

Données d'intérêt général massivement collectées par les géants d'internet, évolution rapide des usages mais aussi des normes et des règles pour la protection des données personnelles... Les enjeux numériques connaissent une accélération liée à la multiplication des initiatives publiques ou privées utilisant massivement nos données. Pour Villes et Territoires, Jacques Priol, Président du cabinet CIVITEO revient sur ces enjeux lors de la Journée Régionale de Réciprocités du 24 novembre 2023.

SOMMAIRE :

1. Le Big Data, c'est quoi ? p.1
 - 1.1 - L'exemple de WAZE p.1
 - 1.2 - Le Big Data et l'action publique p.1
 - 1.3 - La protection des données personnelles et la transparence des données publiques p.2

2. L'irruption de l'IA p.4
 - 2.1 - A quoi sert (ou pourrait servir) l'IA dans les territoires ? p.5
 - 2.2 - Des conditions supplémentaires pour une IA de confiance p.5

3. La donnée et l'habitat social. p.6

4. « Data et quartiers », un programme expérimental et pionnier (2019 - 2022). p.6
 - 4.1 - La fragilité des habitants des QPV face au Covid-19 p.7
 - 4.2 - L'accessibilité à l'emploi des habitants des QPV p.7
 - 4.3 - Une démarche alignée sur les politiques nationales p.8
 - 4.4 - Un rapport officiel important p.8
 - 4.5 - La circulaire du 31 août 2023 p.8
 - 4.6 - Le comité interministériel p.9

5. Contacts p.9



1. Le BIG DATA, c'est quoi ?

1.1 En guise d'introduction : l'exemple des données exploitées par l'application Waze ...



Waze est une application mobile de **navigation GPS gratuite** qui propose sur une cartographie instantanément modifiable par ses utilisateurs, permettant d'identifier en temps et en heure les itinéraires les plus fluides. Elle appartient à **Google**. Il s'agit d'un dispositif de « **crowdsourcing** » très performant.

Parallèlement au service rendu, cette application **collecte les données** des contributeurs usagers pour **vendre des espaces publicitaires**.

Waze dispose aussi du signalement des accidents en France, **4 min 30 avant les services de police et de secours**. Elle souhaite restituer cette donnée, moyennant finances.

« **Waze for cities** » est un service proposé aux collectivités locales. Il prend en compte les plans de circulation, les travaux, les événements, *etc.*

Ce service permet entre autres aux collectivités une circulation plus fluide. Pour waze, c'est un moyen de récolter leurs données.

Le cas de l'application Waze est emblématique car il témoigne de toutes les problématiques du Big Data :

- Collecte des données personnelles.
- Utilisation de ces données à des fins commerciales, de contrôle ou de surveillance.
- Le business modèle de crowdsourcing.
- La puissance et l'efficacité de ce modèle.
- La question de la loyauté des algorithmes.
- La question de la souveraineté publique sur des données d'intérêt général.
- *etc.*

1.2 Le BIG DATA et l'action publique

Le terme « **big data** » (« données massives » en français) a été popularisé dans les années 2010 avec l'accroissement de l'usage des nouvelles technologies. Il désigne la **masse des données numériques générées** par des capteurs, des objets connectés, des travaux de recherche, des productions professionnelles, institutionnelles ou personnelles, *etc.* Les données numériques visent tous les types d'informations qui peuvent circuler : texte, vidéo, audio, données contenues dans les bases de données (données électorales, fichiers usagers...), coordonnées géographiques, transactions, données issues de capteurs, *etc.*

« En 2020, le volume des données produites en une seule journée pour la gestion des villes européennes a été 4 fois supérieur à celui des données produites pour toute l'année 2015 ».

Les usages du Big Data pour le service public local peuvent être regroupés en deux types de familles : **les usages liés aux données territoriales**, comme la smart city : ces données régulent les flux, permettent des économies, et l'optimisation de la consommation d'eau ou d'énergie. Elles s'inscrivent dans une logique de développement et de croissance durables (la gestion connectée de l'espace public de Dijon Métropole par



exemple). La seconde famille d'usages concerne **les données propres aux individus, habitants, citoyens, usagers**. Elle vise à comprendre et anticiper leurs besoins, aspirations. Elle crée ou améliore des services, améliore l'efficacité par la prise en compte des situations individuelles et fait de la prévention grâce à la prédiction.

Si les services publics exploitent les données pour **améliorer les services** rendus aux citoyens, habitants, usagers, les services publics sont aussi **protecteurs des droits**, de la vie privée de ces mêmes citoyens, habitants, usagers. Ces derniers attendent de la part du service public des outils et méthodes qui diffèrent de ceux des enseignes commerciales.

1.3 La protection des données personnelles et la transparence des données publiques

L'exploitation des données massives du Big Data inquiète. Les services publics sont attendus pour **protéger les droits individuels et promouvoir un usage éthique de la donnée**. Les évolutions rapides doivent faire l'objet de stratégies pro-actives de la part des institutions et notamment des collectivités, tant sur la protection des données individuelles, mais également sur la transparence de l'utilisation de ces données.

Les enjeux sont multiples :

→ Des enjeux juridiques : obligations légales :

▪ **Le data Governance Act (2023)**

▪ **La loi pour une République numérique :**

Cette loi instaure l'Open data par défaut dans toutes les communes de plus de 3500 habitants disposant au moins de 50 agents temps plein. Mise en ligne de toutes les bases de données dont la collectivité dispose à l'exception des données personnelles et de sécurité.

Pour aller plus loin :

<https://www.economie.gouv.fr/republique-numerique>

<https://www.cnil.fr/fr/les-collectivites-territoriales-et-lopen-data-concilier-ouverture-des-donnees-et-protection-des>

▪ **Le Règlement Général sur la Protection des Données – RGPD :**

Depuis le 25 mai 2018, le RGPD s'applique à toute organisation, publique et privée, qui traite des données personnelles pour son compte ou non, dès lors :

- qu'elle est établie sur le territoire de l'Union européenne,
- ou que son activité cible directement des résidents européens.

Pour aller plus loin :

<https://www.cnil.fr/fr/RGPD-quel-impact-pour-les-collectivites-territoriales>

<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

Outre les solutions juridiques pour protéger les données personnelles et limiter leur utilisation, il est fondamental de faire prendre conscience aux usagers de leur propre rôle dans la protection des données : **le consentement** peut être la pierre angulaire du système juridique, mais **l'utilisateur doit être conscient** des données qu'il communique, et de son pouvoir de consentement.

Les collectivités sont face à un double enjeu : exploiter les données pour développer des services et rendre plus efficace leur action tout en agissant comme tiers de confiance. D'où l'importance pour les collectivités de se doter d'une stratégie, à l'exemple de la charte métropolitaine de la donnée de Nantes métropole.



Nantes Métropole a lancé en juin 2019 une « charte métropolitaine de la donnée » par laquelle les signataires, publics et privés, s'engagent à respecter des règles de collecte et d'utilisation des données des citoyens. Accéder à la charte sur le site internet de Nantes Métropole :

<https://metropole.nantes.fr/territoire-institutions/projet/ambitions-territoire/la-charte-de-la-donnee>

→ **Des enjeux managériaux :**

De nouveaux métiers :

- Chief data officer / administrateur (général) des données.
- DPO / délégué à la protection des données.
- Data Scientist (dataminer, data analyst...).
- Data architecte.
- Data manager.
- Data ingénieur.

→ **Des enjeux techniques :**

Le choix des infrastructures :

- Les différentes formes de stockage et de traitement des données.
- La cybersécurité.

→ **Des enjeux économique :**

- A qui appartiennent les données ?
- Faut-il acheter les données ?
- Faut-il échanger les données ?
- Quelles retombées pour le territoires ?
- Marketing territorial
- *etc.*

→ **Des enjeux environnementaux :**

La question de la sobriété numérique vs la sobriété DATA :

Selon « The Shift Project », l'utilisation des Data centers est responsable de 19% de la consommation énergétique du numérique en 2017.

→ **Des enjeux éthiques :**

- Quels usages ?
- Avec quels contrôles ?
- Quelle transparence ?
- Smart city vs safe city ?
- *etc.*

→ **Des enjeux démocratiques :**

Les questions de la transparence et de l'accès à la donnée.



2. L'irruption de l'IA

« Il n'y a pas de définition de l'intelligence artificielle, ce qui est certain c'est que l'intelligence artificielle ce n'est pas de l'intelligence. » Cédric Villani.

L'intelligence artificielle (IA) se définit en effet par « **un ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine.** »

On note des **systèmes variés d'intelligence artificielle** tels que l'apprentissage automatique, l'apprentissage profond, le traitement du langage, la vision par ordinateur. Aujourd'hui, ces apprentissages ont donné naissance aux **IA génératives** (ChatGPT, Mid Journey, Dall-E, Copilot, etc.).



Dans les territoires, l'IA se développe également :

- 2021 : 680 exemples d'IA au service de l'action public en Europe dont 40 en France et 5 à l'échelon local.
- 2022 : 19 collectivités françaises (14 métropoles et 5 régions) ont eu recours à l'IA.
- 2022 : Une trentaine de territoires, plus de 40 expérimentations, 3 ou 4 projets déployés en production.



2.1 A quoi sert (ou pourrait servir) l'IA dans les territoires ?

- **Gestion de l'espace public** : « Computer vision », analyse d'images ou de vidéos, émission d'alertes (sécurité, stationnement, dégradations, déchets sauvages...).
- **Relation usagers** : « Chatbots » et robots conversationnels, traitement du langage, compréhension et génération de contenu.
- **Évaluation des politiques publiques** : Utilisation d'outils d'IA pour faire des analyses documentaires préliminaires, identifier des périmètres d'investigation, analyser les données disponibles. L'IA peut être intégrée à des projets pour une évaluation continue.
- **Gestion du patrimoine et des ressources** : Analyse de données en temps réel (alertes) et modélisation prédictive (anticipation) pour la gestion des déchets, la gestion de l'eau (détection de fuites, alertes inondation, modélisation des submersions...), l'éclairage public, la restauration scolaire, l'optimisation des consommations énergétiques, *etc.*
- **Diagnostic territorial** : Analyse d'images pour produire de façon automatique des cartographies (2D/3D) et des indicateurs relatifs à l'occupation du sol et aux enjeux environnementaux (à partir de photos aériennes, satellites, données LIDAR, *etc.*).
Exemples : suivi de la consommation foncière (ZAN), de la végétalisation, identification des îlots de chaleur urbains, du potentiel EnR, *etc.*
- **Production administrative (interne)** : Utilisation d'outils d'IA « générative » pour faire des recherches, produire des analyses, des synthèses, des notes, des discours, des délibérations, *etc.*
- **Mobilité et transports** : Analyse de données en temps réel, modélisation prédictive et pilotage pour la gestion de la congestion (gestion des flux et feux de circulation en temps réel) et pour l'optimisation des infrastructures, du fonctionnement des transports en commun, de la mobilité partagée, du covoiturage, *etc.*

2.2 Des conditions supplémentaires pour une IA de confiance

- **La primauté humaine** : « Derrière la machine, il y a toujours l'homme ».
- **L'autonomie stratégique** : « L'IA publique doit préserver la souveraineté de la France ».
- **La soutenabilité environnementale** : « Il ne faut utiliser l'IA que si ses effets positifs égalent au moins son impact environnemental ».
- **La sécurité** : « Une IA de confiance intègre obligatoirement les enjeux de cybersécurité ».
- **La transparence** : « Une transparence complète : explicabilité du système utilisé, transparence sur les données et leur utilisation ».
- **L'équité et la non-discrimination** : « Prévenir tous les biais pour éviter les discriminations ».
- **La performance** : « Le service doit bien fonctionner »



3. La donnée et l'habitat social

L'IA et l'analyse des données offrent des possibilités considérables pour améliorer la gestion, l'efficacité, la durabilité et la qualité de vie dans l'habitat social. On pense notamment à la donnée pour récolter des informations en terme de :

- **Mixité sociale** :
exemple : <https://inequality.media.mit.edu/> , l'Atlas américain qui détecte les lieux où il y a le plus et le moins de mixité sociale).
- **Visualisation des îlots de chaleur urbains**, sujet très tendance en 2023.
exemple :
<https://ici.radio-canada.ca/info/2022/07/ilots-chaleur-villes-inegalites-injustice-changements-climatiques/>
- **Gestion du patrimoine** pour faciliter la maintenance immobilière.
Objectif : centraliser des données relatives au fonctionnement des immeubles pour en améliorer la maintenance.
<https://morgansindallpropertyservices.com/morgan-sindall-launches-innovative-machine-learning-software-platform/>
- **Action sociale** pour prévenir la précarité énergétique.

4. « Data et quartiers », un programme expérimental et pionnier (2019 - 2022).

Les quartiers Politique de la ville ne bénéficient pas des apports récents de l'usage de la donnée. Pourtant, un certain nombre d'expérimentations déjà appliquées à l'action publique pourraient apporter des éléments de réponses à des problématiques plus spécifiques vécues au sein des quartiers, si elles y étaient testées.

RésO Villes, accompagné par Civitéo et soutenu par l'ANCT, a expérimenté une approche « smart city » dans les quartiers prioritaires de la ville en Bretagne-Pays de la Loire.

→ <https://www.calameo.com/read/007030140656722f8bab6?view=book&page=5>

3 thématiques de travail :

- **Groupe de travail Emploi** : Etude sur la distance entre le lieu de vie des demandeurs d'emploi et la localisation des offres d'emplois qui leur correspondent.
- **Groupe de travail Santé** :
 - Les populations des QPV sont-elles plus fragiles face au Covid-19 ?
 - Le non-recours à la Complémentaire Santé Solidaire (C2S) est-il plus élevé dans les QPV que dans les autres quartiers ?
 - Et à l'intérieur des QPV, existe-t-il des profils de foyers ayant moins recours que d'autres à la C2S ?
- **Groupe de travail Mobilités** :
 - Quels sont les moyens de transports disponibles pour permettre aux habitants des QPV de se rendre sur leur lieu de travail ?
 - Et combien de temps représentent ces trajets ?



4.1 La fragilité des habitants des QPV face au Covid-19

« Toutes les zones fragiles ne sont pas des quartiers de la politique de la ville, mais tous les quartiers QPV sont dans des zones sensibles »

https://www.resovilles.com/wp-content/plugins/covid_map/

→ **La question de départ** : Sur la base de plusieurs études scientifiques internationales parues en 2020, est-il possible de vérifier dans les QPV de Bretagne et Pays de la Loire, si la vulnérabilité des populations face au Covid 19 y est plus importante qu'ailleurs ?

→ **Les données** :

- Des données socio-économiques et socio-démographiques (INSEE).
- Des données de santé (Health Data Hub ; ARS).

→ **Création d'un indicateur Santé** pour rechercher les zones de vulnérabilité face à la contamination à la Covid-19 (données transmises par l'Observatoire régional de santé de la Région Bretagne), basé sur :

- Le taux de population sous traitement antiasthmatique.
- Le taux de population sous traitement antidiabétique.
- Le taux de population souffrant d'une ou plusieurs affections de longue durée.

→ **Création d'un indicateur sociodémographique** pour démontrer que les zones de vulnérabilité correspondent aux quartiers « politique de la ville » (données en open data sur le site internet de l'INSEE), basé sur :

- Le taux de population vivant en logement de type « HLM ».
- Le taux de population bénéficiaire de la CMU-C.
- Le taux de population immigrée.

4.2 L'accessibilité à l'emploi des habitants des QPV

→ **La question de départ** : Quelle est la distance entre le lieu de vie des demandeurs d'emploi et la localisation des offres d'emplois qui correspondent à leurs attentes ?

→ **Les données** :

- Comment identifier les emplois recherchés par les habitants des QPV ? (Pôle emploi)
- Comment identifier toutes les offres d'emploi qui correspondent à ces attentes ? (Randstad)
- Comment automatiser le calcul de distance ? (Tomtom)

→ **Grâce à ces croisements inédits**, nous sommes en mesure d'analyser la réalité des distances « domicile – emploi potentiel » des chercheurs d'emploi de ce QPV :

- Visualisation cartographique de la répartition géographique des offres.
- Visualisation plus fine de la répartition géographique des offres par code ROME, sous forme d'histogrammes.
- Calcul d'une distance moyenne « domicile – emploi potentiel » pour chaque QPV.



4.3 Une démarche alignée sur les politiques nationales

Le programme « Data & Quartiers » est totalement aligné avec les orientations gouvernementales sur la gestion des données (circulaire du 27 avril 2021 relative à la politique des données, des algorithmes et de codes sources et les feuilles de route ministérielles qui en découlent).

La démarche a permis en effet de travailler, parfois avec difficulté mais avec un succès certain :

- ✓ Au développement de l'open data des administrations publiques.
- ✓ Au développement de la circulation des données entre administrations d'Etat et collectivités locales.
- ✓ A l'accès à des données privées d'intérêt général dans des conditions inédites.
- ✓ A la transparence des usages et des algorithmes dans un cadre éthique lui aussi inédit car le programme « Data & Quartiers » est le premier programme en France à avoir obtenu le label EKITIA.

4.4 Un rapport officiel important

« Il est plus que temps de lever le voile d'ignorance pesant sur la connaissance des quartiers en faisant de la data science un outil de la politique de la ville (...). »

L'expérimentation « Data & Quartiers » menée depuis 2019 par Réso Ville et CIVITEO sur 78 quartiers prioritaires de Bretagne et Pays de la Loire a permis d'exploiter des données publiques et des données privées d'intérêt général pour mieux observer les territoires et aiguiller l'action collective.

« Une méthodologie rigoureuse généralisable pour résoudre bien des enjeux des quartiers en dotant la politique de la ville des outils du XXIe siècle ».

Il en ressort deux recommandations innovantes sur la gestion des données :

- ✓ Introduire des « clauses data » dans les futurs contrats de ville.
- ✓ Généraliser « Data & Quartiers ».

4.5 La circulaire du 31 août 2023

« L'évaluation et l'impact des actions financées devront être analysés de manière quantitative mais également qualitative, ce qui pourra donner lieu à l'intégration de clauses relatives à la production, la mise à disposition et l'accès aux données des cosignataires des contrats de villes. Ces données permettront plus généralement une analyse approfondie de l'évolution de la situation socio-économique des quartiers et de leurs habitants. »

→ <https://www.legifrance.gouv.fr/download/pdf/circ?id=45476>



4.6 Le comité interministériel

« Faire de la data science un outil de connaissance territoriale et de territorialisation des politiques publiques. » p.33

→ <https://villes-et-territoires.fr/retour-sur-le-comite-interministeriel-des-villes-%f0%9f%93%a2/>

5. Contacts

Pour tout renseignement :

→ jacques.priol@civiteo.fr

06 74 52 69 23

→ romain.vallee@civiteo.fr

07 63 42 71 17

→ cecile.nonin@villes-et-territoires.fr