

Article 3 de la Loi CNIL : « *sauf désignation expresse par les dispositions législatives ou réglementaires relatives à ce traitement, la personne, l'autorité publique, le service ou l'organisme qui détermine ses finalités et ses moyens* ».



Définition positive de la notion :

Le principal critère pour identifier le responsable de traitement est sa capacité juridique et organisationnelle ainsi que son autonomie à définir les finalités et les moyens du traitement. Le responsable de traitement met en œuvre le traitement dans son intérêt, en son propre nom et pour son propre compte.

Définition négative de la notion :

Le responsable de traitement n'est pas:

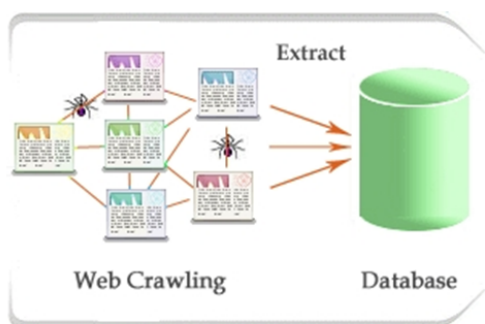
- la personne concernée par le traitement ou le destinataire du traitement;
- Un sous-traitant, qui, au terme de l'article 35 de la loi CNIL est défini comme « *toute personne traitant des données à caractère personnel pour le compte du responsable des traitements* ». Etant précisé que lorsqu'un traitement est sous-traité à un prestataire externe, le dit prestataire ne devient pas le responsable du traitement, lequel conserve toute la responsabilité aux yeux de la loi.
- Un « tiers autorisé » : qui, au terme de l'article 3 II de la loi susvisée est une « *autorité légalement habilitée dans le cadre d'une mission particulière ou de l'exercice d'un droit de communication à demander au responsable du traitement de lui communiquer des données à caractère personnel* ». Pour être qualifié de « tiers autorisé », il faut être habilité expressément par un texte. Tel est le cas de la CNIL légalement habilitée par l'article 44 de la loi CNIL, à accéder aux contenus des traitements de données à caractère personnel.

8. Crawling et scraping

Définition

- **Crawling et Scraping sont deux techniques différentes d'acquisition automatique des données sur le Web.** Ces deux techniques peuvent être utilisées de manière complémentaire
- Un Crawler est un programme informatique dont la fonction consiste à se "promener" de site en site et d'extraire toute l'information présente sur les pages : automatiquement et sans connaissance préalable des sites visités. À partir d'une liste initiale de sites web, il parcourt chaque page de chacun de ces sites pour ensuite suivre les liens qui pointent vers des sites web qui n'étaient pas dans la liste initiale
- Un Scraper est un programme d'extraction du contenu d'un site Web dans le but de le transformer en données structurées pour permettre son utilisation dans un autre contexte. Contrairement au Crawling, le Scraping se prépare en amont : les sites et leurs structures doivent être analysés avant de lancer le recueil des données.

Exemple 1 : Crawling



Exemple 1 : Scraping

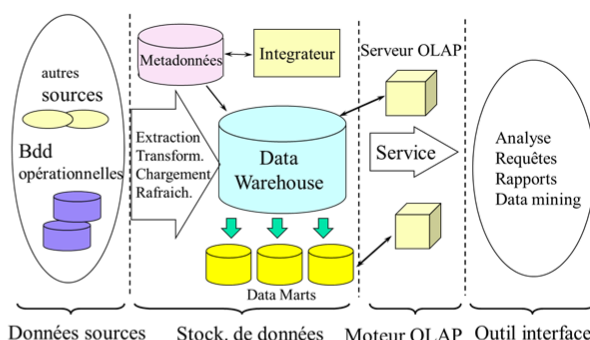
Titre	Auteur	Date	Thème
Ingénieurs, « tous les clignotants sont au vert »	Benoit Floch	10.12.2013	Éducation
Yahoo!, l'un des rares survivants des débuts du Web, fête ses 20 ans	Sarah Bellouezanne	03.01.2014	Technologies

9. Entrepôt de données

Définition

Un entrepôt de données (ou datawarehouse) est un serveur informatique dans lequel est centralisé un volume important de données consolidées à partir des différentes sources de renseignements d'un organisme (bases de données opérationnelles par ex). L'organisation des données est conçue pour que les personnes intéressées aient accès rapidement et sous forme synthétique à l'information stratégique dont elles ont besoin pour la prise de décision.

Illustration :



Exemple 1 : APIDAE

La plateforme SITRA (Système d'information touristique en Rhône-Alpes), rebaptisée Apidae fin 2015, est une base d'informations touristiques en ligne utilisée par les 800 membres du réseau et qui alimente plus de 900 projets numériques. Parmi ceux-ci, on compte de nombreuses applications mobiles, ainsi que des programmes d'animation et des sites diffuseurs de l'offre touristique. Cet entrepôt de données entre dans une logique d'ouverture et de mutualisation de l'information.



Exemple 2 : EDREDON

Le projet EDREDON (Entrepôt de données routières et déplacements engrangés pour leur diffusion ouverte nationale) fait partie des actions retenues en 2014 par le CEREMA. Objectifs : Faciliter la publication et la réutilisation des données de trafic routier open data en temps différé



10. Urbanisation des données

Définition

Travail progressif pour faire évoluer la gestion des données de référence vers des référentiels de données maîtrisés, gouvernés, flexibles, transverses, accessibles et utilisés par l'ensemble des traitements informatiques.

Objectif : accroître la réactivité du SI face aux évolutions

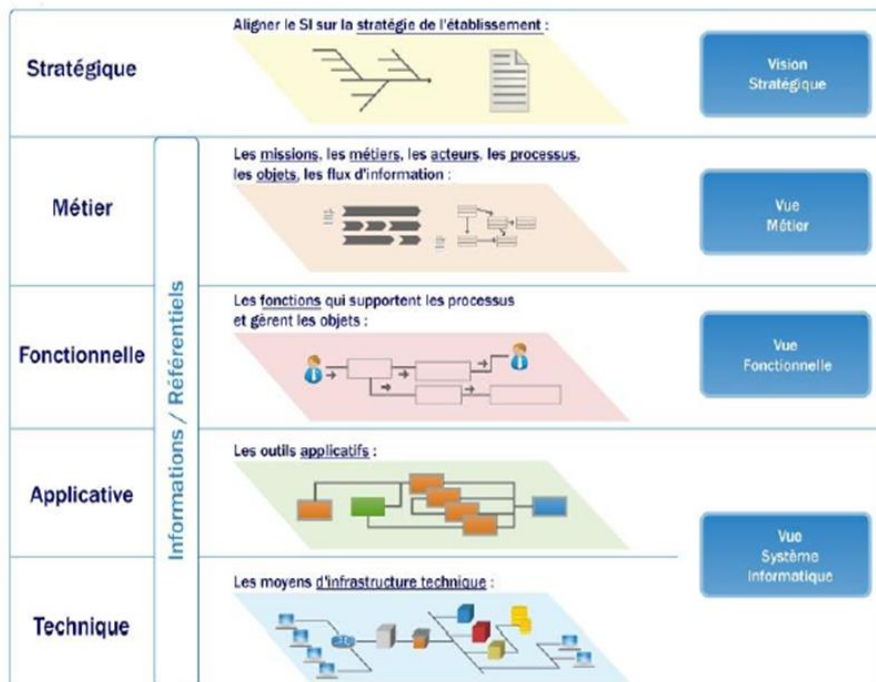
Enjeux :

- Un des axes majeurs de l'optimisation des SI
- Gouvernance de la donnée

Etapes :

- Identifier et isoler fonctionnellement les données transverses
- Impliquer les acteurs métiers à la définition sémantique des données de référence
- Evaluer le patrimoine actuel en matière de référentiels : niveau de satisfaction des utilisateurs, maîtrise des processus d'alimentation et d'actualisation, niveaux d'engagement des acteurs...
- Proposer une cible de rationalisation, de simplification et d'optimisation métier, fonctionnelle et applicative
- Définir une trajectoire de transformation vers la cible

Les six vues du cadre de référence d'urbanisation



11. API

Définition

Applications permettant **d'interfacer des logiciels** entre eux en s'appuyant sur des normes, méthodes et fonctions prédéfinies. L'objectif est de **faire communiquer automatiquement des systèmes entre eux**.

Exemple 1

L'API de Google Maps,
L'application de cartographie de Google, permet à Yelp d'afficher les restaurants critiqués par ses utilisateurs sur la carte de Google.

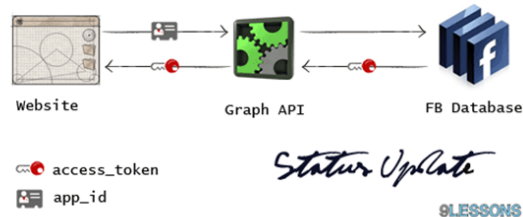
Illustration



Exemple 2

Facebook Graph API
qui permet d'accéder et d'interagir avec les informations Facebook de l'utilisateur

Facebook Graph API Connect



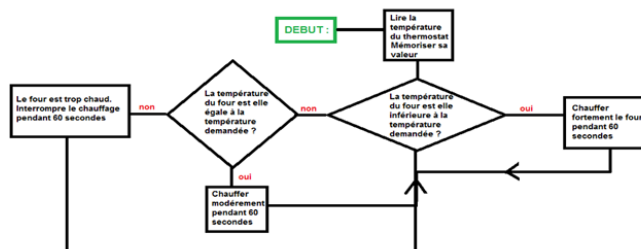
12. Algorithme

Définition

- Un algorithme désigne initialement la **suite de calculs nécessaires pour effectuer une opération complexe**
- Un algorithme informatique est la description dans un langage formel (un langage de programmation) d'une **suite finie et ordonnée de processus qui, à partir de données en entrée, livre des données en sortie en un temps fini, en vue d'un objectif prédéterminé.**

Illustration :

Comment un four se maintient à la bonne T°



Exemple 1 : algorithme de recommandation

Conseiller un livre en fonction des choix précédents de l'utilisateur



3 types de calculs qui correspondent à 3 questions :

- **Similarité** : Quels sont les ouvrages qui abordent un genre similaire à l'ouvrage choisi ?
- **Complémentarité** : Quels sont les ouvrages qui complètent l'ouvrage choisi ?
- **Diversité** : au sein de ce genre, quels sont les ouvrages les plus éloignés de l'ouvrage choisi ?

Puis pondération des réponses par les informations disponibles sur l'utilisateur (âge, localisation, notations d'autres ouvrages).

Exemple 2 : algorithme de prédiction

Prédire les résultats d'une élection : 2 étapes :

- **Identifier les variables, rassembler les données pertinentes, les nettoyer** (INSEE, élections passées, notoriété du candidat...)
- **Pondérer les données** (en ajustant le modèle en comparant ces résultats avec les résultats des élections passées)

13. Crowdsourcing

Définition

Méthode de **production collective de données** qui sont **générées à l'initiative des usagers** souhaitant contribuer à l'efficacité d'un système ou d'une application, de manière anonyme ou non.



Exemple 1

Crowdsourcing actif (l'utilisateur contribue activement)

Dans la mobilité, cette technique est utilisée pour les applications **Moovit**, **Coyote** ou **Waze**.

Le contributeur crée une information nouvelle par simple sélection (accident, travaux...) ou enrichit celles des d'autres.



Exemple 2

Crowdsourcing passif (l'utilisateur a autorisé un objet connecté à envoyer des données en temps réel. Par la suite, il n'a pas à envoyer de push de données).

Waze utilise la géolocalisation des téléphones des utilisateurs pour en déduire une affluence.

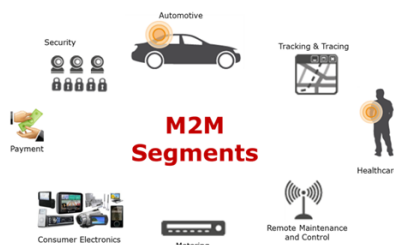
Netatmo utilise les stations météo connectées de ses utilisateurs pour fournir une météo ultra-locale en temps réel.

M2M (Machine to machine)

Définition

Lorsque des appareils connectés échangent des informations sans intervention humaine, on parle de M2M (communication intermachines). Les innombrables données collectées par les appareils connectés sont les M2M data. Cette technologie entre dans le champ de la télématique.

Illustration :



Exemple 1 : les smart cars

Si certaines voitures sont aujourd'hui des capteurs de précision, elles vont devenir des objets communicants et interagir avec les autres véhicules, avec les signaux routiers ou même avec la route. La reconnaissance d'objets et de signaux permet d'optimiser la vitesse et la distance de freinage des véhicules.



Exemple 2 : l'éclairage intelligent

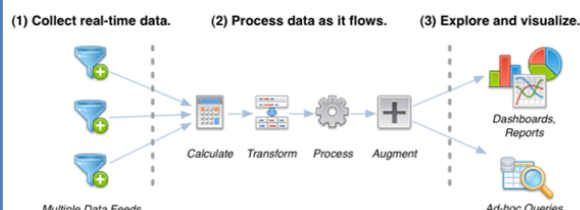
Les lampadaires constituent un réseau bien maillé d'éclairage public qui pourrait se transformer en support multi-usages pour peu qu'on les dote d'une connectivité M2M : ainsi, leur allumage peut être déterminé par la détection d'un signal bluetooth, ou bien ils peuvent communiquer des offres promotionnelles pour les commerces voisins aux passants par le wifi ou la lifi.

14. Suivi en temps réel

Définition

Un système en temps réel est capable de contrôler ou de piloter une évolution de données à la vitesse même où celles-ci sont produites. Il ouvre l'opportunité du monitoring, à savoir le suivi de ces données évolutives. Dans le cas du Big Data, le suivi en temps réel nécessite de fortes capacités de traitement.

Illustration :



Exemple 1 : le monitoring urbain de Nice

L'objectif poursuivi est de superviser, avec un outil commun, plusieurs paramètres environnementaux de la ville et leurs évolutions dans le temps et l'espace. Des capteurs sur candélabres mesurant différents paramètres liés à l'environnement urbain (bruit, température, qualité de l'air, odeurs, comptages de fluides, etc.) ont été mis en place, permettant le suivi en temps réel de ces paramètres. Ce système permet de développer un système de surveillance et un observatoire de l'écologie.



Exemple 2 : l'analyste Pulsar d'Ebay

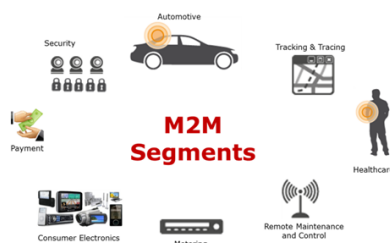
Pulsar peut traiter « des millions d'événements par seconde » pour créer des flux d'information personnalisés en fonction des sessions d'achats des internautes connectés. Le marketing est ainsi adapté à la seconde près aux besoins exprimés par les recherches du client.

15. Blockchain

Définition

Lorsque des appareils connectés échangent des informations sans intervention humaine, on parle de M2M (communication intermachines). Les innombrables données collectées par les appareils connectés sont les M2M data. Cette technologie entre dans le champ de la télématique.

Illustration :



Exemple 1 : les smart cars

Si certaines voitures sont aujourd'hui des capteurs de précision, elles vont devenir des objets communicants et interagir avec les autres véhicules, avec les signaux routiers ou même avec la route. La reconnaissance d'objets et de signaux permet d'optimiser la vitesse et la distance de freinage des véhicules.



Exemple 2 : l'éclairage intelligent

Les lampadaires constituent un réseau bien maillé d'éclairage public qui pourrait se transformer en support multi-usages pour peu qu'on les dote d'une connectivité M2M : ainsi, leur allumage peut être déterminé par la détection d'un signal bluetooth, ou bien ils peuvent communiquer des offres promotionnelles pour les commerces voisins aux passants par le wifi ou la lifi.

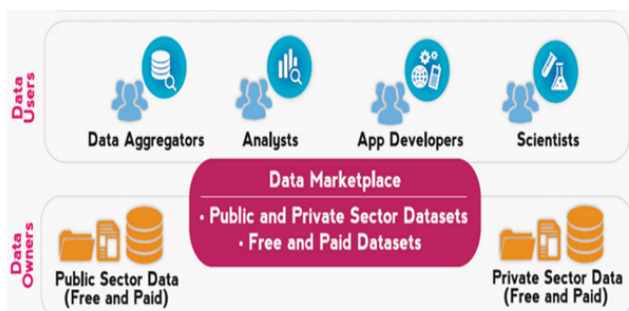
16. Place de marché de données

Définition

Une place de marché de données est une plateforme logicielle permettant la mise en relation de différents acteurs pour le transfert et l'échange de données.

Les places de marché de données organisent l'extraction et la valorisation de données issues de sources hétérogènes ainsi que le système de partage et d'échange entre les producteurs et les ré-utilisateurs.

Illustration :



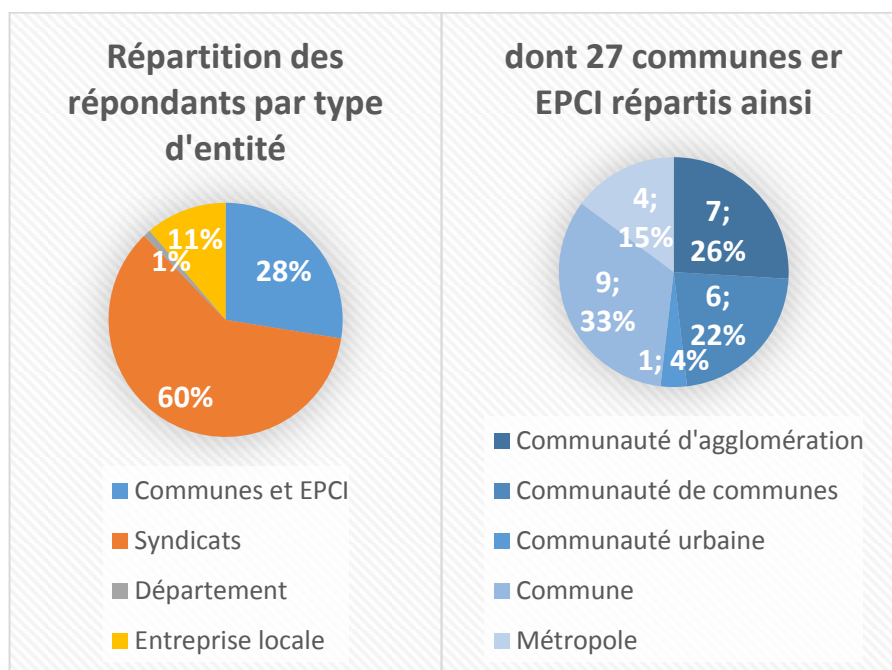
Exemple : OpenDataSoft

OpenDataSoft propose une plateforme permettant aux acteurs publics et aux entreprises de valoriser et de partager leurs données. La plateforme propose des solutions clés-en-main de réponse aux problématiques de publication, ouverture, et ré-utilisation de données de différents types. Celle-ci permet notamment de transformer automatiquement des données en API. L'intégration du service de paiement Stripe permet également aux membres du réseau de monétiser l'accès à leurs données.

B. Résultats de l'enquête flash

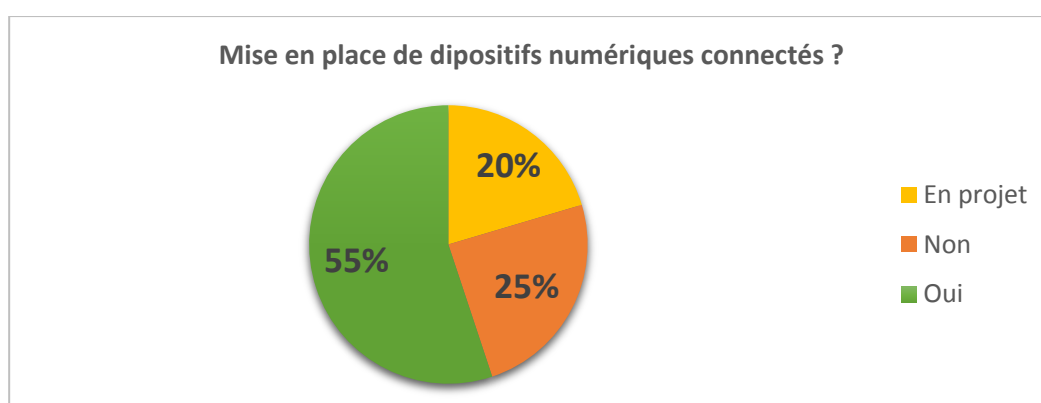
1. Profils des répondants

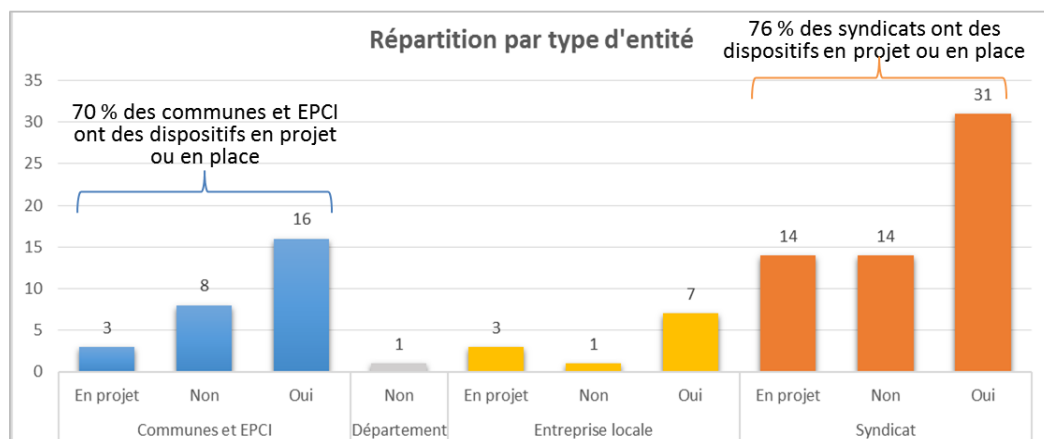
98 réponses ont été enregistrées.



2. Mise en place de dispositifs numériques connectés

3 / 4 des répondants (74) ont mis en place ou ont pour projet des dispositifs numériques connectés.

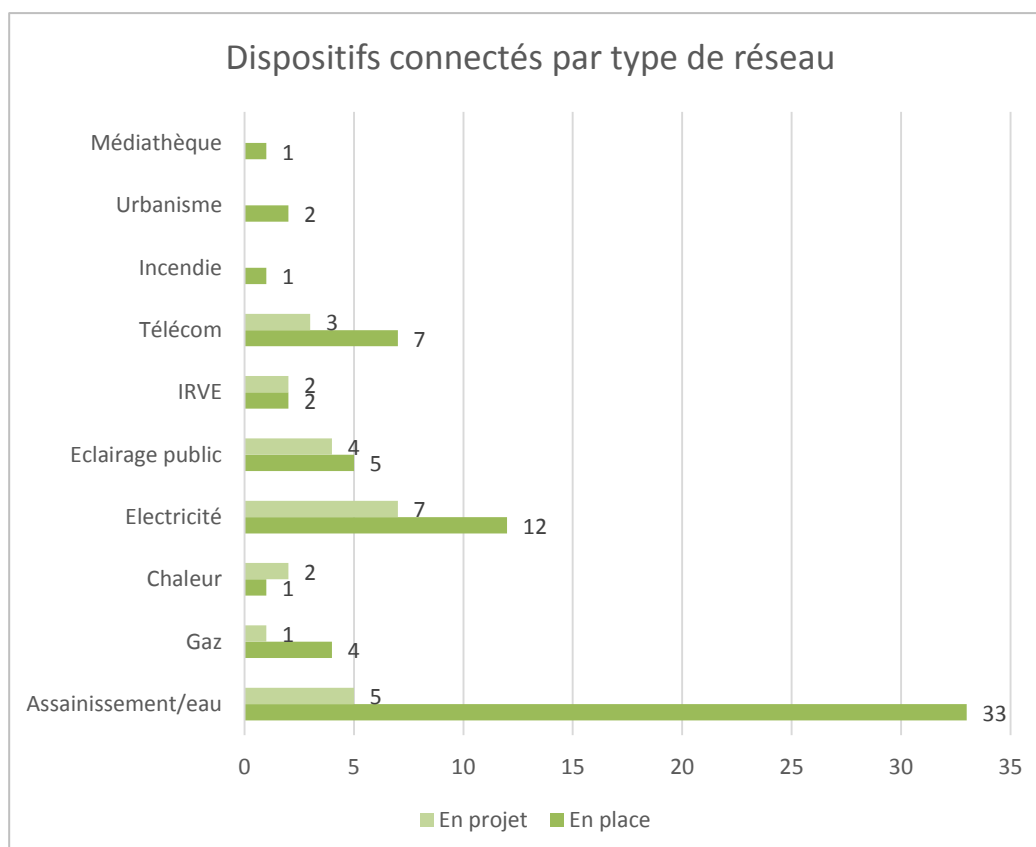




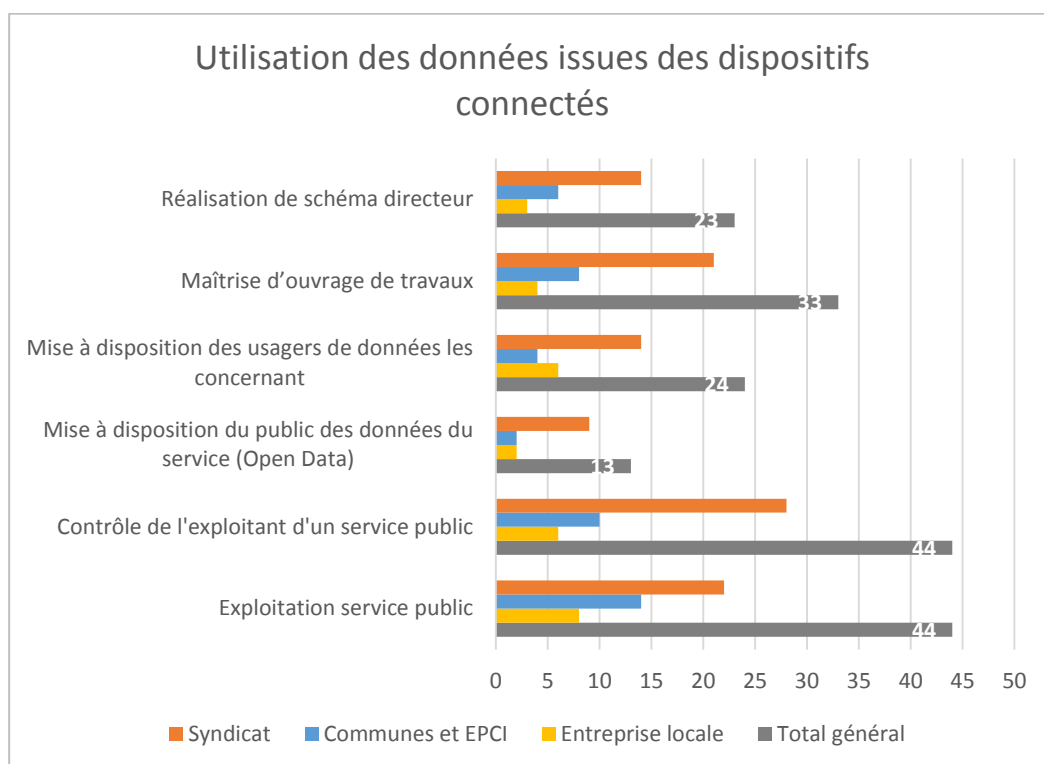
➤ Type de réseaux impactés

Top 3 :

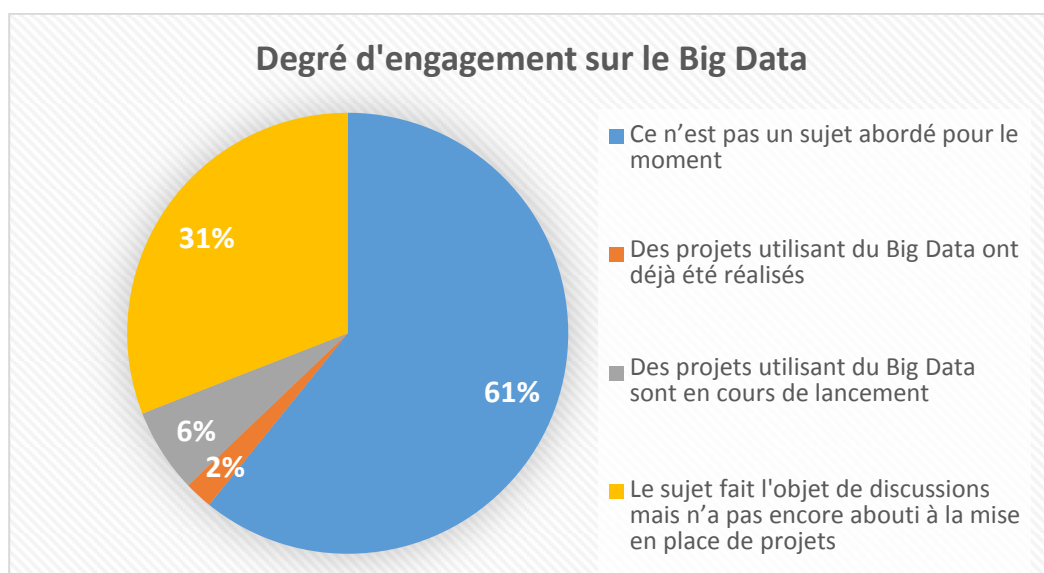
- Assainissement/ Eau
- Electricité
- Télécom

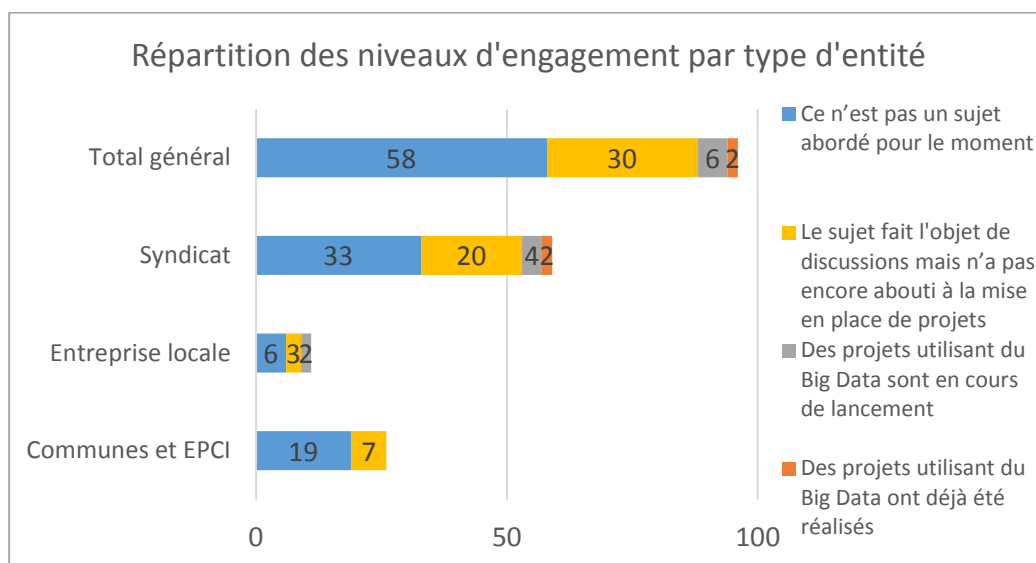


➤ Utilisation des données



3. Degré d'engagement des collectivités sur le big data

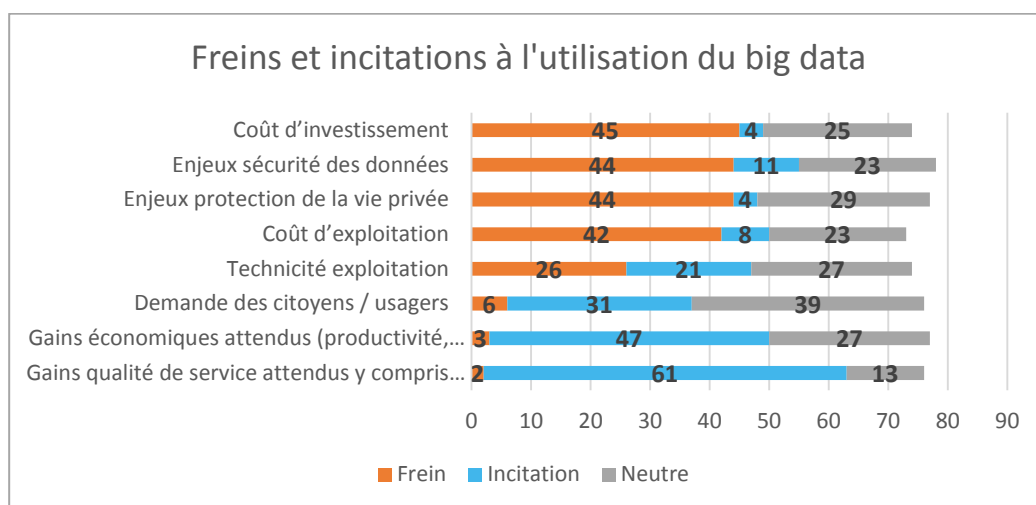




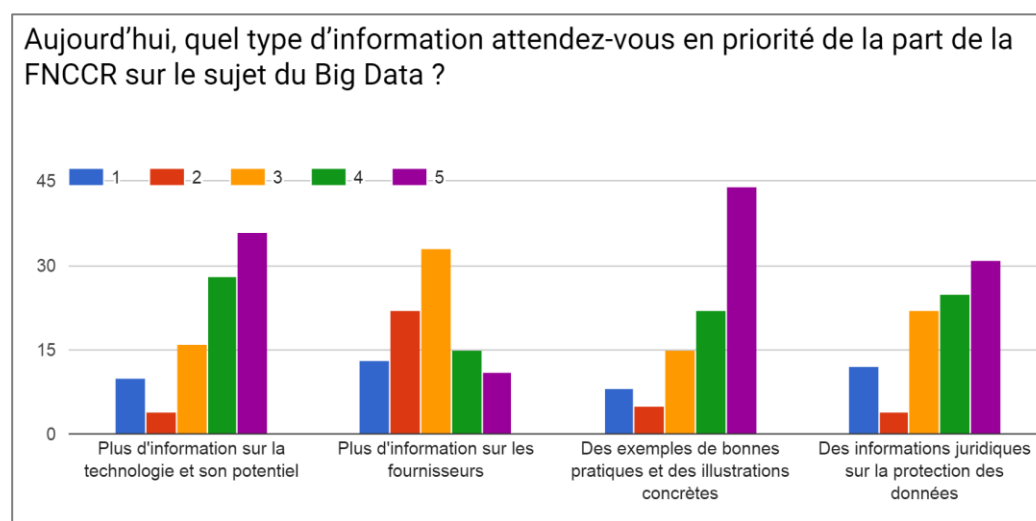
➤ *Domaine impacté*

Domaine impacté	Nombre de projets
Surveillance	8
Optimisation gestion	4
Information sur la consommation	3
Optimisation énergétique	2
Information sur la production	1
Information des usagers	1
Réglementaire	1
SIG	1
Tableau de bord	1
Analyse prédictive	1
Planification	1
Autonomie des systèmes	1
Gouvernance	1
TOTAL	25

➤ Freins et incitations à l'utilisation du big data



4. Attentes des collectivités



Autres :

- Sécurisation des interprétations et de l'utilisation des données (X5)
- Accès aux données dans les cas des conventions de concession (DSP) notamment (x3)
- Modèle économique et partage de la valeur (x3)
- Harmonisation des pratiques pour faciliter les échanges et les compilations de données (X3)
- Gouvernance de la données (x2)
- Accompagnement des petites structures (x2)
- Exemples et partages d'expérience (autres pays européens, structures publiques identiques) (x2)

C. Fiches collectivités

1. Bordeaux Métropole

Perspectives et objectifs sur la donnée

La stratégie numérique de la ville est marquée par une volonté forte de s'affranchir de la dépendance aux éditeurs logiciels et aux délégataires de services publics sur le plan de la propriété des données, en stockant les données du parc applicatif dans un entrepôt après avoir harmonisé les différents modèles pour en faciliter la réutilisation

Modification du SI et état d'avancement

- Le principal chantier qui vient d'être initialisé est celui de la mutualisation et de l'urbanisation du système d'information (ville centre et autres communes), au sein du système d'information de la métropole.
- Une nouvelle plateforme open data est en déploiement (2016).
- Un projet d'entrepôt de données est en gestation (déploiement en 2017), il intégrera la plateforme d'open data.
- La réflexion porte aujourd'hui sur la question de la réutilisation des données, avec la mise en place d'un « espace public » intégrant les informations géolocalisées des différents SI de la métropole. La définition des niveaux d'accessibilité est fonction de la sensibilité des données avec 3 niveaux de diffusion :
 - Un bloc open data en usage libre
 - Des hangars de données, ouverts sur "pass" aux fonctionnaires territoriaux du domaine producteur
 - Un accès ponctuel et réglementé accordé aux partenaires pour prélever de la donnée (associations, étudiants, industriels) dans le cadre de projets ou de partenariats.
- L'heure est aux **négociations avec les délégataires (DSP) pour basculer les données transmises dans l'entrepôt de la collectivité.**
- La création d'API (connecteurs pour open data, tiers applicatifs) se fait au fil des besoins.
- Le PLU sera distribué en open data sous la norme préconisée par le CNIG en 2019 (il sera également chargé sur le Géoportail de l'urbanisme).

Portage en interne et travail avec les services

- La **DSI de la métropole est mutualisée** entre la ville centre et 8 autres communes. L'organisation de cette nouvelle structure est en cours sur le plan de la fusion des équipes, l'harmonisation des méthodes et des pratiques de gouvernance. Certaines villes ne sont pas prêtes à transférer leurs données sur le portail de la métropole.
- Une **Direction Générale du Numérique** est en charge de l'urbanisation des données.
- Bordeaux n'a pas encore de Chief Data Officer.
- Un échange régulier est engagé avec l'association **OpenData France**. Pourtant, l'open data ne suscite plus l'engouement : il n'aurait pas répondu aux attentes en matière de développement de la filière numérique.

Projets et applications

- Le projet le plus visible est la transposition de la carte de vie quotidienne sur une application mobile « **Bordeaux la ville en poche** », qui utilise 6 SI différents : la mise en place d'un serveur relais a été nécessaire (en attendant la construction de l'entrepôt de données de la métropole).
- La récupération de données est bien structurée à partir de nombreux capteurs sur la mobilité, mais elle est en revanche peu développée sur les réseaux sociaux.

2. Métropole de Lyon

Perspectives et objectifs sur la donnée

La Métropole de Lyon coordonne la politique de la donnée sur son territoire afin :

- d'accélérer le développement économique en exploitant la donnée métropolitaine,
- de favoriser la coopération entre acteurs privés et publics et stimuler l'innovation,
- de permettre une réutilisation gratuite des données y compris à des fins commerciales, avec l'impératif d'une conformité à l'intérêt général.

La politique de la donnée alimente la stratégie « métropole intelligente » et le développement de projets sur le territoire.

Modification du SI et état d'avancement

- La mutualisation des Systèmes d'Information de la Métropole et de la Ville de Lyon n'est pas encore effective
- Développement d'une logique territoriale de « Métropole Plateforme » avec la mise en place d'une Plateforme Data depuis 2013 pour l'ouverture des données de la ville et l'hébergement des données des partenaires (ville de Lyon, ville de Chassieux, partenaires privés et délégataires).
 - Plusieurs niveaux d'ouverture de la plateforme : open data, partage restreint à certains acteurs, espace privatif pour l'exploitation en interne
 - La plateforme inclue des services API autour des données
 - Pistes d'évolution :
 - Convergence avec les données SIG du service « géomatique et données métropolitaines »
 - Recensement du patrimoine de données et gestion d'une version interne de la plateforme permettant de cataloguer les données

Portage en interne et travail avec les services

- Première collectivité à créer un poste d'**Administrateur Général des données** en avril 2015, en charge d'animer la communauté des acteurs de la donnée en interne et en externe
- La Direction Innovation Numérique & Systèmes d'Information (DINSI), au sein de la Délégation du Développement économique, Emploi & Savoirs, comprend une équipe dédiée à la gestion des données : 5 personnes sont en charge de la gestion du backoffice de la plateforme data, 1 stagiaire a été recruté pour développer les data sciences
- Développement d'une culture de la donnée au sein des services et mise en place d'interlocuteurs spécifiques pour la question des données au sein des directions métier (15 personnes)
- Gouvernance des données de référence : constitution d'un service « géomatique et données métropolitaines » (anciennement InfoGéo) travaillant à l'ouverture des données de référence et au partage en interne
- Interactions avec les acteurs du territoire : La Métropole se positionne comme le « porteur du service public de la donnée locale sur son territoire »

- 2 niveaux de coopération avec la Région :
 - ⇒ Volet info géographique avec le géoportail de la Région
 - ⇒ Innovation autour de la donnée, en lien avec Cédric Lambert, chargé d'innovation à la Région : amélioration du moissonnage des données par la plateforme régionale et par la plateforme d'Etalab
- Création du TUBÁ en 2014 : lieu dédié à la réutilisation des données publiques et aux expérimentations urbaines, rassemblant entreprises, start-ups et citoyens

Projets et applications

Trois axes majeurs de développement de projets ont été définis par la métropole :

- Energie :
 - Lancement d'expérimentation dans le domaine des smart grids : projets GreenLys Smart Electric Lyon menés en partenariat avec ERDF concernant la gestion des données « aval compteur » et l'amélioration des comportements des usagers en matière de consommation énergétique. Le territoire a par ailleurs été l'un des premiers territoires d'expérimentation pour le développement des compteurs Linky
 - Gouvernance des données de l'énergie : projet Lyon Living Lab sélectionné dans le cadre de l'appel à projets « démonstrateur industriels pour la ville durable : réflexion sur la création d'un opérateur de données énergétiques au niveau de l'agglomération
- Mobilité : projet **Optimod'Lyon** : travail sur l'agrégation et l'interopérabilité de l'ensemble des données de mobilité du territoire (publiques, privées), intégrant un module de prédiction du trafic à une heure.
- Santé : projet d'Observatoire des données de la santé en partenariat avec UrbaLyon : mise en synergies de différents types de données de santé avec des données environnementales : aménagement, qualité de l'air, données d'accompagnement social, données médicales...

Des projets européens liés aux enjeux d'exploitation des données IoT :

- BloTop (Building an Internet of Things Open Innovation Ecosystem for connected smart objects) : création de standards d'échange pour faciliter la création de nouveaux services autour des objets connectés
- Smarter Together : rénovation des logements et apport des données « multi-énergies » pour l'amélioration de la performance énergétique des logements

3. Martigues et Communauté d'Agglomération du Pays de Martigues

Perspectives et objectifs sur la donnée
- Définition en 2016 d'un plan d'action et d'accompagnement autour d'un certain nombre de projets - Mise en place d'un groupe de travail interservices sur la question de la donnée - Martigues fait partie des trois villes (avec Dignes et Marseille) qui vont expérimenter la plateforme Open PACA si la Région acquière OpenDataSoft
Modification du SI et état d'avancement
Migration de 1 200 postes sur Libre Office (terminée)
Portage en interne et travail avec les services
- La municipalité de Martigues et la communauté d'agglomération (CA) étaient sur des problématiques de gestion de data sans pour autant le conscientiser et le structurer comme telles. La question de la donnée était gérée par la DSI (la seule mutualisée dans les Bouches-du-Rhône) et par le SIG qui n'est pas rattaché à la DSI. - Le SIG de la CA et a été intégré à la métropole Aix-Marseille-Provence depuis le 1 janvier 2016. - La DSI est passée des « systèmes informatiques » aux « systèmes d'information » ce qui implique la question de la donnée
Projets et applications
- Constitution de l'entrepôt de donnée (en cours) - Utilisation des solutions de progiciel de gestion (IBM), par exemple, sur les finances et les RH - Le SIG utilise la solution ArcGIS - Des ateliers pratiques de médiation sur l'open data inspirés par le dispositif Infolab de la FING pour apprendre aux citoyens à utiliser l'open data - Travail avec des agents de la collectivité pour développer la culture de la data - Open data (le plus gros producteur de jeux de données sur Open PACA) - Proposition d'organiser une formation sur l'open data pour les responsables de conseils de quartier - Cartopartie de l'OpenStreetMap avec Arsenic pour qualifier le territoire - Ateliers Wikipédia (de production) par 4 ERIC - Organisation de la 1ère semaine de l'Internet citoyen à Martigues : 3-5 ateliers de type infolab de médiation à la donnée, et à l'open data produite par la ville pour l'apprentissage des données et la valorisation des données.

4. Nantes Métropole

Perspectives et objectifs sur la donnée
Changement de paradigme : il s'agit de passer du simple open data à une

gouvernance collaborative de la donnée, pour reconstruire le système d'information de manière structurée et dé-silotée, dans une logique de données.

Modification du SI et état d'avancement

- Le **logiciel de gestion du patrimoine est mutualisé** entre les communes, sur la base d'une convention définissant le propriétaire de la donnée et les droits des utilisateurs.
- Les données sont stockées dans un seul **entrepôt de données** (Nantes Métropole).
- Les services numériques se livrent à des essais de **croisement de données** (publiques et privées), pas forcément de manière ouverte, mais dans une stratégie juridique définie au cas par cas.
- La réflexion porte aujourd'hui sur la gouvernance de la donnée produite par les opérateurs privés en conventions de concession : l'objectif est d'externaliser la prestation mais de **retenir les données produites par le délégataire** dans le domaine informatique de la métropole. Il s'agit d'imposer l'outil de gestion et de formatage des données aux opérateurs en DSP (ce qui est déjà réalisé pour l'éclairage).

Portage en interne et travail avec les services

- La Direction Ressources Numériques est en charge de la cohérence des SI pour tout ce qui relève des compétences de Nantes Métropole.
- Il est prévu de créer prochainement **un service d'urbanisation du SI**, après 1,5 an de discussion et des débats sur l'opportunité de cette fonction. Les porteurs de ce projet ont l'intention de proposer rapidement des services visibles (notamment un Chief Data Officer).
- Il existe une bonne collaboration entre services (en particulier ceux qui produisent des données) mais qui ne fait pas encore l'objet d'une programmation systématique et structurée
- Des référents numériques sont identifiés au sein de chaque département

Projets et applications

- En 2014, la direction en charge de la Stratégie Numérique du Territoire (Nantes Métropole) a lancé un Plan d'action Innovation et Numérique. Le volet open data (intégré à ce plan d'action) semble être de moins en moins convaincant pour les élus et les entrepreneurs, d'où la volonté de développer un service de données plus ambitieux et inclusif.
- La mise en place d'un référentiel de l'immobilier (à l'échelle des bâtiments) a suscité, dans le cadre de ce plan d'action, l'agrégation de 48 fiches services.

5. Métropole Nice-Côte d'Azur

Perspectives et objectifs sur la donnée
• Constitution d'un entrepôt de données métropolitain pour y stocker à termes l'intégralité des données et en faire un point d'entrée unique • La vocation de l'entrepôt est de mettre à jour les données des services et de les diffuser ensuite sur des portails « open data »
Modification du SI et état d'avancement
• La Métropole et la ville centre Nice ont mutualisé toutes les fonctions support y compris l'informatique. La DSI mutualisée assure l'informatique pour la Métropole, la ville de Nice, le CCAS, et aussi pour les établissements publics qui en dépendent, par exemple, la régie de l'eau. • Mise en place de protocoles sur les données issues de capteurs réutilisables par d'autres • Récupération des données via des API Utilisation d'un bus d'échange de données pour modifier le format des données pour d'autres usages (Produit Talend)
Portage en interne et travail avec les services
• Pour assurer le pilotage de la stratégie ville intelligente, un centre d'excellence smart city a été constitué. Il est piloté par un élu et un responsable administratif et s'appuie sur deux organes de gouvernances (un comité de pilotage et un comité technique). C'est dans ce cadre qu'est décidée la stratégie « Smart Métropolis ». • En ce qui concerne l'open data : les directions ont été formées sur la base du volontariat et progressivement elles sont devenues gestionnaires de données, qu'elles mettent à jour • Une convention de partenariat est signée avec certaines communes de la métropole pour mettre à disposition leurs données sur le portail
Projets et applications
• Monitoring urbain sur le volet environnemental • Monitoring prédictif sur les risques de crues (capteurs vidéo installés sur les rivières) • Récupération de données de concessionnaires : travail avec l'aéroport de Nice pour obtenir des données en temps réel des flux entrants et sortants notamment • Echange de données avec l'iMREDD (Institut Méditerranéen du Risque, de l'Environnement et du Développement Durable) pour les partager avec des partenaires dans une perspective de recherche • Bonnes pratiques sur les bâtiments connectés, comme le bâtiment intelligent de Schneider • « Smart data » (bonnes pratiques de remontée, de qualification, d'accompagnement pour faire de la donnée de qualité, pour pouvoir l'automatiser sur de gros volumes)

6. Paris

Perspectives et objectifs sur la donnée

Une politique de la donnée structurée autour de 5 piliers structurants



Modification du SI et état d'avancement

- **Open data** : la ville de Paris a été une des premières à utiliser les outils d'OpenDataSoft, incluant des processus automatiques de mise à disposition des données sur la plateforme, notamment en provenance du SIG
- **Mutualisation SIG** : fusion de la fonction géomatique au sein d'un grand service mutualisé, incluant une réflexion sur la création d'un modèle de données 3D sur le territoire parisien
- Réflexion sur la création d'une **plateforme territoriale de données** partagées, hébergeant les données de la ville et de ses partenaires, notamment les données des concessionnaires ainsi que les données récoltées dans le cadre des **projets IoT** menés par la ville
- Réflexion engagée par la ville sur l'interconnexion des SI et la conception de modèles de gouvernance de la donnée avec les opérateurs de la ville, notamment dans le domaine énergétique

Portage en interne et travail avec les services

- Création d'un poste d'**Administrateur Général des Données** en 2014, dans la lignée des travaux engagés par la Maire de Paris et son adjoint, Jean Louis Missika, pour la constitution d'une démarche « Ville intelligente et durable »
 - Missions de l'Administrateur Général des Données :
 - **Coordination d'une démarche « data »** à l'échelle de la ville
 - **Acculturation des services** de la ville et prise de conscience du rôle de la donnée au sein des différents métiers
 - Directions particulièrement impliquées dans la conduite des projets smart city : Direction de la Voirie et des déplacements, DSI, Direction du Patrimoine, Direction de la Jeunesse et des Sports
- Organisation des services** : présence d'ETP « de coordination » travaillant en appui des directions stratégiques

Projets et applications

Les domaines prioritaires d'action de la ville sont : la mobilité, l'environnement, la

gestion des déchets, l'énergie, la gestion des bâtiments et l'espace public

Exemples d'expérimentations d'envergure menées par la Ville de Paris :

- **Analyse des flux** de déplacement en partenariat avec la start-up Placemeter et Cisco sur la place de la Nation
- **Opération « réinventer les places »**, menée par la Direction de la Voirie et des Déplacements concernant l'analyse des usages de l'espace public et la mise en place de mobilier urbain modulable
- **Réalisation de scénarios de fonctionnement des feux de croisement** en fonction de la densité piétonne autour des carrefours
- **Datasciences** : le concours « DataCity », conduit en partenariat avec le NUMA est ouvert aux startups afin d'apporter des solutions concrètes aux défis urbains dans les domaines de l'énergie, de la mobilité et l'environnement
- **Projet IoT** : création d'une plateforme de données partagées afin de mieux équilibrer les consommations et les productions dans la gestion des pics de consommation à l'échelle d'un îlot ou d'un quartier

7. Rennes Métropole

Perspectives et objectifs sur la donnée

- En tant que ville pionnière de l'open data, Rennes oriente sa stratégie smart city sur des projets à forte dimension participative et privilégie la transparence de l'action publique. Une nouvelle gouvernance de la donnée doit être mise en œuvre autour de l'entrepôt mis en place par la métropole.

Modification du SI et état d'avancement

- Le SI de la métropole comporte **150 jeux de données différents** (très complets pour les SIG, les transports, la culture, l'état civil et le budget, beaucoup moins pour le social, les données économiques, les réseaux en régie et l'environnement), dont les licences d'utilisation et les formats standardisés sont optimisées, du fait d'une « tradition » locale d'open data (notamment sur le volet des transports).
- La réflexion porte aujourd'hui sur le **croisement des données collectées en interne avec les données externes**.
- La plateforme open data de Rennes Métropole (qui concerne 42 communes) va basculer sur OpenDataSoft.
- En ce qui concerne le big Data, l'enjeu est de développer en interne un niveau de compétence suffisant pour internaliser certaines tâches et externaliser la partie data science, sans contrainte sur la souveraineté des données.
- Une **clause open data** est systématiquement insérée dans les marchés publics ou DSP : **la donnée est par défaut diffusée** (à la charge du prestataire).
- Une réflexion est en cours sur les **infrastructures** dont la collectivité doit se doter pour assurer ses missions de service public de la donnée.

Portage en interne et travail avec les services

- Le Service Innovation Numérique est mutualisé entre la ville de Rennes et Rennes Métropole (rattaché au pôle Communication), il compte 4 personnes.
- La collaboration avec les chefs de projet des autres services se fait en particulier avec la Direction déléguée à l'Innovation et à la Smart City, avec le Responsable Aménagement et Usages du Numérique, la DSI et le service SIG.
- Rapport interne en cours d'élaboration sur le montage d'une « régie locale de données » abordant :
 - La dimension technique : stockage, normalisation
 - La dimension politique : souveraineté et partage de la donnée et tiers de confiance
 - La dimension organisationnelle : dé-silotage des services, évolution corollaire du SI.

Projets et applications

- Une **carte multiservices** (sur la base de Korrigo, la carte de transports régionaux), réunissant les services transport et des services de la ville, est en cours de déploiement. Conformément aux recommandations de la CNIL, 10 identifiants sont prévus pour chacun des champs d'application qu'elle pourrait connaître : les 9 préconisés par la CNIL et 1 « champ libre ».
- La ville déploie un **démonstrateur industriel pour la ville durable** basé sur une carte interactive (solution Dassault Systemes). L'enjeu est de greffer sur cette plateforme du BIM et des applications diverses : le Big Cycle Data (suivi des vélos en libre services), une cartographie énergétique des bâtiments, ou encore des données environnementales en temps réel - provenant de capteurs ou du crowdsourcing - dans le cadre d'un projet Open Energie Data.
- La ville a œuvré pour **l'ouverture données budgétaires** assortie de datavisualisations, notamment avec l'application « Qui Sommes-nous » basée sur des données de l'INSEE, qui dresse le portrait statistique des habitants du territoire

D. Fiches entreprises

Data publica



Présentation

Entreprise française spécialisée dans « l'information sur les entreprises » et le marketing prédictif, au départ spécialisée dans le sourcing de données (recueil automatisé de données par des robots essentiellement sur le web).

Offre

- Produit principal : une base de données entreprises (4,5 millions) reposant sur l'aspiration de 600 000 sites web des entreprises en France et sur la récupération des données de base (Infogreffe, DILA, INSEE, INPI)
- Permet de classer les entreprises et d'identifier des écosystèmes (Technologies vertes, Fintech, Start-up...)
- Les technos : recueil automatique et aspiration de données à partir du web (Crawling) et agrégation de données non structurées et de données structurées à partir de sources nationales. Algorithmes de type machine learning
- Clients publics : conseil départemental (évolution de l'écosystème des Fintech) identification de 10 écosystèmes d'entreprises pour la CCI de Paris-IDF

OpenDataSoft



Présentation

Start-up spécialisée dans le traitement et le partage des données, notamment en réalisant des portails open data pour le compte de collectivités et des acteurs privés.

Offre

- Une trentaine collectivités clientes pour l'open data et déjà quelques grands groupes sur le secteur privé (La Poste, SNCF, ERDF)
- Valeur ajoutée : une solution en mode SaaS (cloud) permettant d'agglomérer (via des API) et de rendre compatibles des données venant de canaux très différents
- Entrepôt : pour pilotage interne, surveillance de la production des délégataires ; API qui permet d'agglomérer des données venant de canaux différents et de rendre les jeux de données compatibles ; modèles économiques avec redevance
- Exemple d'usages : la solution intègre des flux de données temps réel (Keolis à Rennes). Certains clients utilisent la plateforme pour effectuer de l'analyse prédictive : savoir où il y aura de la place aux bornes Vélib dans une heure ;



savoir si sa rue est polluée ou non en temps réel ; savoir dans quelle rame monter (Tranquilien).

m2ocity

Présentation

m2ocity a été créée en 2011 par Orange et Veolia EAU et exerce aujourd'hui en tant que société indépendante son activité d'opérateur de télé-relève. Son expertise acquise sur la gestion des réseaux d'eau lui a permis d'élargir son action à d'autres secteurs de la ville durable (ex : bâtiments connectés, stationnement intelligent, pollution...).

Offre

- m2ocity construit des solutions sur mesure qui s'arrêtent soit à la livraison de la donnée soit jusqu'à fournir le logiciel pour exploiter les données :
 - M2ocity fournit à la petite ville de Saint-Amand une solution pour dissuader le stationnement longue durée grâce à l'équipement en capteurs d'une vingtaine de places de parking et un logiciel dédié permettant d'avoir un aperçu
 - M2ocity a mis en place un système de détections des fuites pour la ville de Perpignan
- m2ocity travaille quasi exclusivement sur des réseaux bas débit et dispose aujourd'hui de 1,7 millions de compteurs et 1,8 millions d'objets connectés tous secteurs confondus
- Les données sont analysées, traduites, décompilées pour une livraison efficace

SUEZ Smart solutions (ex Ondeo System)



Présentation

Née en 1986, Ondeo System développe des solutions avancées dans l'ingénierie des systèmes d'information et propose des technologies innovantes pour la gestion intelligente des réseaux.

Offre

Le groupe participe activement à une expérimentation à Saint-Etienne sur une plateforme de gestion des données en prestation. La volonté du groupe est de se positionner avec **une offre complète pour la gestion des données auprès des collectivités territoriales** :

- Cette offre existe déjà en partie à travers un extranet ouvert aux collectivités lorsqu'elles sont en contrat avec SUEZ (ex : toutes les données relatives à l'eau sont fournies qu'il s'agisse de qualité, des rejets, des interventions, etc.)
- Cet outil permet également la dématérialisation des archives pour les migrations lors des changements de prestataires

- Cette plateforme peut directement ouvrir les données ou bien les fournir aux collectivités pour qu'elles les mettent elles-mêmes en open data
- Cette offre de plateforme favorise une mutualisation des moyens technologiques pour les collectivités en contrats de services. C'est Numergie, dont la Caisse des Dépôts est actionnaire, qui héberge les serveurs en France.

Placemeter



Présentation

« Open urban intelligence platform » **spécialisée dans l'analyse de mouvement et de flux de déplacements (véhicules, vélos, piétons...) à partir de contenus vidéo.**

Offre

- Transformation des contenus issus de vidéos en données de trafic en temps réel (comptage, entrées/sorties...) par des algorithmes de reconnaissance de mouvements
- Secteurs d'application très divers : urbanisme, architecture, immobilier, retail...
- Projet mené en collaboration avec la Mairie de Paris : analyse des flux et compréhension des usages urbains en vue d'adapter l'aménagement de la Place de la Nation
- Collaboration avec des entreprises spécialistes du traitement et de l'analyse des données : Cisco, start-ups...

Orange Applications for Business



Présentation

Filiale d'Orange Business Services spécialisée dans le développement de services numériques pour le compte d'entreprises et de collectivités (développement et intégration de solutions de traitement et d'analyse des données, big data, Internet des objets...).

Offre

- Projet big data : Flux Vision analyse les données mobiles pour la mesure de la fréquentation d'une zone et le suivi des déplacements des populations (pour le tourisme, analyse de la fréquentation d'événements culturels ou sportifs par typologie de population ; autres applications en transport, dans la grande distribution...)
- Infrastructure de données et services autour de l'IoT : Data Venue, plateforme de collecte, d'agrégation, de stockage, et d'analyse des données issues de l'IoT
- Smart city : Projet de monitoring urbain environnemental à Nice (consortium Veolia, IBM, Orange) : déploiement de capteurs environnementaux, acheminement vers l'entrepôt de données de la Métropole
- Autres solutions de monitoring urbain : gestion intelligente des flottes de

véhicules, des déchets, de bornes de recharge électrique...

- Développement de **Ma Ville dans ma Poche**, déployée à Nantes, rassemblant l'ensemble des informations utiles aux usagers au sein d'une même application (horaire du prochain bus, fermeture des piscines, disponibilité des parking...)

DC Brain



Présentation

Startup française spécialisée dans la cartographie et le développement d'outils d'aide au pilotage et à l'exploitation des réseaux physiques (eau, électricité, gaz, logistique...). DC Brain est incubée au sein de Telecom ParisTech.

Offre

- L'outil big data DC Brain traite l'ensemble des données émises par les réseaux et permet aux équipes d'exploitation et d'ingénierie de visualiser en temps réel les flux, de diagnostiquer et anticiper les incidents, et de tester des scénarios d'évolution des réseaux
- Secteurs : gestionnaires de réseaux/utilities, infrastructures industrielles, Datacenters et logistique
- Cas d'usages :
 - Mise en place d'un modèle de scoring de câbles et analyse d'impacts prédictifs pour ERDF (lauréat du prix big data d'ERDF)
 - Aide à la priorisation des projets biométhane pour GRDF
 - Gestion de la consommation énergétique des datacenters d'Orange

Process One

Présentation

Société française spécialisée dans le traitement des données temps réel.

Offre

- Cœur de métier : plateformes et applications de messagerie instantanée pour différents types de clients (BBC, Nokia, Whatsapp, ...).
- Solutions smart city : plateforme IoT de traitement des données issues des objets connectés (données environnementales, de qualité de l'air...)

E. Analyses juridiques

1. Liste des principaux textes de loi examinés dans le cadre de cette étude

➤ Textes communautaires

- Directive 2013/37/UE modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public dite « Directive PSI » ;
- Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE.

➤ Textes nationaux

- Loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dite « loi CNIL » ;
- Loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal dite « loi CADA » ;
- Loi n°79-18 du 3 janvier 1979 sur les archives ;
- Décret n°2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, pris pour l'application de la loi n° 78-753 du 17 juillet 1978 ;
- Ordonnance n° 2015-899 du 23 juillet 2015 relative aux marchés publics ;
- Loi n° 2015-912 du 24 juillet 2015 relative au renseignement ;
- Loi n° 2015-990 du 6 août 2015 pour la croissance, l'activité et l'égalité des chances économiques, dite « loi MACRON » ;
- Loi n° 2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République, dite « loi NOTRe » ;
- Loi n° 2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte, dite « loi TECV » ;
- Loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public, dite « loi VALTER » ;
- Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé ;
- Ordonnance n° 2016-65 du 29 janvier 2016 relative aux contrats de concession ;
- Décret n° 2016-86 du 1er février 2016 relatif aux contrats de concession ;
- Décret n° 2016-360 du 25 mars 2016 relatif aux marchés publics ;

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

2. Chaîne de valeur de la donnée : les principaux enjeux juridiques

Si les enjeux pour les collectivités territoriales et les acteurs privés dans le processus de gestion de la donnée sont nombreux – et ont été renforcés dans le cadre de la loi pour une République numérique – ils ne sont pas les mêmes selon que l'on se situe au stade :

- De l'accès aux données (ou « open data ») **(1)** ;
- De la collecte des données **(2)** ;
- Du traitement, de la réutilisation, de l'archivage et du partage **(3)**.

➤ L'accès aux données

Du côté des acteurs publics

Comme cela a été rappelé ci-avant, ces dernières années, le législateur s'est engagé dans une politique volontariste d'ouverture des données publiques.

La plupart des lois adoptées ces dernières années incitent l'administration à mettre les données qu'elle détient à la disposition du public.

Outre les nombreux textes « sectoriels » précités, des textes généraux tels que la Loi n° 2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République (ci-après « Loi NOTRe ») disposait dans son article 106 intitulé « **Transparence des données des collectivités territoriales** » :

*« Les collectivités territoriales de plus de 3 500 habitants ainsi que les EPCI à fiscalité propre auxquels elles appartiennent rendent accessibles en ligne les **informations publiques mentionnées à l'article 10 de la loi n° 78-753 du 17 juillet 1978 lorsque ces informations se rapportent à leur territoire et sont disponibles sous forme électronique**. Ces informations publiques sont offertes à la **réutilisation** dans les conditions prévues au chapitre II du titre Ier de la même loi. »*

Outre le caractère particulièrement général d'une telle disposition, force est de constater qu'elle préfigurait les nouvelles obligations que le législateur allait mettre à la charge des collectivités dans le cadre de la loi pour une République numérique.

En effet, le législateur a estimé nécessaire de modifier le cadre juridique de l'ouverture des données publiques pour passer de l'incitation à **l'obligation pour les administrations de mettre à disposition les données qu'elles détiennent**.

Toutefois et jusqu'à la loi pour une République numérique, la législation distinguait trois cas, soumis chacun à des conditions de mise en œuvre différentes :

- La communication,
- La publication,
- La réutilisation.

La loi introduit davantage de continuité entre ces différentes phases de communication, publication et réutilisation en généralisant la mise à disposition par les administrations des données qu'elles détiennent.

Du côté des acteurs privés

De leur côté, les acteurs privés sont également soumis ces dernières années à davantage d'obligations de mise à disposition d'informations qu'ils détiennent.

Ces obligations sont bien évidemment renforcées lorsque l'on se situe dans le champ des contrats publics et, en particulier des délégations de service public, thématique ayant occupé une large partie des débats ayant eu lieu au Parlement dans le cadre de la loi pour une République numérique.

En effet, la loi crée un nouveau dispositif d'ouverture des données pour renforcer l'information des personnes publiques sur les services qu'elles délèguent.

Enfin, les délégations de service public de certains secteurs jugés « sensibles » sont soumises à des obligations renforcées de mise à disposition d'informations.

A titre d'illustration, dans le secteur des transports publics, les entreprises délégataires doivent, conformément aux dispositions de l'article L. 1115-1 du code des transports, diffuser certaines données « *librement, immédiatement et gratuitement en vue d'informer les usagers et de fournir le meilleur service* » comme les tarifs, les horaires, l'évolution de la fréquentation, etc.

➤ La collecte des données

Du côté des acteurs publics

Il n'est pas toujours aisé pour les collectivités territoriales d'obtenir de la part de leurs partenaires privés, l'ensemble des informations relatives aux services qu'elles délèguent, lesquelles ne sont en général pas accessibles « en ligne ».

Or, l'exploitation d'un contrat public donne aujourd'hui lieu à la production d'un volume croissant de données, en particulier dans les domaines de l'eau ou encore des transports, par exemple.

Il est plus qu'indispensable pour la collectivité de pouvoir disposer de ces données, d'une part, afin d'en assurer leur mise à disposition dans les conditions rappelées au point 1.1 ci-avant, et, d'autre part afin de pouvoir assurer efficacement le suivi, l'évolution mais surtout le renouvellement du contrat de délégation de service public.

Du côté des acteurs privés

De leur côté, les partenaires privés des collectivités territoriales détiennent, dans le cadre de l'exécution des missions de service public qui leur ont été déléguées, des informations de nature diverses et potentiellement sensibles :

- Informations commerciales,
- Informations personnelles,
- Savoir-faire industriel,
- Cyber-sécurité.

Aujourd'hui, certains délégataires ou détenteurs de droits exclusifs tels qu'ERDF, par exemple, masquent les données commercialement sensibles ou comportant des données à caractère personnel.

En effet, ces derniers tendent à se retrancher derrière la responsabilité civile et pénale qui est la leur aux termes des dispositions de la loi CNIL, comme exposé au sein de la fiche consacrée à la responsabilité CNIL.

➤ *Le traitement, la réutilisation et l'archivage*

Du côté des acteurs publics

De nombreuses incertitudes persistent à l'heure actuelle, compte tenu des différentes problématiques que suscite chacune de ces tâches que sont :

- Le traitement,
- La réutilisation,
- Et l'archivage de la donnée.

S'il est renvoyé aux points du rapport consacrés au traitement de données au sens des dispositions de la Loi CNIL, seront successivement exposés les principaux enjeux liés à la réutilisation (3.1.1) et à l'archivage de la donnée (3.1.2).

Les conditions de réutilisation des données publiques

De récentes modifications ont été apportées par le législateur aux conditions de réutilisation des informations publiques.

En effet, il ressort du nouvel article L321-1 du code des relations entre le public et l'administration que :

« Les informations figurant dans des documents produits ou reçus par les administrations mentionnées à l'article L. 300-2, quel que soit le support, peuvent être utilisées par toute personne qui le souhaite à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus.

Lorsqu'elles sont mises à disposition sous forme électronique, ces informations le sont, si possible, dans un standard ouvert et aisément réutilisable, c'est-à-dire lisibles par une machine. »

Cet article tel qu'issu de l'ordonnance n° 2016-307 du 17 mars 2016 (ordonnance portant codification des dispositions relatives à la réutilisation des informations publiques dans le code des relations entre le public et l'administration prise en application notamment de la loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public) définit la réutilisation d'informations comme **leur utilisation à d'autres fins que celles de la mission de service public pour les besoins de laquelle ces informations ont été produites ou reçues.**

Etant précisé que, aux termes des dispositions de l'article L. 322-2 du même code et conformément aux prescriptions de la loi CNIL, les informations publiques comportant des données à caractère personnel peuvent faire l'objet d'une réutilisation soit lorsque la personne intéressée y a consenti, soit si l'autorité détentrice est en mesure de les rendre anonymes ou, à défaut d'anonymisation, si une disposition législative ou réglementaire le permet.

Enfin, si elle ne l'a pas rendue obligatoire, la loi n° 2015-1779 du 28 décembre 2015 précitée a affirmé la faculté pour les administrations de soumettre à licence de réutilisation les informations publiques à titre gratuit.

Une licence de réutilisation a pour objet de fixer les conditions spécifiques de réutilisation des informations publiques. Elle permet également, comme le rappelait la CADA dans un récent avis, de rappeler de manière pédagogique les règles générales relatives à la réutilisation des informations publiques figurant désormais aux articles L. 322-1 et L. 322-2 du CRPA³⁹.

Ces nouvelles dispositions ont ouvert un nouveau champ d'action aux collectivités territoriales qui ne sera pas sans générer des incertitudes quant à leurs modalités de mise en œuvre.

L'archivage des données publiques

L'ensemble du régime juridique applicable à l'archivage de manière générale figure au sein du Titre 1er du livre II du code du patrimoine lequel régit, notamment, la définition et la justification des archives, le réseau des archives de l'État, l'exercice du contrôle scientifique et technique de l'État sur les archives publiques, les modalités de collaboration entre les services producteurs et les services d'archives, les versements dans les services publics d'archives, l'accès aux archives et au régime de communication des archives ainsi que les sanctions pénales.

Par ailleurs, le décret n° 2012-479 du 12 avril 2012 a créé auprès du Premier ministre deux nouvelles instances dédiée à la problématique de l'archivage : **le délégué et le comité interministériel aux archives de France dont le directeur, chargé des Archives de France assure le secrétariat.**

En premier lieu, l'article L211-4 dudit code définit les archives comme :

- Les documents qui procèdent de l'activité, dans le cadre de leur mission de service public, de l'Etat, des collectivités territoriales, des établissements publics et des autres personnes morales de droit public ou des personnes de droit privé chargées d'une telle mission ;
- Les minutes et répertoires des officiers publics ou ministériels.

En deuxième lieu, les articles L. 212-6 et L. 212-7 du code du patrimoine prévoient que les collectivités territoriales et leurs groupements sont « propriétaires de leurs archives ». Elles en assurent elles-mêmes la conservation et la mise en valeur.

Toutefois, les régions et la collectivité territoriale de Corse peuvent également confier la conservation de leurs archives, par convention, respectivement au service d'archives du département où se trouve le chef-lieu de la région ou de la collectivité territoriale de Corse.

Les groupements de collectivités peuvent, quant à eux, les déposer au service départemental d'archives compétent.

En troisième lieu, les articles L. 213-2 et L. 213-3 du code du patrimoine organisent le régime de communicabilité des archives publiques en appliquant des délais différenciés selon la nature de ces archives. Ils ne précisent cependant pas le régime de réutilisation des informations publiques figurant dans ces archives.

À défaut d'interdiction ou d'encadrement de la réutilisation des informations publiques contenues dans les archives publiques, la cour administrative d'appel de Lyon a jugé que :

³⁹ CADA, conseil 20155079 du 19 novembre 2015, Secrétariat d'État chargé du Numérique.

« les informations publiques communicables de plein droit, figurant dans les documents détenus par les services d'archives publics, qui constituent des services culturels au sens des dispositions de l'article 11 de la loi du 17 juillet 1978, relèvent de la liberté de réutilisation consacrée de façon générale par cette loi, dans sa rédaction issue de l'ordonnance du 29 avril 2009 »⁴⁰

La cour administrative d'appel a toutefois estimé qu'il appartenait :

« à l'autorité compétente, saisie d'une demande de réutilisation de ces documents, de s'assurer que cette réutilisation satisfai[sai]t aux exigences qu'imposent les dispositions de l'article 13 de cette loi qui, s'agissant d'informations publiques comportant des données à caractère personnel, renvoient aux dispositions de la loi n° 78-17 du 6 janvier 1978 ».

En dernier lieu, le code du patrimoine prévoit un certain nombre d'infractions pénales réprimant la violation des règles régissant les archives.

Du côté des acteurs privés

Dans l'hypothèse où une entreprise produit de la donnée, cette activité entrera dans le champ concurrentiel : son traitement sera valorisé et les modalités de réutilisation et de partage commercialisées.

Il est renvoyé à cet égard à l'étude menée en parallèle de la présente étude par la DGE et traitant plus précisément de ces aspects.

➤ Synthèse : Schéma récapitulatif des différents enjeux

ENJEU N° 1 : L'ACCES AUX DONNEES	Du côté des acteurs publics	Vers la généralisation massive de la diffusion et mise à disposition de données publiques collectées, notamment, dans le cadre de services publics ainsi que dans le cadre de contrats publics (DSP). Passage d'une logique de demande d'accès à une offre d'accès.
	Du côté des acteurs privés	Obligation de fournir aux collectivités les données collectées et produites à l'occasion d'un service public: Loi PRN / Loi TECV / Loi NOTRE / Directives marchés.
ENJEU N° 2 : LA COLLECTE DES DONNEES	Du côté des acteurs publics	Difficultés pour les collectivités d'obtenir auprès de leurs partenaires privés des informations nécessaires au fonctionnement des services publics.
	Du côté des acteurs privés	Contraintes particulières liées à la préservation et à l'anonymisation des données personnelles ou commercialement sensibles et, notamment, à l'obtention, en temps réel des données de consommation.
ENJEU N° 3 : LE TRAITEMENT, LA REUTILISATION ET LE PARTAGE	Du côté des acteurs publics	Incertitudes liées aux conditions de réutilisation des données publiques, notamment dans le cadre de licence de réutilisation ou encore au stade de l'archivage des données.
	Du côté des acteurs privés	Dans l'hypothèse où une entreprise produit de la donnée, cette activité entrera dans le champ concurrentiel : son traitement sera valorisé et les modalités de réutilisation et de partage commercialisées.

⁴⁰ CAA Lyon 4 juillet 2012, Département du Cantal, req. N° 11LY02325.

3. Les responsabilités administrative et pénale liées à la protection des données personnelles

Le respect, par les collectivités locales, des règles de protection des données à caractère personnel est un facteur de transparence et de confiance à l'égard des usagers. C'est aussi un gage de sécurité juridique pour les élus qui, responsables des fichiers mis en œuvre, doivent veiller à ce que la finalité de chaque traitement informatique et les éventuelles transmissions d'informations soient clairement définies, les dispositifs de sécurité informatique précisément déterminés et les mesures d'information des administrés appliquées.

Après avoir rappelé les conditions de licéité des traitements de données à caractère personnel **(1)**, les conditions de mise en jeu de la responsabilité administrative et/ou pénale des collectivités territoriales au titre de ce dispositif seront exposées **(2)**.

➤ *Conditions de licéité des traitements de données à caractère personnel*

Le Chapitre II de la loi CNIL consacré aux conditions de licéité des traitements de données à caractère personnel peut être exposé autour des cinq notions clefs figurant au sein des articles 6, 7 et 34 de la loi, à savoir :

- **La finalité** : les données devant être collectées pour des finalités « *déterminées, explicites et légitimes* » et ne doivent pas être traitées ultérieurement « *de manière incompatible avec ces finalités* » ;
- **La pertinence** : seules doivent être collectées les données « *adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont collectées et de leurs traitements ultérieurs* » ;
- **La conservation** : les données doivent être conservées « *pendant une durée qui n'excède pas la durée nécessaire aux finalités pour lesquelles elles sont collectées et traitées.* » ;
- **Le respect des droits des personnes** : un traitement de données à caractère personnel doit avoir reçu le consentement de la personne concernée ou satisfaire à l'une des conditions suivantes :
 - Le respect d'une obligation légale incombant au responsable du traitement ;
 - La sauvegarde de la vie de la personne concernée ;
 - L'exécution d'une mission de service public dont est investi le responsable ou le destinataire du traitement ;
 - L'exécution, soit d'un contrat auquel la personne concernée est partie, soit de mesures précontractuelles prises à la demande de celle-ci ;
 - La réalisation de l'intérêt légitime poursuivi par le responsable du traitement ou par le destinataire, sous réserve de ne pas méconnaître l'intérêt ou les droits et libertés fondamentaux de la personne concernée.
- **La sécurité** : le responsable du traitement est tenu de prendre toutes précautions utiles, au regard de la nature des données et des risques présentés par le traitement, pour préserver la sécurité des données et, notamment, empêcher qu'elles soient déformées, endommagées, ou que des tiers non autorisés y aient accès.

➤ Conditions de mise en jeu de la responsabilité administrative et pénale des collectivités

La CNIL, dès lors qu'elle est compétente, dispose d'un pouvoir de sanction en cas de réutilisation contraire à une disposition de la loi CNIL conformément au chapitre VII de cette dernière.

Les décisions prononçant une sanction peuvent faire l'objet d'un recours de pleine juridiction devant le Conseil d'État, en vertu du dernier alinéa de l'article 46 de cette même loi.

La loi CNIL renvoie enfin, aux termes de son article 50, pour les sanctions pénales, aux dispositions des articles 226-16 à 226-24 du code pénal, relatifs aux atteintes aux droits de la personne résultant des fichiers ou des traitements informatiques.

A cet égard, les collectivités territoriales et leurs groupements ne sont responsables pénalement que des infractions commises dans l'exercice d'activités susceptibles de faire l'objet de conventions de délégations de services publics, comme cela est rappelé par l'article 121-2 du code pénal :

« Les personnes morales, à l'exclusion de l'Etat, sont responsables pénalement, selon les distinctions des articles 121-4 à 121-7, des infractions commises, pour leur compte, par leurs organes ou représentants.

Toutefois, les collectivités territoriales et leurs groupements ne sont responsables pénalement que des infractions commises dans l'exercice d'activités susceptibles de faire l'objet de conventions de délégation de service public.

Aux termes du dernier alinéa de cet article, cette responsabilité n'exclut pas toutefois la responsabilité pénale des personnes physiques auteurs ou complices des mêmes faits (art. 121-2 du code pénal, par renvoi de l'art. 226-24).

Les principales infractions réprimées par le code pénal peuvent être résumées de la manière suivante :

AU STADE DE LA COLLECTE	5 ans de prison / 300 000 € d'amende	Fait de collecter des données à caractère personnel par un moyen frauduleux, déloyal ou illicite (art. 226-18).
AU STADE DE LA CONSERVATION	5 ans de prison / 300 000 € d'amende	- Conservation en mémoire informatisée, non prévue par la loi, ou non autorisée par l'intéressé, de données à caractère personnel concernant notamment les origines raciales, ethniques, opinions politiques... les infractions, condamnations, mesures de sûreté (art. 226-19). - Conservation au-delà de la durée prévue par la loi (art. 226-20).
AU STADE DU TRAITEMENT / DE LA RÉUTILISATION / DU TRANSFERT	5 ans de prison / 300 000 € d'amende 3 ans de prison / 100 000 € d'amende pour la divulgation prévue à l'art. 226-22	- Traitement de données à caractère personnel sans le consentement des personnes concernées ou malgré leur opposition, à des fins de prospection notamment commerciale (art. 226-18-1) ou dans le domaine de la santé (art. 226-19-1); - Détournement des informations à caractère personnel de leur finalité <u>lors de leur enregistrement, classement, transmission ou toute autre forme de traitement</u> (art. 226-21); - Fait de porter à la connaissance de tiers n'ayant pas qualité pour recevoir les données à caractère personnel, des données dont la divulgation porterait atteinte à la considération de l'intéressé ou à sa vie privée (art. 226-22); - Transfert de données à caractère personnel faisant l'objet ou destinées à faire l'objet d'un traitement vers un Etat hors UE (art. 226-22-1).

4. Encadrement juridique des algorithmes

➤ *Rappels des enjeux juridiques*

Comme le relevait le Conseil d'Etat dans son Etude annuelle 2014 consacrée au numérique et aux droits fondamentaux, il devient aujourd'hui essentiel de prendre la mesure du rôle d'intermédiation joué par les algorithmes et de concevoir leur encadrement juridique et leur utilisation.

A cet égard, le Conseil d'Etat préconisait trois méthodes d'encadrement :

- Assurer l'effectivité de l'intervention humaine dans la prise de décision au moyen d'algorithmes.
- Mettre en place des garanties de procédure et de transparence lorsque les algorithmes sont utilisés pour prendre des décisions à l'égard d'une personne.
- Développer le contrôle des résultats produits par les algorithmes, notamment pour détecter l'existence de discriminations illicites.

Or, aujourd'hui de nombreuses décisions individuelles sont prises par l'administration assistée de traitements algorithmiques, c'est-à-dire d'outils soumettant les données entrées par l'administration à une suite d'opérations ou d'instructions permettant d'aboutir à un résultat.

De tels instruments sont en particulier utilisés pour gérer de grandes masses de données, faisant intervenir de multiples facteurs à prendre en compte comme des listes de vœux ou des critères à croiser, afin d'optimiser les solutions.

Comme le relevait la CNIL dans son avis sur le projet de loi pour une République numérique⁴¹, la loi CNIL, apporte déjà certaines garanties, au titre notamment de son article 10, lequel interdit qu'une « *décision produisant des effets juridiques à l'égard d'une personne physique* » puisse résulter de la seule mise en œuvre d'un traitement de données « *destiné à définir le profil de l'intéressé ou à évaluer certains aspects de sa personnalité* », sans intervention humaine.

Toutefois, ce dispositif demeure limité aux traitements de données à caractère personnel.

➤ *Le nouveau dispositif introduit par la loi pour une République numérique*

L'article 4 de la loi introduit un nouvel article L. 311-3-1 au sein du code des relations entre le public et l'administration intitulé « *Droit d'accès aux règles et caractéristiques de l'algorithme intervenu dans la prise d'une décision individuelle* ».

Ce nouvel article a vocation à créer, au bénéfice de l'utilisateur qui a fait l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, un droit d'accès aux règles définissant le traitement algorithmique ainsi qu'aux principales caractéristiques de sa mise en œuvre.

⁴¹ CNIL, délibération n° 2015-414 du 19 novembre 2015 portant avis sur un projet de loi pour une République numérique.

Ce nouveau dispositif s'étendrait donc à toutes les personnes, physiques ou morales, ainsi qu'à toutes les décisions administratives individuelles prises sur le fondement d'un traitement algorithmique.

En outre, à l'initiative, notamment du Député Christian Paul, l'Assemblée nationale a introduit en séance publique un nouvel article L. 312-1-3 du CRPA visant à la publication des règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de leurs missions lorsqu'ils fondent des décisions individuelles. Cet ajout a été conservé par le Sénat.

5. Les enjeux liés à la notion de redevance

La directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public (dite « directive PSI ») avait été transposée par l'ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.

Afin de préciser le champ d'application de la directive PSI, d'encadrer le traitement des demandes de réutilisation des informations publiques, ainsi que la mise en place de redevances ou la conclusion d'accords d'exclusivité, la directive 2013/37/UE du Parlement européen et du Conseil a été adoptée le 26 juin 2013 **(1)**.

Cette Directive a été transposée par la loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public (ci-après « Loi VALTER ») **(2)** dont la mise en œuvre suscitera des interrogations **(3)**.

➤ *Principales dispositions de la Directive PSI*

En premier lieu, aux termes de son article 4, la directive n'impose aucune obligation pour les États membres d'autoriser la réutilisation de documents.

Elle ne fait que créer des règles minimales qui s'appliquent lorsque la réutilisation est autorisée.

En deuxième lieu, les redevances pouvant être perçues sont encadrées par la directive.

Aux termes des dispositions de son article 6, le total des recettes provenant de la fourniture et des autorisations de réutilisation des documents ne peut ainsi dépasser les coûts de collecte, de production, de reproduction et de diffusion, tout en permettant un retour sur investissement raisonnable.

La directive 2013/37/UE a modifié ce principe : **désormais lorsqu'une réutilisation est soumise à redevance cette dernière sera basée uniquement sur les coûts marginaux de reproduction, de mise à disposition et de diffusion des données.**

En troisième lieu, les organismes publics peuvent autoriser la réutilisation des documents sans conditions ou peuvent imposer des conditions, le cas échéant par le biais d'une licence. Etant précisé que plusieurs articles de la directive 2013/37/UE renforcent l'obligation de transparence imposée aux administrations. C'est le cas notamment de son article 7 s'agissant de l'indication de la méthodologie utilisée pour la détermination du montant des redevances.

➤ *La transposition de la directive par la loi « Valter »*

La loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public (ci-après « Loi VALTER ») a consacré en droit français le principe de gratuité de l'utilisation et de la réutilisation des informations publiques **(2.1)** tout en prévoyant deux dérogations **(2.2)**.

L'affirmation du principe de gratuité

L'ensemble des acteurs sont conscients depuis plusieurs années de l'enjeu important que représente les données publiques et de la nécessité de les mettre gratuitement à disposition d'utilisateurs et de ré-utilisateurs pour renforcer la démocratie, développer l'économie et moderniser l'action publique.

Toutefois, de nombreuses administrations ont maintenu et/ou institué des redevances de réutilisation.

Comme le relevait Monsieur le Député Luc BELOT, rapporteur du projet de Loi VALTER déposé à l'Assemblée Nationale, le montant des principales redevances s'élève, selon le périmètre retenu, à une trentaine de millions d'euros, mais il semblerait qu'il connaisse une baisse continue au cours des dernières années.

RECETTES TIRÉES DES REDEVANCES DE RÉUTILISATION, PAR SERVICE BÉNÉFICIAIRE

Service bénéficiaire	2012
Institut national de la statistique et des études économiques	9 981 000 €
Institut national de l'information géographique et forestière	9 940 748 €
Ministère de l'Intérieur	3 865 282 €
Institut national de la propriété intellectuelle	2 744 054 €
Ministère économique et financier	1 955 234 €
Météo-France (hors recettes refacturées à sa branche commerciale)	1 585 000 €
Service hydrographique et océanographique de la marine	1 300 000 €
Direction de l'information légale et administrative	892 326 €
Service de l'observation et des statistiques	580 000 €
Agence technique de l'information sur l'hospitalisation	543 719 €
FranceAgriMer	300 000 €
Cour de cassation	264 120 €
Conseil d'Etat	231 508 €
Office national d'information sur les enseignements et les professions	155 143 €
Ministère de l'éducation nationale	131 091 €
Institut français du cheval et de l'équitation	81 671 €
Institut national de l'origine et de la qualité	79 265 €
Agence de services et de paiement	53 480 €
Ministère de l'agriculture	16 700 €
Commission d'accès aux documents administratifs	5 000 €
TOTAL	34 705 341 €

Source : M. Mohammed Adnène Trojette, rapport sur l'ouverture des données publiques, juillet 2013.

Il ressort de nombreuses études réalisées ces dernières années et, en particulier, du rapport de M. Mohammed Adnène Trojette, « Ouverture des données publiques, les exceptions au principe de gratuité sont-elles toutes légitimes ? »⁴² que les effets de la gratuité de l'utilisation et de la réutilisation des informations publiques sont, à terme, extrêmement bénéfiques pour la société.

Sur la base de ce rapport, le Premier ministre a présenté au comité interministériel pour la modernisation de l'action publique (CIMAP) du 18 décembre 2012 la doctrine d'ensemble du Gouvernement en matière d'exception à la gratuité et les premières décisions de suppressions de redevances pour 2014.

Dans le prolongement de ces réflexions, le législateur a donc saisi l'occasion de la transposition de la Directive 2013/37/UE susvisée pour aller au-delà de la stricte

⁴² M. Mohammed Adnène Trojette, rapport au Premier ministre, « Ouverture des données publiques, les exceptions au principe de gratuité sont-elles toutes légitimes ? », juillet 2013.

transposition en inscrivant dans la loi le principe de gratuité décidé par le CIMAP le 18 décembre 2012.

Ainsi, l'article 5 de la loi VALTER désormais inséré au sein du nouvel article L. 324-1 du code des relations entre le public et l'administration prévoit que :

« La réutilisation d'informations publiques est gratuite. »

Les dérogations

L'article 5 susvisé institue deux dérogations au principe de gratuité.

D'une part, l'Etat, les collectivités territoriales ainsi que les autres personnes de droit public ou les personnes de droit privé chargées mission de service public peuvent établir une redevance de réutilisation « *lorsqu'elles sont tenues de couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service public.* ».

L'article 5 désormais inséré au sein du nouvel article L. 324-1 précité du code des relations entre le public et l'administration prévoit que :

« Le produit total du montant de cette redevance, évalué sur une période comptable appropriée, ne dépasse pas le montant total des coûts liés à la collecte, à la production, à la mise à la disposition du public ou à la diffusion de leurs informations publiques. »

Cette précision signifie que les redevances ne pourront donc plus prendre en compte, comme cela était le cas avant l'entrée en vigueur de la loi VALTER, la rémunération raisonnable des investissements, comprenant, le cas échéant, une part au titre des droits de propriété intellectuelle.

Au cours des débats parlementaires, ce choix a été justifié par le fait qu'une administration n'avait pas pour vocation de générer des bénéfices.

D'autre part, aux termes du nouvel article L. 324-2 du code des relations entre le public et l'administration, la réutilisation peut également donner lieu au versement d'une redevance dans un cas plus spécifique, à savoir :

« Lorsqu'elle porte sur des informations issues des **opérations de numérisation des fonds et des collections des bibliothèques, y compris des bibliothèques universitaires, des musées et des archives**, et, le cas échéant, sur des informations qui y sont associées lorsque ces dernières sont commercialisées conjointement. Le produit total du montant de cette redevance, évalué sur une période comptable appropriée, ne dépasse pas le montant total des coûts de collecte, de production, de mise à disposition ou de diffusion, de conservation de leurs informations et d'acquisition des droits de propriété intellectuelle. »

Là encore il est précisé que les redevances ne pourront donc plus prendre en compte, comme cela était le cas avant l'entrée en vigueur de la loi VALTER, la rémunération raisonnable des investissements, comprenant, le cas échéant, une part au titre des droits de propriété intellectuelle.

Dans les deux cas, conformément à la directive et au principe de non-discrimination d'ores et déjà inscrit dans la loi, le montant de la redevance devrait être fixé selon des **critères objectifs, transparents, vérifiables et non discriminatoires**.

En outre, l'article 5 de la loi a institué le principe d'une révision de ces montants au moins tous les cinq ans.

Les modalités de fixation de ces redevances sont renvoyées à un décret en Conseil d'État, pris après avis de la CADA.

➤ *Les interrogations suscitées par la loi VALTER*

Il est regrettable, que, pour des raisons liées au retard du gouvernement français dans la transposition de la Directive susvisée, le sujet des conditions financières de réutilisation des informations publiques n'ait pas été traité dans le cadre de la loi pour une République numérique.

Le manque de précision, tant dans la loi que dans les travaux parlementaires sur les modalités de mise en œuvre des redevances traduit la rapidité avec laquelle le parlement a du traiter un sujet aussi majeur.

En effet, comment interpréter la notion *« part substantielle des coûts liés à l'accomplissement de leurs missions de service public. »* ?

Comment calculer le montant total des coûts *« liés à la collecte, à la production, à la mise à la disposition du public ou à la diffusion de leurs informations publiques »* ?

Comme cela a été relevé par les deux rapporteurs du projet de loi pour une République numérique à l'Assemblée Nationale et au sénat, alors que le rapport Trojette précité incitait l'Etat à engager une réflexion sur les « modèles économiques » qui pourraient être envisagés pour les administrations, ni la loi VALTER, ni la loi pour une République numérique n'en font mention.

Pourtant, et comme le relevait Monsieur Trojette dans son rapport : *« les opérateurs dont la mission même est de produire des données doivent rechercher des modèles économiques leur permettant de faire face à un paysage économique en profonde reconstitution »*.

Ainsi, en l'absence de précisions du législateur, il appartiendra donc aux collectivités territoriales de rechercher des modèles stimulant l'innovation autour de leurs données, à la fois pour les entrepreneurs innovants, mais également pour les citoyens eux-mêmes.

6. La cybersécurité : nouvel enjeu de sécurité nationale

Afin de profiter des progrès des technologies de l'information et de la communication, les infrastructures critiques reposent de plus en plus sur des systèmes d'information complexes et se sont largement connectées à des réseaux publics.

Ces interconnexions rendent les infrastructures interdépendantes et les exposent à l'ensemble des vulnérabilités des systèmes informatiques.

➤ *Présentation du programme PEPIC*

Le Programme européen pour la protection des infrastructures critiques (dit « programme PEPIC »), lancé en 2004 par le Conseil européen, vise à identifier et à protéger les infrastructures critiques, entendues comme toute sorte d'infrastructure (y compris les services), considérée comme essentielle au fonctionnement de l'économie et de la société.

Plus précisément, il s'agit des installations physiques, des technologies de l'information, des réseaux, des services et actifs qui, en cas d'arrêt ou de destruction, peuvent avoir de graves incidences sur la santé, la sécurité ou le bien-être économique des citoyens ou encore le travail des gouvernements des Etats membres.

Indispensables au bon fonctionnement du pays, elles constituent des cibles privilégiées : il s'agit notamment de :

- La distribution d'énergie électrique (auprès d'autres infrastructures : hôpitaux, etc.) ;
- La production d'énergie électrique en particulier nucléaire ;
- Des réseaux d'alimentation et de production des raffineries ;
- La distribution et production d'eau douce ;
- Des réseaux de transport (réservations billets d'avions, contrôle aérien, réseaux de signalisation des voies ferrées, etc.) ;
- Des réseaux de communication (téléphone filaire, cellulaires, réseau Internet, etc.) y compris ceux des forces de police et de la défense.

L'objectif du programme PEPIC était alors de fixer un cadre commun, au niveau de l'Union européenne, pour la protection des infrastructures critiques en Europe afin de s'assurer que tous les États membres offrent des niveaux de protection suffisants de ces infrastructures.

Dans ce cadre, la Commission s'est ainsi vue confier la tâche de :

- Recenser et diffuser les informations relatives aux meilleures pratiques en matière de protection de ces infrastructures,
- Adopter des normes communes au niveau de chaque secteur,
- Evaluer les menaces et les risques.

Il en résulte un cadre réglementaire composé des éléments suivants :

- Une procédure pour l'identification et la désignation des infrastructures critiques européennes et une approche commune pour évaluer le besoin d'amélioration de leur sécurité ;

- Des mesures destinées à faciliter l'amélioration du programme incluant un plan d'action, un réseau d'alerte concernant les infrastructures critiques (CIWIN), l'établissement de groupes d'experts de la protection des infrastructures critiques (PIC) au niveau de l'UE, des procédures de partage d'informations concernant la PIC et l'identification et l'analyse des liens de dépendance ;
- Un soutien aux États membres en ce qui concerne la sécurité des infrastructures critiques nationales, sur leur demande et des plans d'intervention ;
- Une dimension extérieure ;
- Des mesures financières d'accompagnement, et en particulier le programme spécifique « *Prévention, préparation et gestion des conséquences en matière de terrorisme et autres risques liés à la sécurité* » de l'UE pour la période 2007-13, qui fournira des opportunités de financement pour les mesures liées à la protection des infrastructures critiques.

Tout en gardant à l'esprit que la protection des infrastructures critiques nationales incombe aux propriétaires, aux exploitants et aux États membres, la Commission prévoit néanmoins un soutien en la matière, à la demande des États.

Chaque État membre est par ailleurs encouragé à établir son programme national de protection incluant les éléments suivants :

- Le classement des infrastructures, tenant compte des effets suivant l'arrêt ou la destruction de chaque infrastructure (ampleur de la zone géographique touchée et gravité des conséquences) ;
- Le recensement des liens de dépendance géographique et sectorielle ;
- L'établissement de plans d'intervention.

➤ *Le cadre juridique national*

Si l'Etat français a atteint un haut degré dans la diffusion et l'usage des systèmes d'informations, il n'a sans doute pas accordé suffisamment d'importance à la sécurité de ces systèmes.

Les entreprises et les opérateurs d'importance vitale français demeurent encore insuffisamment sensibilisés à la menace liée aux attaques contre les systèmes d'information.

Ce constat, dressé par le rapport Lasbordes en 2006⁴³, reste encore largement d'actualité.

Or, face à l'espionnage informatique, la problématique de la sécurité des systèmes d'information des entreprises - et notamment de celles des secteurs jugés stratégiques - représente un enjeu majeur.

Seuls douze secteurs d'importance vitale ont été identifiés, regroupant environ deux cents trente opérateurs ou entreprises, issus du secteur public ou du secteur privé.

⁴³ Rapport intitulé « *La sécurité des systèmes d'information - Un enjeu majeur pour la France* », Pierre Lasbordes, remis au Premier ministre Dominique de Villepin le 13 janvier 2006.

Enfin, reste la question centrale des opérateurs d'importance vitale.

Indispensables au bon fonctionnement du pays, les opérateurs d'importance vitale représentent aujourd'hui des cibles particulièrement vulnérables aux attaques informatiques.

La principale difficulté tient cependant à la très grande diversité des opérateurs d'importance vitale.

On constate, en effet, de fortes différences entre les secteurs concernés, qu'il s'agisse de l'existence ou non d'une autorité de régulation, en termes de réglementation ou encore de relations avec la puissance publique.

Ainsi, dans certains secteurs, à l'image du secteur bancaire, de l'aviation civile ou encore de l'énergie nucléaire, les préoccupations de sécurité sont majeures et l'autorité de régulation joue un rôle important.

Mais il n'en va pas de même dans tous les secteurs.

Or, l'Agence Nationale de Sécurité des Systèmes d'Information (ANSSI) n'a pas les moyens d'assurer la protection de tous les opérateurs d'importance vitale et il est donc indispensable d'encourager les opérateurs, sur une base sectorielle, à renforcer les mesures de protection de leurs systèmes d'information.

De nombreux pays, à l'image des Etats-Unis ou de l'Allemagne, ont fait de la protection des infrastructures d'importance vitale une priorité nationale.

Or, dans ce domaine, la France accuse encore un réel retard par rapport à nos principaux alliés et partenaires.

Comme cela a été confirmé par l'ensemble de ses interlocuteurs, les échanges entre l'ANSSI et les opérateurs d'importance vitale sont très limités et on constate une méconnaissance réciproque.

Ensuite, en raison de leur diversité, la protection des systèmes d'information n'est clairement pas une priorité pour la plupart de ces opérateurs.

Surtout, la plupart des opérateurs d'importance vitale ne sont pas organisés pour répondre efficacement à un grave incident informatique et l'ANSSI n'a pas les moyens de faire face à une crise générale paralysant un secteur entier du pays.

Enfin, la France ne dispose pas de capacités de protection et de systèmes permanents de détection des attaques informatiques à l'entrée des réseaux des opérateurs d'importance vitale.

A ce jour, il apparaît donc indispensable de faire de cette question une priorité nationale.

L'État a la responsabilité, en relation avec les représentants des secteurs stratégiques économiques, de la protection de ces infrastructures vitales.

Le pilotage général de la protection des infrastructures vitales est confié au Secrétariat général de la Défense nationale, avec un rôle particulier pour le COSSI (centre opérationnel en SSI qui englobe le CERTA).

Un des objectifs de ce nouveau dispositif est d'arriver à un nombre de points d'importance vitale sensiblement inférieur à celui des actuelles installations et points sensibles, afin de mieux les protéger.

➤ *L'insuffisante sensibilisation des collectivités territoriales en matière de sécurité informatique*

Les collectivités territoriales n'ont pas encore suffisamment pris conscience de la nécessité de sécuriser leurs systèmes d'informations, alors même qu'elles sont bien évidemment des cibles potentielles.

A ce jour, un faible nombre de collectivités a organisé des formations pour sensibiliser les agents et moins de 15% d'entre-elles ont admis ne pas avoir pris connaissance du Référentiel Général de Sécurité (RGS) de l'ANSSI (l'Agence nationale de la sécurité des systèmes d'information), auquel elles doivent pourtant se conformer, depuis mai 2013, en ce qui concerne les certificats électroniques.

Par ailleurs, aucune d'entre elles n'a encore eu recours au chiffrement des données.

Lors d'un colloque organisé le 16 mars 2016 par la Mission Ecoter, association ayant en charge le développement des systèmes de communication et d'information dans les collectivités territoriales, les intervenants ont aussi pris le temps d'expliquer que les collectivités territoriales étaient bien des cibles potentielles d'attaques, contrairement à ce que beaucoup d'entre elles tendent à penser.

Ces offensives pourraient pourtant déboucher sur la perte de données sensibles.

Primo France, association dédiée à la gouvernance et à la gestion du risque public, a conduit une enquête portant sur l'exposition des collectivités publiques locales au cyber risque, laquelle a conclu aux enseignements suivants :

- Les directeurs généraux des services, par leur vision transversale, restent les acteurs les plus impliqués en matière de gestion des risques ;
- Malgré une sensibilité aux risques qui ne cesse de croître, les mécanismes mis en place pour la gestion des cyber risques demeurent insuffisants ;
- L'exposition croissante des collectivités aux cyber risques est due, d'une part, à l'augmentation des accès connectés en interne et en externe (guichet unique, sites des collectivités, données sensibles, etc.) et, d'autre part, à la montée d'un cyber terrorisme, idéologique ou purement mercenaire ;
- L'achat d'assurance reste encore un outil de protection* peu utilisé par les collectivités qui appréhendent surtout des freins politiques, techniques et réglementaires liés à sa mise en œuvre.

Pour anticiper les problèmes, l'Agence Nationale de Sécurité des Systèmes d'Information (ANSSI) a mis en place une démarche d'homologation en neuf étapes permettant aux collectivités territoriales de sécuriser leurs systèmes d'information.

Ce processus assez lourd doit impérativement être adapté aux enjeux, au contexte d'emploi, à la nature des données et aux utilisateurs.

Le contexte cyber est encore jeune, puisque les principales attaques médiatisées datent pour la plupart de moins d'un an. On ne peut qu'anticiper une augmentation de ces attaques et de leurs conséquences.

Cependant, le constat est bien présent : les collectivités peinent à atteindre un niveau minimal de protection de leurs données, et sont, de fait, mal préparées aux futures révolutions apportées par le guichet unique et la ville connectée.

Les acteurs privés du secteur informatique poussent à centraliser les données, tout en les délocalisant (cloud, gestion électronique à distance, etc.), envoyant ainsi des signaux contradictoires et exposant les collectivités locales à un risque accru, nécessitant la mise en place d'une couverture efficace afin de réduire les conséquences néfastes d'une cyber-attaque.

7. Gouvernance de la mission locale de service public de la donnée

Au regard, tant des enjeux que des comptes rendus d'entretiens ayant eu lieu dans le cadre de la présente étude, il existe aujourd'hui une très grande hétérogénéité dans les processus de gestion de la donnée par les collectivités territoriales sur le territoire national.

Or, la plupart des retours d'expériences locales font majoritairement ressortir :

- La nécessité d'une mutualisation pour la gestion de la donnée territoriale ;
- L'impossibilité matérielle pour certaines collectivités territoriales, et en particulier les petites collectivités, de prendre en charge seules, un tel processus.

Dès lors, des réflexions doivent être menées, tenant compte des outils et/ou dispositifs d'ores et déjà existants, de sorte à faciliter la mise en œuvre d'un cadre de gouvernance partagé le plus efficace à l'échelle locale.

Si la tendance paraît s'orienter vers un pilotage « régional » de la gestion de la donnée - échelon susceptible d'être le plus approprié pour accompagner les collectivités (1) – des interrogations persistent sur les modalités juridiques les plus adaptées pour confier la gestion et le traitement de la donnée à un guichet unique institué à un niveau régional (2).

➤ *Le choix de la maille régionale*

Compte tenu des difficultés tenant à la superposition des autorités publiques productrices de données, d'une part, et susceptibles de vouloir en prendre en charge la gestion, d'autre part, (36.767 communes, 2.456 EPCI à fiscalité propre et 101 départements gérant des domaines publics ou privés), il est nécessaire de s'interroger sur l'échelon le plus pertinent pour la mise en place de dispositifs de mutualisation de la gestion de la donnée.

Il n'existe pas un scénario général adapté à la diversité des situations des collectivités ainsi qu'à la complexité des sujets se rapportant à la gestion de la donnée territoriale. L'idée qui ressort de l'étude est de faire émerger une structure de gestion des données à l'échelon régional qui pourrait mettre en réseau plusieurs structures.

En effet, la maille régionale est le niveau d'implantation le plus cohérent pour la mise en place d'une structure d'animation et de coordination commune.

Cette intervention à l'échelon régional s'inscrirait d'ailleurs parfaitement dans le cadre de la loi n° 2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République, dite « loi NOTRe », laquelle précise que la région a pour mission de contribuer au développement économique notamment par la « *coordination, au moyen d'une plateforme de services numériques qu'elle anime, de l'acquisition et de la mise à jour des données géographiques de référence nécessaires à la description détaillée de son territoire ainsi qu'à l'observation et à l'évaluation de ses politiques territoriales, données dont elle favorise l'accès et la réutilisation* » (art. 1^{er} de la loi).

Plusieurs modèles régionaux existent d'ores et déjà : GIP littoral aquitain, GIP e-bourgogne, CRIGE PACA...

La structure régionale créée pourrait ainsi regrouper des structures départementales ou métropolitaines qui adhèreraient à la structure régionale.

Tous les échelons territoriaux ne sont pas capables de faire de la gestion de la donnée, de sorte qu'il convient de créer une structure de mutualisation à l'échelon régional.

Le service régional couvrirait ainsi l'ensemble du territoire en s'appuyant sur les opérateurs existants et en intégrant les syndicats de mutualisation, les Métropoles, les départements.

La structure ainsi choisie pourrait avoir les caractéristiques suivantes :

- Une structure au service du secteur public et des acteurs économiques (une **structure ouverte**) ;
- La mise en place d'une **infrastructure de données publiques** (alphanumérique et géographique) ;
- Un **Chief Data Officer**, pièce centrale, pour piloter et organiser ;
- Des **compétences en maîtrise d'ouvrage sur la donnée** ;
- Une **gestion de l'ouverture des données publiques**, des relations avec les acteurs privés, de l'organisation des filières économiques qui en expriment le besoin ;
- Un **travail d'information, de sensibilisation et d'accompagnement des structures locales** dans le déploiement de leur propre outil et pour alimenter les outils mutualisés mis en place ;
- Une **bonne réutilisation des API et des données de référence** ;
- La **constitution d'un schéma des données d'intérêt général sur le territoire**, actualisable en concertation avec les producteurs locaux, et d'un magasin d'API et d'un annuaire de méta données correspondant (structuration des données de références ou d'intérêt général local) ;
- Un **regard sur la formation** (Chief Data Officer, data sciences, business intelligence, etc.);
- La **commande de services structurants sur le territoire régional** ;
- Un **chantier sur les modèles de données (coordonné) et sur les fonctions de business intelligence** qui pourraient y être associées ;
 - Entre les données structurées et non structurées, il y a tout un travail à faire sur les modèles de données. Cela pourrait être un chantier coordonné et également sur les fonctions de business intelligence qui pourraient y être associées.
- La **coopération avec l'échelon national, les autres régions et les métropoles**.

➤ *La constitution d'une structure de coordination*

Dans ce cadre régional et au vu des échanges ayant eu lieu dans le cadre de la présente étude, le Groupement d'Intérêt Public (GIP) nous paraît être la structure d'intervention et de coopération la plus adaptée pour le portage de la future plateforme applicative, du fait de sa grande souplesse.

Par ailleurs, il permettrait d'accueillir de nombreux partenaires pour la mise en œuvre d'un véritable big data territorial.

Le recours aux structures existantes ou aux plateformes d'ores et déjà en place au niveau local ne devra néanmoins pas être négligé.

Seront donc successivement exposées :

- Les modalités de création d'une structure prenant la forme d'un groupement d'intérêt public (2.1) ;

- Les modalités de fonctionnement et de gouvernance (2.2) ;
- Les missions confiées au GIP (2.3).

La mise en place d'une structure de portage instituée sur la base du volontariat : le groupement d'intérêt public (GIP)

Le GIP, personne morale de droit public dotée de l'autonomie administrative et financière, est une structure de coopération institutionnelle entre plusieurs personnes morales de droit public (au moins une) et de droit privé, lesquelles mettent en commun des moyens en vue d'exercer des activités d'intérêt général à but non lucratif.

Il ne peut être constitué entre des collectivités territoriales ou leurs groupements pour exercer des missions relevant des établissements publics de coopération intercommunale (EPCI).

Les GIP sont définis par l'article 98 de la loi n° 2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit (ci-après dite « **Loi Warsmann** »), selon lequel :

« Le groupement d'intérêt public est une personne morale de droit public dotée de l'autonomie administrative et financière. Il est constitué par convention approuvée par l'Etat soit entre plusieurs personnes morales de droit public, soit entre l'une ou plusieurs d'entre elles et une ou plusieurs personnes morales de droit privé.

Ces personnes y exercent ensemble des activités d'intérêt général à but non lucratif, en mettant en commun les moyens nécessaires à leur exercice ».

Le GIP est donc caractérisé par plusieurs éléments :

- Il dispose de la personnalité morale ;
- Il est une personne morale de droit public ;
- Il dispose de l'autonomie administrative et financière.
- Il vise la mise en commun de moyens entre des partenaires tant publics que privés ;
- Son objectif consiste en l'exercice d'activités d'intérêt général à but non lucratif en mettant en commun les moyens nécessaires à leur exercice, ce qui les distingue des groupements d'intérêt économique (GIE) qui visent, quant à eux, à faciliter ou à développer l'activité économique de leurs membres (article L. 251-1 du code de commerce).

S'il peut être constitué, à l'origine, sur une zone géographique définie, rien ne lui interdit ensuite d'intervenir sur l'ensemble du territoire.

En effet, les textes applicables au GIP imposent que la convention constitutive vise la « zone géographique couverte par le GIP ».

Trois solutions juridiques sont dès lors envisageables :

- Soit les statuts prévoient que la zone géographique visée par les textes est celle de la ou des collectivité(s) membre(s) fondateur(s) du GIP à titre principal et, à titre accessoire, toute autre entité participant du développement des objectifs du GIP dans le domaine de la donnée ;
- Soit les statuts prévoient explicitement que la zone géographique comprend celle de la ou des collectivités membres fondateurs du GIP. Cette solution aurait le mérite d'être parfaitement en ligne avec les textes régissant les GIP

mais aurait l'inconvénient d'être peu évolutive au regard des nouveaux entrants ;

- Soit enfin les statuts prévoient que la zone géographique considérée recouvre l'ensemble du territoire national. Si cette solution peut sembler plus risquée car le territoire national ne paraît pas véritablement correspondre à la notion de « zone géographique », elle nous semble pouvoir être juridiquement argumentée. En pratique, certains GIP ont d'ailleurs adopté cette rédaction extensive. Si cette solution pourrait être considérée comme compatible avec les textes régissant les GIP, il n'en reste pas moins que, lors de la création du GIP, la nature des membres fondateurs devront, à notre sens, justifier que sa zone géographique soit nationale.

Préalablement à la création du GIP, il est nécessaire que (i) les futurs partenaires aient été choisis, (ii) que ceux-ci aient donné leur accord et (iii) que la convention constitutive du GIP soit négociée, rédigée et approuvée par l'ensemble des adhérents.

La convention constitutive du GIP a pour objet de préciser les modalités de fonctionnement du groupement⁴⁴.

Elle doit ensuite être approuvée par l'Etat selon les modalités définies par le décret n° 2012-91 du 26 janvier 2012 relatif aux GIP⁴⁵. Un délai de quatre mois pour l'obtention de l'arrêté interministériel doit donc être pris en compte⁴⁶.

Le GIP jouit de la personnalité morale à compter de la publication de la décision approuvant sa convention constitutive⁴⁷.

Un GIP peut être créé dans des délais relativement courts (de six à huit mois).

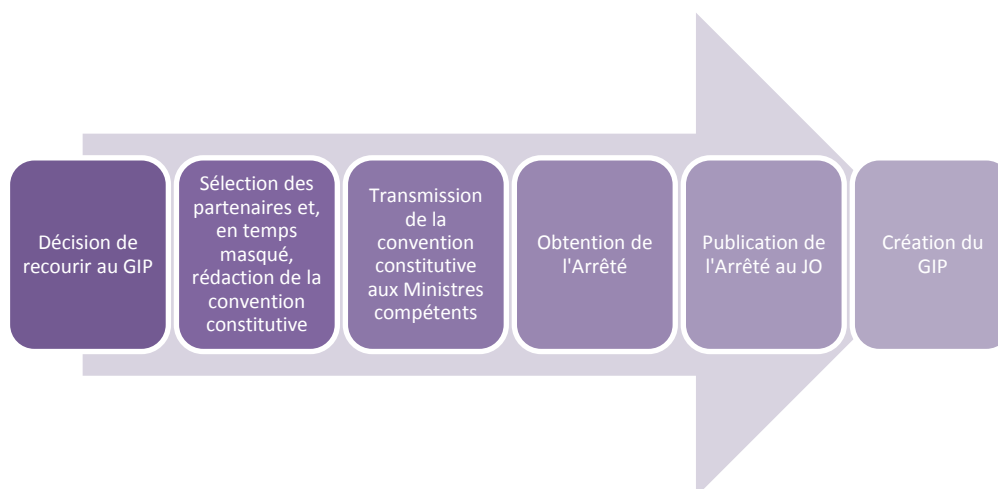
Synthèse des étapes de création d'un GIP :

44 Article 99 de loi Warsmann

45 Art. 100 de la Loi Warsmann

46 Il doit être précisé qu'a priori, la constitution du GIP comprendra des établissements publics nationaux ou autre entité publique à caractère national. Dans une telle hypothèse, le décret de 2012 précité ne prévoit pas la saisine du directeur régional ou départemental des finances publiques

47 Art. 4 du décret n° 2012-91 du 26 janvier 2012 relatif aux groupements d'intérêt public : « Le groupement d'intérêt public jouit de la personnalité morale à compter de la publication de la décision approuvant sa convention constitutive ».



Le recours au GIP nous semble présenter de nombreux avantages :

- Il apparaît comme la structure d'intervention et de coopération la plus souple, pérenne et compatible avec les problématiques liées à l'impossibilité de contraindre les collectivités territoriales à l'externalisation de la gestion de leurs données ;
- Les membres du GIP peuvent partager les coûts d'achat et de développement des services mutualisés ;
- Les collectivités membres du GIP peuvent recevoir des services du GIP sans le mettre en concurrence.

Les modalités de fonctionnement et de gouvernance du GIP

Les modalités de fonctionnement du GIP sont précisées dans la convention constitutive librement négociée entre ses membres.

Plusieurs institutions doivent être créées afin de gérer le GIP :

- **L'assemblée générale du GIP :**

Il s'agit de l'organe délibérant du GIP laquelle est compétente pour prendre « toute décision relative à l'administration du groupement, sous réserve des pouvoirs dévolus à d'autres organes par la convention constitutive »⁴⁸. Cette assemblée doit être majoritairement représentée – en voix ou en capital – par les personnes morales de droit public et les personnes morales de droit privé chargées d'une mission de service public.

Celle-ci est composée de l'ensemble des membres du GIP. Chaque membre dispose d'une voix, sauf si la convention constitutive en dispose autrement.

Seule l'assemblée générale a le pouvoir de modifier ou de renouveler la convention constitutive, de transformer le GIP en une autre structure ou de le dissoudre par anticipation.

- **Le Conseil d'administration :**

La loi Warsmann donne également la faculté à la convention constitutive d'établir un conseil d'administration, outre l'assemblée générale.

⁴⁸ Article 105 de la loi Warsmann

- **Le Directeur du GIP :**

Le fonctionnement du GIP est assuré par un directeur, qui peut être placé sous le contrôle de l'assemblée générale ou du conseil d'administration et dont la désignation se fait dans les conditions prévues par la convention constitutive. Le directeur peut également exercer les fonctions de président du conseil d'administration, si la convention constitutive le prévoit.

- **Personnel du GIP :**

Selon le choix opéré par le GIP et quelle que soit la nature de ses activités, tous les personnels, y compris le directeur, sont soumis soit aux dispositions du code du travail, soit à un régime de droit public, sous réserve des dispositions relatives à la mise à disposition prévues par le statut général de la fonction publique.

Le choix du régime du personnel est donc fixé librement par la convention constitutive du GIP, le décret n° 2013-292 du 5 avril 2013 relatif au régime de droit public applicable aux personnels des groupements d'intérêt public disposant que :

« En ce qui concerne les GIP créés après la publication de la loi du 17 mai 2011 et après la publication du décret du 5 avril 2013, ils sont soumis aux dispositions du II de l'article 110 de la loi, dans les conditions prévues par les dispositions du décret du 26 janvier 2012 qui régissent la constitution d'un groupement et, partant, l'approbation de sa convention constitutive et la détermination du régime juridique des personnels du GIP ».

Le régime du personnel sera donc librement déterminé par le GIP, au moment de sa création, mais devra cependant respecter les régimes en place, notamment, dans le cas des fonctionnaires mis à disposition.

Pour effectuer ses missions, le GIP pourra donc recourir à trois possibilités lui permettant de se doter de personnel :

- La mise à disposition d'agents publics ou de salariés de droit privé,
- Le recrutement de personnel en vue de se doter de certains profils spécifiques,
- Le détachement de personnel.

8. Proposition de clausier type : éléments susceptibles d'être intégrés au sein des contrats de la commande publique

Il appartient aujourd'hui aux collectivités territoriales de décider du degré de contrôle qu'elles souhaitent exercer sur la gestion des données des services publics délégués, en particulier.

Après avoir rappelé le contexte dans lequel s'inscrit ce nouvel objectif, il sera exposé dans quelle mesure le dispositif de protection des données à caractère personnel peut s'articuler avec celui d'un contrat de délégation de service public pour enfin lister les éléments susceptibles d'être intégrés au sein des contrats de délégation de service public.

➤ *Éléments de contexte*

Ces dernières années, des collectivités ont rencontré des difficultés pour récupérer les données de leurs services publics.

Il n'est pas toujours aisé pour les collectivités territoriales d'obtenir de la part de leurs partenaires privés, l'ensemble des informations relatives aux services qu'elles délèguent lesquelles ne sont en général pas accessibles « en ligne ».

Or, l'exploitation d'une DSP donne aujourd'hui lieu à la production d'un volume croissant de données, en particulier dans les domaines de l'eau ou encore des transports par exemple.

Il est plus qu'indispensable pour la collectivité de pouvoir disposer de ces données, d'une part, afin d'en assurer leur mise à disposition dans les conditions rappelées ci-avant, et, d'autre part afin de pouvoir assurer efficacement le suivi, l'évolution mais surtout le renouvellement du contrat de délégation de service public.

➤ *Articulation du dispositif de protection des données personnelles avec celui institué dans le cadre de contrats de délégation de service public*

La notion de données à caractère personnel telle que régie par les dispositions de la loi CNIL semble constituer une limite juridique qui interdit, sauf consentement de l'intéressé, fondement légal ou anonymisation, la communication ou la réutilisation des données.

En effet, le dispositif de protection des données personnelles se préoccupe peu des relations susceptibles d'exister entre un délégant et son délégataire.

En outre, le droit pénal constitue un obstacle supplémentaire en ce que la « responsabilité pénale » ne saurait se transférer ou se déléguer.

Il convient donc de s'interroger sur le lien contractuel qui pourrait être institué avant la conclusion du contrat entre le délégataire et le délégant et sur le niveau de responsabilité qui pourrait être mis à la charge de l'administration en sa qualité de responsable de traitement.

En conclusion : il conviendrait d'établir un lien entre le fait que la collectivité reste toujours responsable de l'organisation de son service public et qu'elle ne se contente que d'en déléguer la gestion, conformément à l'esprit même d'un contrat de délégation de service public.

En tout état de cause, ces difficultés pourraient être réduites si l'administration acceptait de prendre l'engagement de s'assurer du respect des dispositions de la loi CNIL au risque de se voir suspendre son droit de réutilisation, supprimer ou demander le rapatriement du jeu de données au profit du délégataire, s'il s'avère qu'il présente un risque pour le respect de la vie privée.

➤ Types de clauses susceptibles d'être insérées dans les contrats publics

La nécessité pour les collectivités d'insérer dans leurs contrats des clauses relatives à la gestion des données et bases de données collectées ou produites à l'occasion de l'exécution du contrat s'avère aujourd'hui nécessaire.

Il convient à cet égard de distinguer, les types de clauses qui pourraient être introduites selon qu'elles ont vocation à régir :

- Les modalités d'ouverture et de mise à disposition des données à destination des administrés (3.1) ;
- Le régime de propriété des données (3.2) ;
- La responsabilité liée au traitement desdites données (3.3).

Modalités d'ouverture et de mise à disposition des données à destination des administrés

Si le législateur s'est engagé ces dernières années dans une politique volontariste d'ouverture des données publiques allant même dans le cadre de la loi pour une République Numérique **à instituer une obligation pour les administrations de mettre à disposition les données qu'elles détiennent**, ces obligations se trouvent plus particulièrement renforcées dans le cadre des contrats publics, en particulier, des délégations de service public.

Or, l'étude approfondie des textes fait ressortir un empilement des différentes obligations dans des textes différents et tel que résumé dans le schéma ci-dessous :

Dans le cadre de la loi CADA	Dans le cadre de l'ordonnance et du décret concession	Dans le cadre du CGCT	Dispositions sectorielles
Article L. 311-1 du CRPA	Article 53 de l'ordonnance + article 34 du décret	Nouvel article L. 1411-13 du CGCT	Codes des transports
Si l'administration avait à sa charge de communiquer les documents administratifs qu'elles détiennent à la suite d'une demande d'accès, le projet de LPRN tend à inciter le passage à une logique de mise à disposition spontanée	Obligation de mise à disposition sur le profil acheteur des « données essentielles » avant le début d'exécution du contrat mais également chaque année à l'occasion de chaque avenant.	Dans les communes de + de 3,500 habitants les collectivités sont tenues de mettre à disposition du public en mairie des « documents relatifs à l'exploitation des services publics délégués ».	Obligation de diffusion libre immédiate et gratuite à la charge du délégataire (tarifs, horaires, évolution de la fréquentation, ...)

Outre cet empilement des textes, d'une façon générale, il appartiendra aux collectivités territoriales de s'interroger sur les modalités de mise en œuvre des nouvelles obligations d'open data qui ont été mises à sa charge d'autant plus que souvent, les différents échelons de collectivités territoriales n'ont pas tous appréhendés dans la même mesure ces nouvelles obligations.

Une fois que les collectivités auront en quelque sorte « défini » leur « politique » en matière d'open data, et dès lors que les dispositions de la loi CNIL sont respectées, il pourrait être inscrit dans les contrats que toutes données, produites ou reçues dans le cadre dudit contrat pourront être mises à la disposition du public dans les conditions définies par le Délégrant et après l'accord préalable de ce dernier.

Le régime de propriété des données

Il nous semble important qu'il soit désormais indiqué dès le lancement d'une procédure de passation que la collectivité souhaite être propriétaire de toutes les données et bases de données nécessaires à l'exploitation du service public.

Pour ce faire, il appartient aux collectivités d'inscrire au sein de leurs contrats que lesdites données relèvent de la catégorie juridique des « biens de retour ».

Pour rappel, la notion de bien de retour a été clairement définie par le Conseil d'Etat a dans un important arrêt d'Assemblée du 21 décembre 2012 (CE Ass 21 décembre 2012, *Commune de Douai*, req. n° 342788).

Les biens de retours correspondent aux biens affectés au service public et nécessaires à son exploitation et qui sont considérés comme étant, dès leur acquisition ou leur réalisation, propriété du délégant, et ce même dans les cas où ils ont été financés par le délégataire.

Aussi, il conviendrait d'indiquer expressément dans les futurs contrats que l'ensemble des données et bases de données produits ou reçus par le délégataire dans le cadre des missions qui lui ont été confiées constituent des biens de retour dès lors qu'ils sont par ailleurs nécessaires à la continuité du service public.

Il pourrait être par ailleurs envisagé que soient consenties au délégataire des licences non exclusives d'exploitation de ces bases de données pour toute la durée du contrat.

La responsabilité liée au traitement des données

Afin que les collectivités disposent d'un pouvoir de contrôle plus étendu sur la gestion de l'ensemble des données du service public, elles disposent de la possibilité de revêtir la qualité de « responsable de traitement » au sens des dispositions de l'article 3 de la loi CNIL lequel définit le responsable de traitement comme étant « *sauf désignation expresse par les dispositions législatives ou réglementaires relatives à ce traitement, la personne, l'autorité publique, le service ou l'organisme qui détermine ses finalités et ses moyens* ».

En l'occurrence, dès lors que la collectivité délégante détermine les finalités et les moyens de mise en œuvre du traitement des données du service, elle nous semble pouvoir être considérée comme responsable du traitement correspondant et

assumer à ce titre l'ensemble des obligations prescrites par la Loi CNIL précitée mais également les responsabilités civiles et pénales qui en découlent.

Dans l'hypothèse où la collectivité serait considérée comme responsable du traitement, il reviendrait alors au concessionnaire, en qualité de « sous-traitant » au sens des dispositions de la loi CNIL, d'assurer la confidentialité et la sécurité des données du service conformément aux engagements qui seront inscrits dans le contrat pour la couverture des risques résiduels. En sa qualité de sous-traitant, le concessionnaire ne pourrait alors agir que sur instruction du concédant.

En effet, il sera rappelé qu'aux termes de l'article 35 de la loi CNIL, un sous-traitant est défini comme « *toute personne traitant des données à caractère personnel pour le compte du responsable des traitements* ».

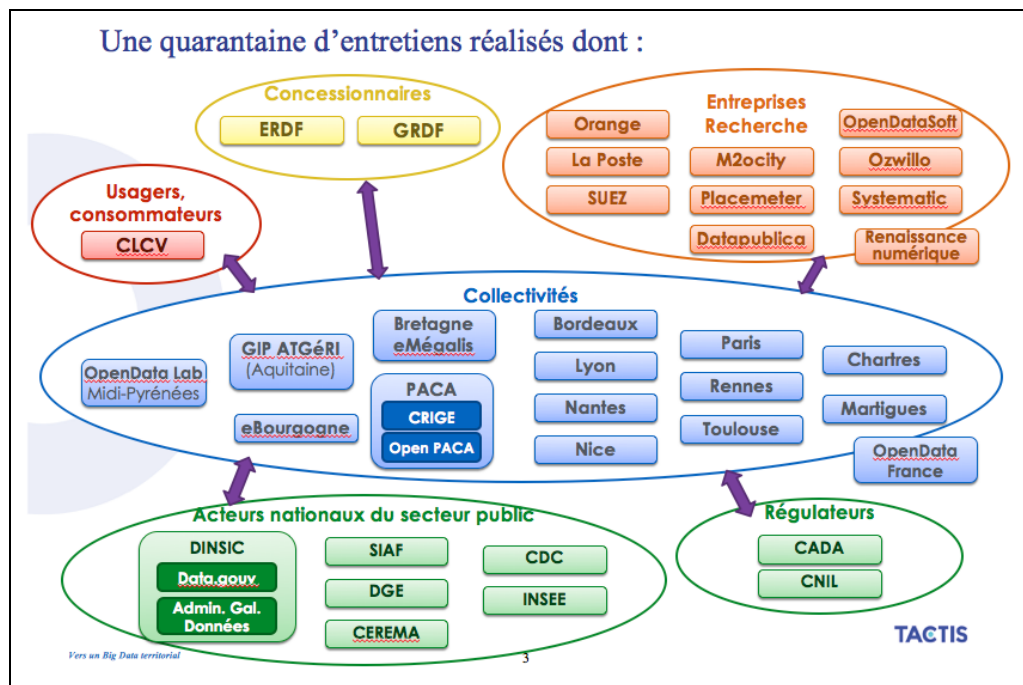
Dès lors, lorsqu'un traitement est sous-traité à un prestataire externe, ledit prestataire ne devient pas le responsable du traitement pour autant étant donné que seul ledit responsable de traitement conserve toute la responsabilité du traitement au yeux de la loi.

Ainsi, au regard de ces différents éléments, il appartient aux collectivités, compte tenu des conséquences attachées en termes de responsabilité à la qualité de responsable de traitement, de décider d'inscrire ou non un tel principe dans le projet de contrat de délégation de service public en cours de négociation.

➤ **Synthèse : schéma récapitulatif des différents types de clauses susceptibles d'être introduites au sein des contrats de concession**

AU TITRE DE L'ACCES A LA DONNEE (OU OPEN DATA)	Sur le fondement de la loi CADA	L'ensemble des documents collectés ou produits par le délégataire dans le cadre de ses missions de service public sont des documents administratifs dès l'origine.
	Sur le fondement de dispositions sectorielles	Dès lors qu'il ne s'agit pas de données personnelles, il doit être prévu que toutes données collectées et/ou reçues dans le cadre du contrat pourront être mises à la disposition du public après accord préalable du délégant
AU TITRE DE LA RESPONSABILITE DU TRAITEMENT	Principe : La collectivité responsable du traitement	- Le délégant détermine les finalités et les moyens de mise en œuvre du traitement des données du service; - Le délégant assume en conséquence l'ensemble des obligations et des responsabilités au titre de la Loi CNIL.
	Mise en œuvre	Si la collectivité est responsable de traitement, afin de ne pas « déresponsabiliser » le délégataire, il conviendra de lui en « sous-traiter » le traitement sous le contrôle du délégant
	Dérogation : Le délégataire responsable de traitement	En fonction des « traitements » mis en œuvre dans le cadre de l'exploitation du service, le délégataire pourra revêtir la qualité de responsable de traitement et en assumera la responsabilité. Cette responsabilité n'emporte aucune conséquences sur la propriété du délégant sur les données concernées
AU TITRE DE LA PROPRIÉTÉ DE LA DONNEE	Au stade de la négociation du contrat	L'ensemble des données nécessaires à l'exploitation du service ainsi que l'ensemble des éléments du SI doivent être propriété du délégant et constituer des biens de retour en ce qu'ils sont nécessaires pour assurer la continuité du service public
	A la fin du contrat de DSP	Le Délégataire doit s'engager à ce que tous les contrats de services informatiques et de licence soient transférables à leur échéance au délégant ou du nouvel exploitant dès lors qu'ils sont nécessaires à l'exploitation du service

F. Personnes et entités interrogées



Entité	Nom du contact	Fonction du contact
Bordeaux Métropole	Christophe COLINET	Chargé de mission Smart Métropole
Bordeaux Métropole	Arnaud BURAY	Chef de service valorisation de la donnée
Métropole de Lyon	Nathalie VERNUS-PROST	Chief Data Officer
Métropole Nice-Côte d'Azur	Hervé PAUL	Président de la Commission eau, assainissement et énergie de la Métropole Maire de Saint-Martin-du-Var
Métropole Nice-Côte d'Azur	Aurélien BOZZA	Responsable Data, DSI de la Métropole
Nantes Métropole	Claudine BARON	Responsable informatique
Paris	Jean-Philippe CLEMENT	Chief Data Officer
Rennes Métropole	Bernadette KESSLER	Responsable du service innovation numérique, Rennes Métropole et Ville de Rennes

Toulouse Métropole	Sandrine MATON	Chef du service Administration des SI de la Mairie de Toulouse / Toulouse Métropole
CLCV - Consommation Logement et Cadre de Vie	Olivier GAYRAUD	Chargé de mission consommation
OpenData France	Jean-Marie BOURGOGNE	Délégué général
ARSENIC	Stéphane DELAHAYE	Délégué régional ARSENIC et Conseiller municipal à Martigues
Renaissance numérique	Henri ISAAC	Président de RN et VP Université Paris-Dauphine
Caisse des Dépôts	Cédric VERPEAUX	Responsable pôle « ville numérique »
DATA PUBLICA	François BANCILLON	CEO
La lyonnaise	Frédéric CHARLES	Stratégie & Gouvernance du SI - Data, Architecture & Relations Numériques chez SUEZ environnement
M2ocity	Edouard SAINT-GEOURS	Responsable des offres, Direction de marketing
OpenDataSoft	Mathieu CAPS	Chief Public Affairs Officer
Orange	Laurent LONDEIX	Délégué régional, Président du Sophia Club Entreprises, ex président du Pôle SCS
Ozwillo	Sophie HOUZET	Déléguée générale de l'association
Placemeter	Martin LAGACHE	Business Development Europe
ERDF	François BLANC	Directeur national du programme numérique d'ERDF
GRDF	Isabelle DROCHON	Responsable du programme "traitement et valorisation des données de consommation"
Orange	Christophe GUION	Directeur de Projets Innovants dans le Sud-Est
La Poste		
CADA	Bruno RICARD	Membre de la Commission
CEREMA Nantes	Laurent DAVID	Expert TIC

CNIL	Maurice RONAI	Commissaire Secteurs : NTIC, communications électroniques, innovations technologiques
CNIL	Delphine CARNEL	Juriste e-administration - Direction de la conformité
DGE	Cédric MORA	Chargé de mission Cloud Computing & Big Data
DINSIC/ Etalab	Henri VERDIER	Administrateur général des données
Etalab	Simon CHIGNARD	Data Editor du portail data.gouv.fr
IGN	Michel SEGARD	Directeur des programmes civils
IGN	Pascal LORY	Chargé de mission CNIG
INSEE	Françoise DUPONT	Directrice du projet Référentiel de métadonnées statistiques
Open Data Lab Midi-Pyrénées	Benoît CHABRIER	Chargé de mission Economie numérique auprès du Préfet de région
SIAF	Claire SIBILLE	Sous-directrice de la politique archivistique
Systematic	Philippe DESBATS	Responsable Programme Energie & Ville Numérique - CEA
e-Megalis	Hervé LE NORCY	Directeur Général
GIP ATGeRi (PIGMA)	Pierre MACE	Directeur
GIP e-Bourgogne	Patrick RUESTCHMANN	Directeur adjoint
Institut d'aménagement et d'urbanisme de la région Île-de-France	Daniel THEPIN	Chargé de mission TIC à l'IAU-ÎdF
PACA (CRIGE & Open PACA)	Christine ARCHIAS et Thomas BEKKERS	Directrice du CRIGE et Chef de projet Open Data CR PACA