



Réseau
Villes

DATA & QUARTIERS : le livre blanc

Retour sur trois ans d'expérimentation



agence nationale
de la cohésion
des territoires






DATA & QUARTIERS : le livre blanc

Retour sur trois ans d'expérimentation

Sommaire

Édito	7
Introduction	10
➔ PARTIE 1 : Le programme « Data & Quartiers »	17
À l'origine...	12
Les objectifs initiaux du programme « Data & Quartiers »	14
Objectif n°1 : utiliser des outils de data science et de type « smart city » dans les quartiers	14
Objectif n°2 : améliorer la connaissance et l'observation des quartiers grâce à la data	14
Objectif n°3 : documenter la méthodologie pour favoriser la duplication des expérimentations	14
Objectif n°4 : expliquer ces nouveaux outils, ces nouvelles données, ces nouveaux usages, ces nouveaux enjeux	15
Les partenaires du programme « Data & Quartiers »	16
La Méthodologie « Data & Quartiers »	17
Étape n°1 : sensibiliser	17
Étape n°2 : cadre les besoins et définir les « cas d'usage » de la donnée	17
Formuler des objectifs	17
Identifier les données utiles	17
Étape n°3 : valider le cas d'usage / la preuve de concept	18
Accéder aux données	18
Préparer et explorer les données	18
Modéliser	18
Évaluer le modèle	19
Étape n°4 : prototyper la solution	19
Étape n°5 : mettre en production	19
Synthèse	20

➔ PARTIE 2 : Les prototypes « Data & Quartiers »	21
Data & Emploi dans les quartiers : une étude sur les distances domicile-travail	22
Le cadrage du projet	22
Les étapes de travail	23
L'obtention des données	23
La preuve de concept	24
La mise à l'échelle	25
Le prototype data	26
Data & Santé dans les quartiers : une étude sur la fragilité des habitants face au Covid-19	29
Le cadrage du projet	29
Les étapes de travail	30
L'obtention des données	30
La réalisation cartographique et le prototype data	31
Data & Santé dans les quartiers : une étude sur le non-recours à la Complémentaire Santé Solidaire	32
Le cadrage du projet	32
Les étapes de travail (Phase 1 : estimation du non-recours)	33
L'obtention des données	33
La preuve de concept	33
Le prototype Data	35
Les étapes de travail (Phase 2 : modélisation des profils)	35
L'obtention des données	35
La preuve de concept	37
Data & Mobilités dans les quartiers : une étude sur les modalités de déplacement Domicile-Emploi	41
Le cadrage du projet	41
Les étapes de travail	41
L'obtention des données	41
La preuve de concept	41
Le prototype Data	42
Conclusion : politique de la ville, éthique et data	43
Généraliser « Data & Quartiers » ?	43
Des enjeux nombreux	43
Une indispensable éthique de la donnée	44
Une démarche reconnue et labellisée	44

	PARTIE 3 : Paroles d'acteurs	45
	Aminata Keita et Christophe Girard - ANCT	46
	Jean-Aimé Rakotoarisoa - Conseiller citoyen	48
	Benoît Blondet - Harmonie Mutuelle	49
	Francky Trichet - Nantes Métropole	50
	Émilie Sarrazin - Rés0 Villes	52
	Patrick Vanoli - Randstad France	53
	Louis-Vincent Potura - Valoway	54
	Cécile Nonin - Villes et Territoires Occitanie	55
	Renaud Epstein - Sociologue	56
	ANNEXES	58
	Annexe 1. Exemple de calcul des « unités de consommation » et du « plafond de ressources » pour l'IRIS GRASLIN COMMERCE (Nantes)	59
	Annexe 2. QPV et IRIS : à la recherche de données introuvables	61
	REMERCIEMENTS	66



**Près de 230 000 habitants
dans les deux Régions
vivent dans les
78 quartiers prioritaires
répartis sur 32 communes**



Édito

Lorsqu'en 2019 RésO Villes lance le programme « Data & Quartiers », c'est avec l'intuition forte que les quartiers de la politique de la ville ne doivent pas être les grands oubliés des « villes intelligentes ». Si ce terme est sans doute un peu « fourre-tout », évoquant la gestion des éclairages publics comme la gestion des flottes de trottinettes en passant par des solutions d'affichage dynamique, il caractérise surtout la gestion des villes avec de la donnée. Et nous avons pu, pendant 3 ans, expérimenter une approche *smart city* pour nos quartiers.

Soutenue par l'ANCT et accompagnée par les experts de CIVITEO, notre association a pris le temps : le temps de s'acculturer à ces sujets, le temps de trouver et de convaincre des partenaires, le temps de choisir des thématiques utiles, le temps de sélectionner et de dialoguer avec des territoires. Un rythme inhabituel peut-être, car les projets « numériques » sont souvent synonymes de rapidité, voire parfois de précipitation. Du temps, nous en avons eu besoin pour poser les bonnes questions, lever les doutes, constater des erreurs et affiner les certitudes. Nous avons mis à profit ce temps pour créer des passerelles solides, parfois insolites, entre deux univers qui ne se connaissaient pas : celui de la politique de la ville et celui de la data.

Nous avons trois objectifs.

Expérimenter ces outils pour mieux observer les quartiers prioritaires de Bretagne et des Pays de la Loire. La politique de la ville est une politique sensible à l'observation, à l'évaluation, à la prise en compte des données, notamment statistiques. Nous avons souhaité aller plus loin en utilisant des méthodes et des données encore non exploitées ou peu exploitées par la politique de la ville.

Sensibiliser les acteurs publics, privés mais aussi les citoyennes et les citoyens à l'utilisation de la *data science* au profit des territoires les plus fragiles. Nous faisons le pari que ces nouveaux outils sont des alliés précieux pour la prise de décision et les actions qui en découlent. Cette phase de sensibilisation est une étape essentielle : il faut faire disparaître les peurs et dissiper les doutes autour de ces sujets tout en mettant tous les partenaires au même niveau de connaissance.

Offrir la possibilité de dupliquer les expérimentations mises en œuvre dans le cadre de ce programme aussi expérimental qu'ambitieux. Nous avons souhaité documenter l'intégralité de la méthodologie sans omettre aucun point. C'est l'objet de ce livre blanc. Utilisez-le comme un guide pratique pour mettre en place vos propres outils. Si ce document peut vous faire gagner du temps, dans la prise en compte de ce sujet ou dans la mise en place à l'échelle de votre territoire, alors nous aurons atteint cet objectif.

Ces 3 années d'expérimentation ont aussi eu un impact sur RésO Villes. Si nous ne devons retenir qu'une seule chose de cette expérimentation, c'est que le sujet de la donnée n'est pas une problématique technologique réservée à quelques experts. Bien au contraire. C'est une question beaucoup plus large qui doit être traitée avec exigence.

Une exigence politique car les nouveaux usages de la donnée permettent aux décideurs d'accompagner leurs choix de manière plus objective. Les données collectées puis retravaillées rendent compte de la réalité de la vie des habitants de façon inédite. Grâce aux données, nous pouvons mettre le doigt sur des problèmes propres à chaque quartier et proposer des actions concrètes aux résultats mesurables. Les quartiers ne forment pas un tout homogène. Un quartier connaît des difficultés liées à l'emploi pour des raisons de formation quand un autre a d'abord un problème d'accès aux transports.

Une exigence démocratique car manipuler massivement des données n'est pas anodin. Ces méthodes obligent à questionner la gouvernance de la donnée. Qui dispose des données ? Comment sont-elles collectées ? Qui peut y accéder ? Avec quelles contraintes ? Avec quelle éthique ? Nous avons constaté de nombreuses inégalités entre territoires. Il faut mettre les citoyens au cœur de ces questions sensibles, c'est notre plus grand défi, celui de la confiance.

Une exigence économique car au-delà des innovations possibles dans un cadre expérimental, et sans nécessairement chercher un *business model*, il faut imaginer comment généraliser l'usage de ces outils, avec des financements, un modèle de gestion, des règles d'accès et de traitement des données.

Aujourd'hui RésO Villes est en pleine mutation. Ce programme nous a fait comprendre que la transformation de nos pratiques professionnelles était nécessaire. Nous devons adapter nos réponses au plus près des réalités objectives et mieux comprendre les dynamiques territoriales en cours : la data est un outil formidable qui, j'en suis sûre, ouvre la voie à une meilleure compréhension du monde qui nous entoure.

Un dernier mot enfin pour remercier les équipes de RésO Villes qui ont su s'emparer du sujet et qui n'ont pas hésité à se remettre en cause lorsque c'était nécessaire. Je remercie également les élus et membres de l'association qui ont soutenu ce projet dans cette période particulière. Je remercie enfin nos partenaires, publics ou privés, qui nous ont accompagnés. La place manque pour remercier chacune et chacun, mais sachez que je suis très heureuse d'avoir pu mener à bien cette expérimentation avec vous.

J'espère que la lecture de ce livre blanc vous permettra de trouver l'inspiration pour démarrer, vous aussi, votre projet « data ».

Bonne lecture.



Nathalie Gosselin,
Présidente de RésO Villes



Introduction

Ce livre blanc a plusieurs objectifs.

Il a été imaginé dès le début du programme « Data & Quartiers » comme un récit retraçant les enjeux, les méthodes, les découvertes, les difficultés, les obstacles, les solutions, les réussites et les échecs de ce programme atypique de gestion expérimentale des données.

Pour comprendre la manière dont il a été rédigé, il faut remonter presque 4 ans en arrière lorsque les premières discussions ont été engagées entre l'association RésO Villes et l'Agence nationale pour la cohésion des territoires (ANCT). Le sujet de la data dans l'action publique n'était pas du tout identifié comme il l'est aujourd'hui. L'*open data* venait tout juste d'être reconnu par la loi, le RGPD entrait à peine en vigueur, les méthodes de la *data science*, les prototypes data n'étaient pratiqués que par quelques grandes métropoles. **Il fallait documenter les travaux et garder la trace du projet.**

L'idée des dirigeants de RésO Villes était audacieuse ! Elle bousculait les habitudes d'un univers, celui de la Politique de la Ville, qui avait fait de la donnée statistique sa boussole. N'était-ce pas en fonction de données précises que des quartiers étaient classés (ou non) comme prioritaires ?

Partie prenante des discussions dès l'origine, exerçant ensuite un rôle dans la conception du programme et son animation, l'équipe du cabinet CIVITEO a été à la fois actrice et spectatrice privilégiée du déroulement du programme. **En prenant la plume et en accompagnant la production de ce livre blanc nous avons voulu témoigner de ce qu'est concrètement un projet autour de la donnée.**

Il faut de la méthode, des données (bien sûr !), des experts, des outils (souvent moins sophistiqués que ce que l'on croit) mais il faut surtout connaître ce dont on parle, et pour ce programme, il faut connaître les quartiers !

Ce livre blanc raconte aussi des anecdotes. Elles sont utiles ! Surtout lorsqu'il s'agit d'expliquer des difficultés, voire des échecs. Il n'y a rien de « magique » avec la donnée et avec des algorithmes. Il n'y a que des outils qui permettent d'accéder à des informations qui doivent être utiles aux acteurs des territoires : aux élus, aux services des villes, aux conseillers citoyens.

Ce livre blanc s'adresse aussi aux décideurs nationaux. Il montre, à travers de simples prototypes, que **la Politique de la ville a tout à gagner à se saisir de ces outils nouveaux.** Les indicateurs qui président au classement des quartiers ont sans doute leur utilité au moment de la contractualisation. Mais ils ne décrivent que médiocrement la réalité de la vie des habitants, les spécificités de chaque quartier et leurs enjeux. Ce livre blanc expose à plusieurs reprises une difficulté qui suscite à vrai dire de la part de ses auteurs une forte incompréhension : **la plupart des données nécessaires à nos travaux (sur l'emploi, la mobilité, la santé, le social...) ne sont pas disponibles à l'échelle des quartiers.** Pire, aucun outil public (idéalement disponible en *open data*) ne permet de reconstituer ces données par exemple à partir des unités statistiques de base que sont les IRIS. **Comme si, une fois le classement effectué pour quelques années, les décideurs publics se privaient de toute capacité d'observation...**

Puissent ces quelques prototypes servir une prise de conscience car, dans les quartiers comme ailleurs, la donnée lorsqu'elle existe, peut être mise au service de l'intérêt général et au service des politiques publiques.

Gageons que ce livre blanc contribue à le démontrer.



Jacques Priol

**Président du cabinet CIVITEO
Président de l'Observatoire
Data Publica**

*Tout au long de ces 4 années aux côtés des équipes de RésO Villes le cabinet CIVITEO a joué plusieurs rôles. Tantôt formateurs, animateurs, négociateurs ou chefs de projet, **Jordan Esnault** et **Romain Vallée** ont consacré beaucoup de temps à « Data & Quartiers » et à la production de ce livre blanc. Merci à eux.*

Le programme « Data & Quartiers »



PARTIE 1



À l'origine...

En 2018, plus de 600 personnes ont été réunies à Rennes pour la deuxième édition du forum Cités Cap', un événement organisé par RésO Villes, Centre de ressources de la politique de la Ville de Bretagne et des Pays de la Loire.

Au cœur de cet événement dédié au développement économique dans les quartiers, des élus locaux, des cadres associatifs ou des représentants des habitants ont été invités à prendre part à **un atelier intitulé « Le big data au service des quartiers »**. Plusieurs prototypes avec des usages innovants de la donnée y ont été présentés, notamment un outil algorithmique de modélisation prédictive des parcours des bénéficiaires du RSA testé par le Conseil départemental du Finistère et un outil d'analyse du marché de l'emploi développé par le Groupe Randstad.

Les experts de la politique de la ville et les acteurs du terrain présents ont encouragé l'idée d'une expérimentation inscrite dans la durée et permettant aux professionnels comme aux habitants intéressés de s'initier aux nouveaux usages et aux nouveaux outils de

la donnée. Le Conseil d'administration de l'association RésO Villes, présidé alors par Isabelle Melscoët, adjointe au Maire de Brest, s'est emparé du sujet, posant ainsi les bases du programme qui allait s'appeler « Data & Quartiers ».

Mais encore fallait-il savoir de quoi il en retourne. De quelles données parle-t-on ? Quelles sont les données disponibles ? Comment sont-elles collectées ? Qu'est-ce qu'un algorithme ? **Un temps d'acculturation aux enjeux de la donnée a été organisé à l'attention des membres du Conseil d'administration de RésO Villes** (composé principalement des élus à la politique de la ville des collectivités adhérentes). Une matinée complète a permis de détailler des exemples d'expérimentations ayant recours à la *data science* au service de l'action publique locale. Elle a aussi permis d'expliquer le cadre juridique de ces expérimentations, notamment la diffusion progressive de l'*open data*.

Très vite, des idées ont fusé. **Les élus ont exprimé des attentes et des besoins d'informations et de connaissances sur les quartiers**. Et il apparaissait



Forum Cité Cap, Rennes, juin 2018.

que, sans doute, ces nouveaux outils et ces nouvelles données pourraient y répondre. Les premières thématiques de travail étaient identifiées : emploi, mobilité, santé, loisirs, culture...

Pour aller plus loin un consultant spécialiste des sujets data au service des politiques publiques locales, a conduit **une trentaine d'entretiens avec des experts de la politique de la ville, les équipes de RésO Villes et des acteurs des quartiers** de Bretagne et des Pays de la Loire. Cette phase d'amorçage du programme a été étalée sur trois mois. Elle a aussi servi à expliquer le projet « Data & Quartiers » et à **identifier des partenaires publics et privés** intéressés par cette démarche visant à mettre la donnée, quel que soit son statut initial (donnée publique ou privée, donnée publiée ou non, anonymisée ou non...) au service de l'intérêt général. Avec le soutien de l'**Agence Nationale pour la Cohésion des Territoires**, très tôt mobilisée aux côtés de RésO Villes, **le programme a été officiellement lancé en avril 2019** par une première réunion d'un groupe de travail sur l'emploi.

Avant de détailler le déroulé du programme et d'en présenter les résultats, il est utile de revenir quelques instants fin 2018 pour en rappeler les objectifs. **L'année 2018 est une année clef pour les acteurs publics en matière de gestion des données.** En mai, **le Règlement général européen sur la protection des données (RGPD) est promulgué.** Il conforte les règles en vigueur en France depuis 1978 avec la loi Informatique et libertés. Bien sûr, le programme « Data & Quartiers » veillera strictement au respect des données personnelles des habitants. Ce point sera détaillé par ailleurs. En octobre 2018, l'une des principales dispositions de **la loi pour une République numérique de 2016¹ entre aussi en application** : il s'agit des obligations faites aux acteurs publics de publier leurs données. Cet **open data obligatoire** va progressivement rendre accessibles de très nombreuses données utiles à la compréhension de la vie des quartiers.

En 2018, **la mode des smart cities** est aussi en plein développement. Dans toutes les métropoles de France, on expérimente des nouveaux systèmes dits « intelligents » pour réduire la consommation électrique ou la consommation d'eau, pour collecter les déchets, pour faciliter la mobilité. Mais ces prototypes qui sont l'objet d'intenses campagnes de communication et de promotion de la part des entreprises, des *start up*, mais aussi des territoires, ne concernent pas les quartiers.



Atelier sur le *big data* lors du forum Cités Cap.

¹ LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique, dite « loi Lemaire ».

Les objectifs initiaux du programme « Data & Quartiers »

[+] Objectif n°1 : utiliser des outils de *data science* et de type « *smart city* » dans les quartiers

La ville intelligente ou *smart city* vise à améliorer la qualité de vie des habitants en créant de nouveaux services et en rendant les villes plus « efficaces ». De nouvelles technologies se diffusent. Elles reposent sur l'utilisation de capteurs et l'usage massif de la donnée. La *smart city* trouve des applications dans de nombreux domaines : les infrastructures publiques (bâtiments, mobiliers urbains, domotique, etc.), les réseaux (avec la gestion et l'optimisation des flux d'eau, d'électricité, de gaz, etc.), les transports (transports publics, routes et voitures intelligentes, co-voiturage, mobilités douces) ou encore la e-administration (facilitation des démarches administratives grâce à la dématérialisation, établissement d'un lien continu entre les services de la ville et les habitants via des applications, etc.). Si l'expression *smart city* désignait au départ uniquement les villes, on utilise désormais le terme pour d'autres territoires, y compris non urbains.

En 2018, les Quartiers Prioritaires ne font pas partie des terrains d'expérimentations de ces nouveaux usages. RésO Villes a fait le choix d'engager une réflexion sur l'utilité des outils de la *data science* pour résoudre des problématiques spécifiques des quartiers. **L'enjeu central du programme « Data & Quartiers » va être de mobiliser des données disponibles à l'échelle des villes**, quelle qu'en soit la source, pour proposer de nouvelles méthodes d'observation des quartiers. **Les outils utilisés dans des quartiers « intelligents » des centres villes arrivent dans les quartiers périphériques.**

[+] Objectif n°2 : améliorer la connaissance et l'observation des quartiers grâce à la data

L'idée n'est pas de remplacer les analyses traditionnelles ou les études des différents observatoires qui interviennent sur les périmètres des quartiers prioritaires, mais de les enrichir via l'angle de vue inexploré de la *data science*.

Les entretiens conduits auprès d'experts de la politique de la ville et les ateliers menés auprès des partenaires des différents groupes de travail ont rapidement mis en évidence une situation qui peut se résumer ainsi : **il existe des indicateurs de suivi des quartiers**, notamment auprès de l'INSEE et de l'Observatoire national de la politique de la ville. Ces indicateurs permettent de **suivre des statistiques utiles** : démographie, emploi, pauvreté... **Mais cette approche ne permet pas de connaître et de comprendre la vie quotidienne des habitants.** Et pourtant des données existent, elles sont même de plus en plus nombreuses.

Le programme « Data & Quartiers » va donc expérimenter un travail d'identification de données, et de traitement à l'échelle des quartiers, pour **décrire une réalité quotidienne en complément des statistiques traditionnelles.**

[+] Objectif n°3 : documenter la méthodologie pour favoriser la duplication des expérimentations

« Data & Quartiers » est un programme expérimental et, à ce titre, il mérite d'être documenté. En 2018, les projets recourant à la *data science* ne sont pas très fréquents chez les acteurs publics, de surcroît à l'échelle de deux régions et dans une démarche partenariale associant des acteurs privés. Pour ces raisons, les partenaires de RésO Villes (voir ci-après) ont très vite insisté sur la nécessité de documenter les processus. Leur engagement dans le programme visait autant à comprendre (et à apprendre) la méthode qu'à accéder à des résultats inédits.

L'ensemble des expérimentations du programme « Data & Quartiers » et le travail menés dans différents groupes thématiques (santé, emploi et mobilité) **sont présentés dans le présent livre blanc.** Chaque étape est décrite de façon précise. Nous avons essayé de ne rien omettre, y compris les erreurs commises, les obstacles rencontrés, les voies sans issues, qui seront probablement utiles au lecteur qui ambitionne de mettre en œuvre des projets « pilotés par la donnée » dans d'autres territoires et au service d'autres quartiers.

Tous les acteurs du programme ont été sensibilisés à cette démarche, y compris les prestataires réalisant des travaux de data science. Il leur a constamment été demandé de documenter non seulement les méthodes et les outils retenus, mais aussi ceux qui ont été testés et laissés de côté. Précision importante : RésO Villes a fait le choix de publier ce document sous licence Creative Commons CC BY-NC 3.0² c'est-à-dire que tous les travaux présentés dans ce document peuvent être enrichis ou intégralement reproduits par d'autres territoires ou dans d'autres centres de ressources, à condition d'en citer la source et à l'exclusion de tout usage commercial. En revanche, toutes les données utilisées pour ce programme ne sont pas disponibles, et ne peuvent pas l'être pour des raisons qui seront détaillées plus loin. Certaines données sont certes des données publiques, et nous avons exploité largement l'open data. Mais d'autres données ont été partagées par des partenaires de façon restreinte. Si elles présentent à l'évidence un caractère d'intérêt général, elles ne sont pas publiques.

Objectif n°4 : expliquer ces nouveaux outils, ces nouvelles données, ces nouveaux usages, ces nouveaux enjeux

Outre le travail de documentation des expérimentations, « Data & Quartiers » avait un objectif plus large d'acculturation et de sensibilisation des différents acteurs du secteur public en général, et de la politique de la ville en particulier, sur les sujets data.

À ce titre, diverses actions ont été menées auprès de publics variés ayant des niveaux de maturité numérique et data très disparates. Le récit du programme et le détail des expérimentations deviennent alors des supports concrets pour expliquer les enjeux et défis de la gestion des données dans notre société, et au service de l'intérêt général. Entre 2019 et 2022, de nombreux événements ont été organisés :

- Deux conférences (en format webinaire) à l'attention des conseillers citoyens de Bretagne et des Pays de la Loire.
- Une conférence à l'attention des responsables politique de la ville et/ou numériques des collectivités partenaires
- Deux conférences ouvertes à des acteurs territoriaux de la France entière organisées par l'Agence Nationale de la Cohésion des Territoires et le Centre National de la Fonction Publique Territoriale
- Une conférence grand public à l'initiative du Breizh Data Club
- Des retours d'expérience auprès de différents centres de ressources politiques de la ville (Occitanie, PACA, Centre Val-de-Loire...)
- Différents communiqués et des articles repris dans la presse (notamment deux articles sur la santé et l'emploi dans les quartiers publiés dans les pages régionales de Ouest-France)
- Un article expert dans le dossier spécial « Big data dans les territoires » de la revue Horizons Publics
- ...



² <https://creativecommons.org/licenses/by-nc/3.0/fr/>

Inventés le plus souvent à d'autres fins, notamment commerciales, ces outils de modélisation peuvent s'avérer des alliés précieux pour les agents publics dont le métier consiste à observer et comprendre les territoires avant d'y déployer des politiques publiques.

Exemples de publications mettant en avant le programme «Data & Quartiers»

LES PARTENAIRES DU PROGRAMME

« Data & Quartiers »

Le programme a été conçu, animé et porté par RésO Villes et l'équipe de CIVITEO, avec le soutien constant, bienveillant et exigeant, de l'ANCT.

RésO Villes

RésO Villes est un centre ressources de la politique de la ville au statut d'association loi 1901. Il rassemble 32 communes des régions Bretagne et Pays de la Loire, représentant 78 quartiers prioritaires. L'association accompagne les expertises locales en matière de politique de la ville et contribue au débat public à travers une réflexion politique, économique et sociale sur l'avenir des quartiers. Son Conseil d'administration est constitué d'élus des collectivités adhérentes. RésO Villes pilote depuis 2019 le programme expérimental « Data & Quartiers ».

CIVITEO

CIVITEO est un cabinet nantais qui développe depuis 2016 une expertise sur la gestion des données publiques. Ses consultants interviennent auprès des acteurs des territoires, collectivités locales et administrations, mais aussi de nombreuses entreprises qui agissent dans le cadre de délégations ou qui ont un rôle dans le développement et l'aménagement des territoires, afin de les conseiller autour des problématiques liées aux nouveaux usages de la donnée. CIVITEO a assuré la direction scientifique du programme et accompagné sa gestion au quotidien, encadrant notamment les travaux de l'entreprise de *data science* Valoway.



L'Agence nationale de la cohésion des territoires (ANCT) est un établissement public ayant pour mission de conseiller et d'accompagner les collectivités territoriales et leurs groupements dans la conception, la définition et la mise en œuvre de leurs projets. L'ANCT constitue un guichet unique pour les collectivités territoriales afin de leur permettre de bénéficier d'un appui technique et/ou financier de la part de l'Etat. L'ANCT est le principal financeur du programme « Data & Quartiers » et a assuré un suivi constant tout au long des différents projets.

De nombreux partenaires ont contribué au programme « Data & Quartiers ».

Tout d'abord les collectivités adhérentes qui ont mis à disposition leur « terrain de jeu », mobilisé leurs équipes et participé au financement du projet :



Des partenaires publics et privés ont apporté des contributions significatives au programme en mettant à disposition des ressources essentielles : des données bien sûr, mais aussi du temps, des experts, des équipes de *data science*, de la méthodologie...

- Agence régionale de santé des Pays de la Loire
- Caisse nationale de l'assurance maladie
- Groupe Randstad France
- Harmonie Mutuelle
- Observatoire régional de la santé de Bretagne
- Pôle emploi Bretagne
- Pôle emploi Pays de la Loire
- Poligma
- Université de Rennes 2



La Méthodologie

« Data & Quartiers »

Accéder à des données nouvelles, les stocker, les trier, les croiser... tous les acteurs de l'action publique le savent, **cela ne s'improvise pas**. A fortiori lorsque l'on souhaite expérimenter des méthodes nouvelles qui vont compléter des outils statistiques traditionnels par des outils algorithmiques susceptibles de modéliser la manière dont vivent les quartiers.

Dès le départ, le programme « Data & Quartiers » a été doté d'une méthodologie sous la forme d'un **protocole d'expérimentation**. Ce protocole est issu de travaux classiques de la *data science*.

✓ Étape n°1 : sensibiliser

L'étape de sensibilisation est primordiale.

Au démarrage de chaque projet, nous avons pris le temps d'**acculturer les partenaires aux enjeux de la gestion des données**. On parle aussi de **data littératie** et de *data mindset*.

Collecter massivement des données ne s'improvise pas. Il y a des **enjeux techniques**. Mais aussi des **enjeux juridiques et éthiques** (notamment la protection des données personnelles). Voire des **enjeux démocratiques et politiques**. Les acteurs de la politique de la ville ont en outre, à juste titre, **une posture prudente concernant la manipulation de certaines données qui peuvent s'avérer sensibles** : des données sociales, des données de santé, des données sur les origines ethniques, des données sur la délinquance...

L'étape de sensibilisation qui a été proposée à tous les acteurs du programme avait deux objectifs. Un objectif général d'acculturation : la donnée est aujourd'hui au cœur de tous nos actes personnels, en tant que consommateurs, usagers mais aussi citoyens. Un objectif ciblé : définir ensemble les conditions du partage et de la mise à disposition de données en travaillant en commun les finalités des expérimentations à venir. **La sensibilisation dissipe les craintes éventuelles mais permet aussi de fixer les limites !**

L'acculturation peut se faire de différentes façons : une formation « théorique » présentant les enjeux, des ateliers de mise en situation des participants ou des séquences en ligne de type webinaire. Pour « Data & Quartiers » nous avons tout essayé !

Evidemment, bien que nous ayons décidé de la présenter comme « étape 1 », la sensibilisation est un effort continu, elle se poursuit tout au long du projet.

✓ Étape n°2 : cadrer les besoins et définir les « cas d'usage » de la donnée

Formuler des objectifs

La deuxième étape vise à définir les objectifs d'une expérimentation. Tous nos travaux ont débuté par **une séance d'idéation** durant laquelle les participants ont proposé puis affiné de la manière la plus précise possible des objectifs à atteindre. Nous déciderons ainsi de chercher les données qui nous permettront de connaître *le temps moyen de trajet entre le domicile et le lieu de travail des habitants des quartiers*, plutôt que de travailler sur un objectif trop général et trop vague « d'amélioration de la connaissance des habitudes de déplacements des habitants ».

Pour **formuler des objectifs précis**, une méthode de travail a été proposée par le cabinet CIVITEO. Pour chaque sujet, quel que soit le thème, nous avons essayé de suivre un cheminement rigoureux.

- 1. Partir d'un constat ou d'une réalité.** Nous débutons d'une photographie de la réalité éprouvée dans les quartiers prioritaires. Elle est fondée sur la connaissance du terrain des experts de la politique de la ville. À ce stade, il n'est pas encore question de data...
- 2. Formuler une hypothèse.** Une hypothèse est formulée pour améliorer la situation décrite préalablement. On commence à aborder la donnée comme un levier de résolution des difficultés rencontrées. On formule par exemple l'hypothèse que des données existent pour améliorer la connaissance des habitudes des habitants des quartiers (des données pour appréhender les difficultés d'accès à l'emploi, pour comprendre la prévalence de problématiques de santé, pour connaître les habitudes d'alimentation, pour connaître les pratiques culturelles...).
- 3. Poser une problématique.** La problématique prend la forme d'une question précise. C'est l'objet des travaux. Concise et directe, elle décrit les ambitions fixées par le groupe de travail. Par exemple : les habitants des quartiers sont-ils plus exposés que d'autres au Covid-19 ?
- 4. Définir des objectifs.** Les objectifs constituent la feuille de route de chaque groupe de travail : identifier des données, créer des conditions pour y accéder (s'il ne s'agit pas d'*open data*), les collecter, les croiser, produire des datavisualisations...

Identifier les données utiles

Chacun des projets du programme s'est appuyé sur une ou plusieurs séances d'exploration de données, que l'on appelle aussi le *datamining*.

Cet exercice est bien connu des experts de la donnée, mais il est peu utilisé pour les politiques publiques locales, et encore moins à l'échelle des quartiers. Organisés sur une demi-journée, nos travaux d'exploration ont fait appel à l'imagination des participants. Les sources de données potentielles sont infinies... **Certaines données sont en *open data*, mais la plupart ne le sont pas.** Au fil des groupes de travail, nous avons envisagé recueillir des **données produites par des entreprises des territoires**, qu'elles agissent pour le compte de missions de service public (les sociétés de transport par exemple) ou pas : les données du groupe Randstad pour l'emploi, les données produites par les pharmacies ou les médecins des quartiers pour la santé, les données des restaurants pour l'alimentation, sans oublier **les données produites par les habitants eux-mêmes**, notamment à travers leurs téléphones mobiles.

Ces données sont **parfois librement accessibles, parfois payantes**. Parfois autorisées, mais parfois très protégées, et c'est tant mieux ! À chaque étape, les partenaires de « Data & Quartiers » ont eu à se pencher sur **les règles d'accès aux données**. Et parfois à se fixer eux-mêmes des **limites éthiques**. Ce n'est pas parce que quelque chose est autorisé (par exemple acheter des données de géolocalisation des habitants à partir des téléphones portables) qu'il faut le faire !

✓ Étape n°3 : valider le cas d'usage / la preuve de concept

Accéder aux données

Au fil des mois, l'équipe de « Data & Quartiers » a développé une certaine expertise pour organiser l'accès aux données. Il y a d'un côté les données librement accessibles, et notamment les données en *open data* mises à disposition par l'INSEE ou publiées par les communes. Mais il y a aussi d'autres données publiques non accessibles. Il s'agit de données produites par des acteurs publics, mais non rendues publiques. Parfois pour de bonnes raisons (par exemple certaines données de santé), parfois parce que l'*open data* n'est pas encore bien en place (y compris auprès des services du Ministère en charge de la Politique de la ville, nous y reviendrons), parfois pour de réelles raisons techniques (les données existent dans un système de gestion mais ne sont extractibles à l'échelle des QPV). Il y a aussi des données privées. Elles sont produites par des entreprises qui acceptent de les mettre à disposition

de Rés0 Villes au nom de l'intérêt général. Citons parmi les partenaires du programme : le Groupe Randstad, Harmonie Mutuelle ou encore la société Poligma.

Publiques ou privées, les organisations sont souvent frileuses et s'inquiètent des usages qui pourraient être faits de leurs données. **Créer les conditions du partage n'est pas chose aisée.** L'expérience de « Data & Quartiers » nous a même montré qu'il est parfois nécessaire de réorienter les travaux, faute d'accès aux données.

Dans tous les cas, **un processus rigoureux a été mis en place** : définition partagée des objectifs, préparation d'une **convention de mise à disposition** de données, définition de **règles techniques** (modalités de transmission par exemple) et **juridiques** (vérifications du respect du RGPD, règles d'accès, règles de conservation et de suppression). Cette étape peut prendre beaucoup de temps...

Préparer et explorer les données

Une fois les données collectées, il convient de s'assurer qu'elles sont exploitables au regard des objectifs. Les experts data prennent alors la main et vont explorer les différents fichiers et les bases de données mises à disposition. Ils vont notamment se pencher sur leur qualité : le format et le type de données, leur granularité (c'est-à-dire leur précision et leur fréquence de mise à jour), leur distribution et leur complétude (a-t-on bien toutes les données utiles ?), leur qualité (données aberrantes, incohérences...), etc. Cette phase peut donner lieu à des allers-retours entre les « fournisseurs » et les « explorateurs » de données afin de les enrichir et de les compléter.

À vrai dire : **le travail de *data science* débute souvent par des heures de « toilettage » de tableaux Excel !**

Pour le programme « Data & Quartiers », la principale difficulté à laquelle nous avons été confrontés a été de **vérifier que nous pouvions utiliser les données transmises à l'échelle des quartiers**. Nous verrons plus loin que ça n'a pas été toujours le cas, et que c'est d'ailleurs sans doute là un enseignement utile de notre programme expérimental.

Modéliser

En fonction des données disponibles et des objectifs à atteindre, les *data scientists* vont **appliquer une ou plusieurs techniques de modélisation**. Il peut s'agir de modèles statistiques classiques ou d'algorithmes et de techniques plus avancés, y compris ayant recours à l'intelligence artificielle. Les *data scientists* vont ainsi **tester les jeux de données** et voir si les informations que l'on espère en tirer vont en découler. Cette phase, la

plus technique du projet, peut parfois apparaître comme une « boîte noire », notamment lorsque les techniques et les modèles utilisés sont difficiles (voire impossibles) à expliquer. C'est par exemple le cas lors de l'utilisation de *deep learning* ou « apprentissage profond » qui repose sur l'utilisation de réseaux de neurones artificiels et dans lesquels la prise de décision de l'**algorithme** est très complexe à expliquer à des non-experts. Pour « Data & Quartiers » nous nous sommes limités à des outils plus simples, et leur fonctionnement sera expliqué.

Évaluer le modèle

Les *data scientists* appliquent leurs modèles sur les jeux de données de tests afin d'en comparer les performances. Les résultats obtenus sont ensuite présentés aux « experts métiers » afin d'en valider la pertinence. **Nous allons vérifier que les informations obtenues correspondent aux besoins et sont cohérentes avec la connaissance que l'on a des quartiers.** Parfois, il faut corriger les choses, par exemple en collectant des données différentes. Ce n'est qu'une fois la démarche validée, que l'on se lance dans un traitement à grande échelle des données. Appliquée à « Data & Quartiers », cette méthode nous a par exemple conduit à travailler sur 5 ou 6 quartiers sur des données de santé, avant d'étendre le travail aux 78 QPV de Bretagne et des Pays de la Loire.

✓ Étape n°4 : prototyper la solution

Une fois vérifiée la pertinence des données et démontrée l'utilité de la modélisation (avec ou sans intelligence artificielle), il faut réfléchir à la manière de rendre les données « actionnables », c'est-à-dire **trouver une forme de traitement des données qui va permettre aux acteurs de s'en servir de façon concrète.**

De nombreuses questions se posent : faut-il un outil que chacun peut manipuler à distance ? Faut une représentation statique (un tableau, une carte...) ?

Comment représenter des données complexes à l'échelle de chaque quartier ? A l'échelle des villes ou des régions ? Les données doivent-elles être figées dans le temps ou faut-il prévoir de les actualiser ? Avec quel degré d'automatisation ?

La meilleure façon de procéder est de concevoir collectivement l'outil en listant d'abord les fonctionnalités indispensables puis celles qui pourront être ajoutées lors de futures itérations. **Les experts data doivent ici s'effacer au profit des professionnels qui expriment leurs besoins** et des développeurs qui se chargent de les mettre en œuvre.

On emploie parfois le terme de *product owner* pour désigner la personne en charge de lister et de prioriser les fonctionnalités à mettre en place. Dans le cadre de « Data & Quartiers » ce rôle a été tenu par un chef de projet, qui a assuré le lien entre RésO Villes, les différents métiers et les spécialistes data.

✓ Étape n°5 : mettre en production

Cette cinquième et dernière étape est mentionnée à des fins d'information car le programme « Data & Quartiers » dans sa version 2019-2022 était un programme expérimental. Il s'est arrêté à l'élaboration des prototypes. Mais l'objectif logique d'une démarche de *data science* est de **fournir ensuite des outils qui seront intégrés dans les processus de travail au quotidien.** A la différence des nombreuses études statistiques traditionnelles qui aident les acteurs des quartiers mais qui sont mises à jour de façon ponctuelle (notamment lors des renouvellements des contrats de ville), **les nouveaux outils data pourront être mis à disposition en continu auprès des équipes des villes, des travailleurs sociaux, des conseillers citoyens, etc.** Les modèles et les interfaces développés seront conçus de façon à être intégrés dans les systèmes d'information existants.

Ceci nécessitera la participation des équipes informatiques des différents acteurs des territoires afin de **résoudre des questions techniques** concernant l'automatisation des flux de données ainsi que leur stockage et leurs traitements. Cette implication des équipes informatiques devra être faite le plus en amont possible, idéalement dès la phase de cadrage du projet, afin d'anticiper les défis techniques.

Les **modalités économiques de mise à disposition des données** devront aussi avoir été négociées au préalable. Une chose est de mettre à disposition des données pour un prototype, une autre de les laisser en libre accès sur la durée... L'exemple du Groupe Randstad dans « Data & Quartiers » illustre bien cet enjeu. Le groupe a mis à disposition des données extrêmement précieuses concernant le marché de l'emploi pour notre expérimentation, mais dans quelles conditions pourrait-il le faire de façon pérenne ? Avec potentiellement un flux de données mis à jour quotidiennement à l'échelle de tous les bassins d'emploi de Bretagne et des Pays de la Loire...

Synthèse

Avant de détailler les trois grands projets mis en œuvre dans le cadre de « Data & Quartiers », un schéma permet de synthétiser la méthode que nous avons utilisée. Certes, il a fallu parfois s'adapter et conserver un peu de souplesse. Mais nous avons essayé, et plutôt réussi, à respecter cette méthodologie tout au long de ces trois années.

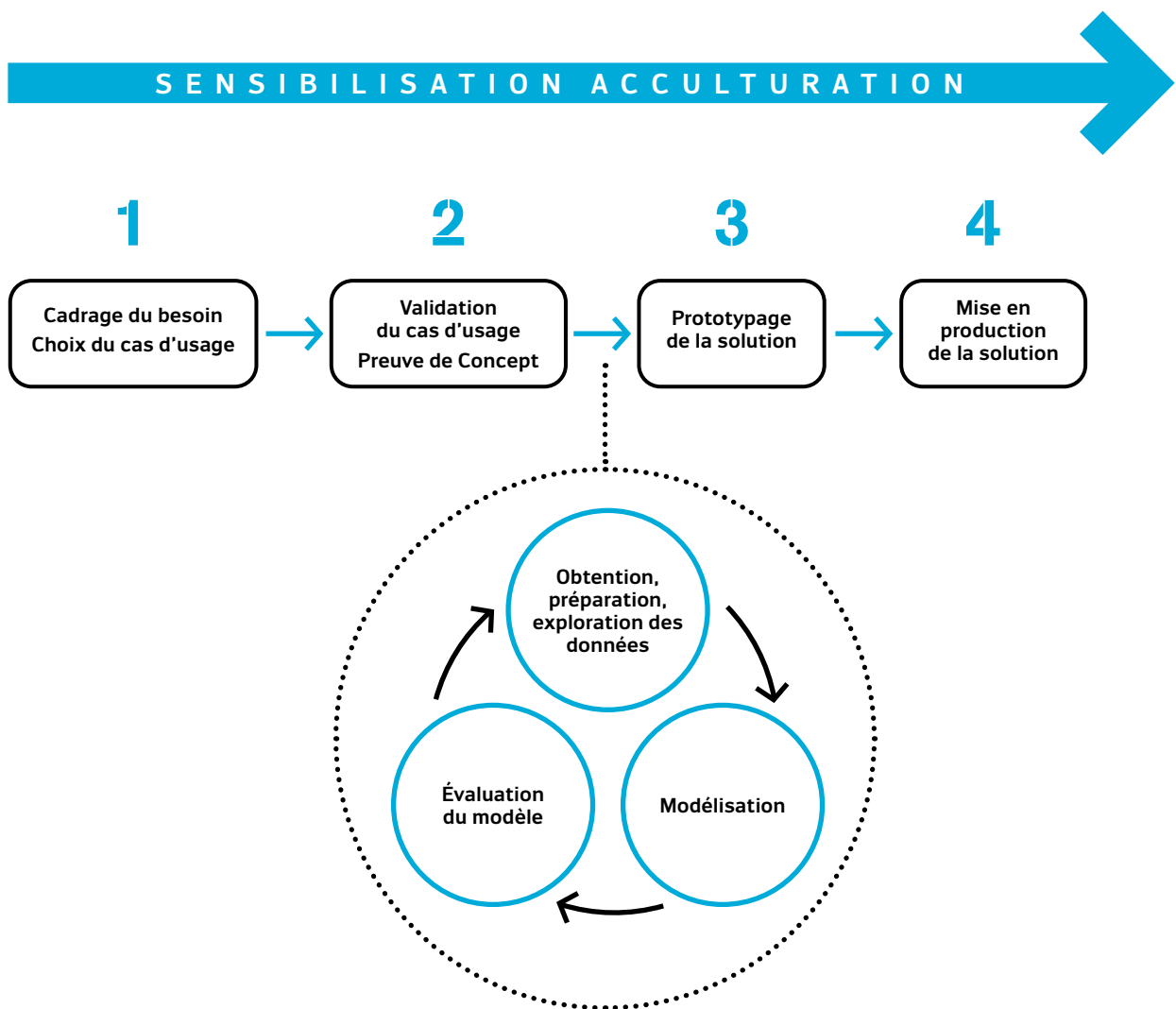


Schéma de la méthodologie projet «Data & Quartiers»

Les prototypes « Data & Quartiers »



PARTIE 2



Data & Emploi dans les quartiers : une étude sur les distances domicile-travail

Le cadrage du projet

Le groupe de travail « Emploi » est le premier à avoir démarré son action. Les partenaires initiaux étaient les suivants : l'équipe de **RésO Villes, Pôle emploi** (au niveau national et les deux directions régionales de Bretagne et des Pays de la Loire), le **Groupe Randstad France**, ainsi que **Nantes Métropole** et **Brest Métropole**.

Définir un projet sur la question de l'emploi dans les quartiers n'a pas été simple. La **profusion de données disponibles (ou supposées disponibles)** a facilité la formulation d'intentions très variées. Les membres du groupe de travail ont envisagé conduire des travaux sur la lutte contre le non-recours aux droits, l'accompagnement dans la recherche d'emploi, les habitudes de mobilité des habitants des QPV, les dispositifs d'accompagnement solidaires par les habitants eux-mêmes, etc.

À l'issue du premier atelier, deux thématiques principales ont émergé. La première visait à identifier et objectiver les difficultés d'accès aux droits qui seraient spécifiques aux QPV et qui expliqueraient une difficulté d'accès à l'emploi. Ainsi par exemple : la méconnaissance des dispositifs d'aide à la mobilité est-elle réelle ? Est-elle mesurable ? Est-elle plus forte dans les quartiers qu'ailleurs ? Est-elle un frein pour accéder à un emploi ? La seconde thématique concernait la mobilité avec l'idée d'identifier des caractéristiques propres aux QPV qui seraient un frein pour l'accès à l'emploi. Par exemple : est-ce que la distance moyenne domicile-travail est plus élevée dans les quartiers ? Quel impact sur les déplacements des situations de multi-employeurs (par exemple pour les métiers de l'entretien) ? Peut-on mesurer l'impact des politiques tarifaires sur l'usage, dans les quartiers, des transports publics ?

À la rentrée 2019, le groupe de travail s'est réuni une deuxième fois pour **formuler de façon précise une seule problématique** pour laquelle l'accès à des données nouvelles, et peut-être des procédés algorithmiques nouveaux pourraient apporter des réponses. La problématique retenue a été choisie pour son originalité et l'absence de travaux déjà existants, mais aussi pour sa transversalité. La voici : **« dans les quartiers de la politique de la ville, quelle est la réalité de la distance entre le lieu de vie des demandeurs d'emploi et la localisation des offres d'emplois qui correspondent à leur recherche ? »**.

Une seconde question a été mise en réserve : « quels sont les moyens de transport à la disposition des habitants pour pouvoir accéder aux emplois qui correspondent à leur recherche ? ». Cette seconde question sera l'objet d'un prototype en 2021 qui exploitera les données du groupe de travail emploi.



Une réunion du groupe de travail sur l'emploi à Nantes (juin 2019)

✓ Les étapes de travail

L'obtention des données

Pour répondre à la question posée, il fallait évidemment accéder à des données concernant le chômage et les demandeurs d'emploi des quartiers. **De nombreuses statistiques existent. Elles sont disponibles à de larges échelles auprès de l'INSEE par exemple.** Mais la source des données c'est Pôle emploi. Et la question posée n'est pas simple. **Il ne s'agit pas d'obtenir des informations sur le taux de chômage, ni le profil des demandeurs d'emploi.** La question posée nécessite d'**accéder à une information qui n'est pas disponible en open data** : identifier, quartier par quartier, les emplois que recherchent les demandeurs d'emploi. Cette information, **Pôle emploi** en dispose. Les demandeurs d'emploi indiquent (notamment lors du premier entretien en agence) leurs souhaits.

Avant de parler distance et temps de trajet, il faut aussi **disposer d'informations sur les offres d'emploi qui vont correspondre aux métiers recherchés par les habitants de chaque quartier.** Il s'agit là d'accéder à des données massives. Pôle emploi seul n'y suffit pas car une grande majorité des offres d'emploi des entreprises transitent par d'autres canaux. L'idée du groupe de travail a été de s'adresser à une entreprise privée : le **Groupe Randstad France**. Connu pour ses agences de travail temporaire, le Groupe Randstad développe aussi des outils *big data* d'analyse des bassins d'emploi. Ces outils lui servent notamment à réaliser des études pour des collectivités locales ou des entreprises dans des logiques d'implantation. Les données qui intéressaient RésO Villes étaient issues d'**un dispositif algorithmique qui récupère de façon automatique**, - dans le jargon de la data, on dit qu'il *scrappe* -, **toutes les annonces publiées quotidiennement sur près de... 11 000 sites internet !** Dédoublonnées, ces données constituent sans doute l'une des premières bases d'offres d'emploi en France. En outre, Randstad s'assure de leur conversion dans des formats permettant le croisement avec la liste des métiers recherchés fournie par Pôle emploi.

Jusqu'ici, tout va bien. Mais **de nombreuses semaines ont été nécessaires pour organiser l'accès aux données de Pôle emploi et de Randstad.** En tant qu'organisme public, Pôle emploi est soumis à des obligations d'*open data*. Mais les données que nous souhaitions obtenir ne sont pas des données triviales et immédiatement accessibles. Il faut définir

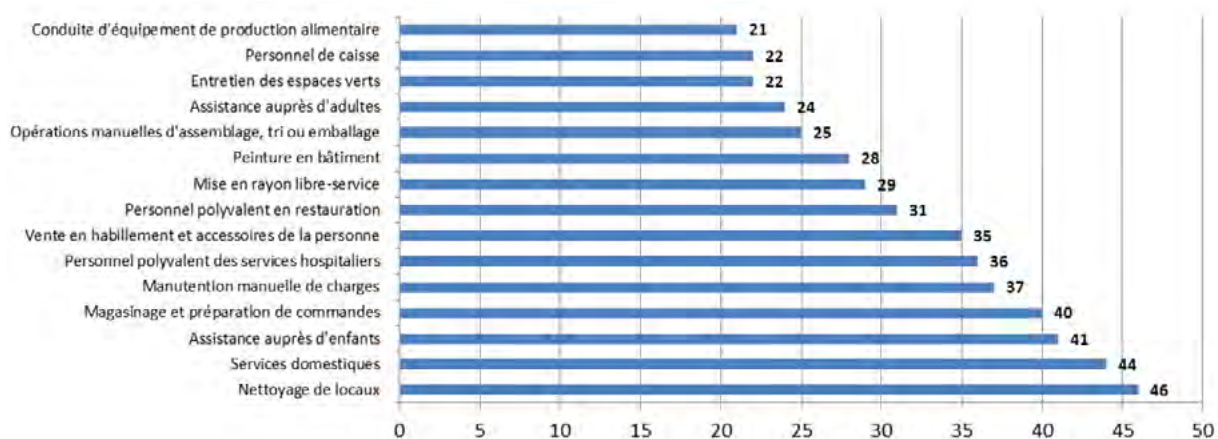
précisément ce qui est attendu, en l'occurrence **une extraction pour chacun des 78 QPV de Bretagne et des Pays de la Loire**, de la liste des métiers les plus demandés. Ce traitement a été effectué par chacune des deux directions régionales. Avec des limites malgré tout : dans certains quartiers, le nombre de demandeurs d'emploi recherchant tel ou tel métier était inférieur au seuil du « **secret statistique**³ ». En d'autres termes : leur nombre étant très faible, un risque pouvait exister qu'à partir d'une donnée statistique on puisse identifier une personne. « *L'idée principale est de ne pas pouvoir identifier une personne physique par croisement de données au sein de nos productions* » nous a précisé dans un mail le responsable du service « Statistiques, Études et Évaluations » de Pôle emploi Pays de la Loire. La donnée statistique anonyme deviendrait alors une donnée personnelle. Et Pôle emploi n'a bien évidemment pas le droit de nous la communiquer.

Mais une autre difficulté a surgi. Si la direction nationale de Pôle emploi (et son équipe data) a accepté de fournir les données à RésO Villes, **la perspective d'un croisement opéré avec les équipes de data science de Randstad n'allait pas de soi.** Les relations entre l'opérateur public et les grands groupes de travail temporaire ne sont pas simples. Leurs actions sont souvent complémentaires, mais parfois concurrentes. Les équipes de Pôle emploi voulaient **avoir l'assurance que les données transmises ne serviraient qu'au prototype de « Data & Quartiers », et qu'elles ne seraient pas utilisées à des fins économiques** par exemple. Une convention a été signée entre RésO Villes et Randstad. Elle précise que le « *Groupe Randstad France se verra transmettre directement les jeux de données utiles en vue de la réalisation de l'expérimentation par les partenaires du groupe de travail. Pour réaliser cette prestation, le Groupe Randstad France traitera et stockera des jeux de données qui sont la propriété d'autres partenaires (...). Les jeux de données transmis ne pourront faire l'objet d'un traitement que dans le cadre de l'expérimentation et ne seront pas conservés au-delà de la durée de la présente convention* ».

Préparer une telle convention, s'accorder sur les principes, faire valider les termes par les directions juridiques... tout ceci prend du temps. Mais **les données transmises par Pôle emploi pour chacun des 78 QPV de Bretagne et des Pays de la Loire, furent à la hauteur de nos attentes.** Nous avons la liste des quinze métiers (codes ROME⁴) les plus recherchés par les habitants de chacun des quartiers prioritaires. Ces données nous ont été transmises dans un premier temps pour une période d'un mois (les

³ Le secret statistique est une règle qui garantit aux personnes concernées par des données que leurs informations ne seront pas utilisées d'une façon susceptible de leur porter tort. Source : Guide du secret statistique INSEE 2022.

⁴ Le code ROME correspond à un code identifiant chaque métier au sein d'une classification nationale. Le Répertoire opérationnel des métiers et des emplois comporte plus de 10 000 métiers.



Exemple de visualisation des codes ROME les plus demandés par les habitants des QPV, à l'échelle régionale des Pays de la Loire

demandeurs d'emploi inscrits au mois de novembre 2019). Elles ont été transmises aux équipes de *data science* de Randstad qui ont pu en valider la qualité et réaliser les premiers croisements avec leurs données propres. Ultérieurement, nous récupérerons les données pour toute l'année 2020.

Notons ici que le fait de disposer de la liste des métiers recherchés dans chaque quartier est déjà riche d'enseignements (ou de confirmations) pour les acteurs des territoires. Les métiers les plus recherchés ne sont pas les mêmes à Malakoff à Nantes et Villejean à Rennes, qu'à Ville-Ouest à Saint-Nazaire, Ker Uhel à Lannion ou Moulin Liots à Fontenay-le-Comte.

En parallèle, **le Groupe Randstad a fourni les données des offres d'emploi disponibles dans chacun des 32 bassins d'emploi ayant des QPV en Bretagne et en Pays de la Loire**. Il devenait donc possible de croiser ces données pour mettre en regard, quartier par quartier, les emplois recherchés et les offres disponibles.

Restait alors à **calculer la distance entre ces offres et le domicile des demandeurs**. Nous n'avions pas, bien sûr, l'adresse de chacun puisqu'il s'agit de données personnelles. Nous avons fait le choix d'**utiliser comme point géographique de référence le « centroïde » des quartiers**, c'est-à-dire un point central qui fera office de repère pour l'ensemble du quartier. Ces centroïdes sont souvent utilisés pour des travaux d'analyse et de recherche. Leurs coordonnées GPS nous ont été transmises par l'**Observatoire national de la politique de la ville (ONPV)**.

Pourquoi des coordonnées GPS ? Pour calculer automatiquement la distance entre les offres d'emploi et les demandeurs. Et ceci a été fait en utilisant **un outil algorithmique bien connu du grand public : le GPS Tom-Tom**.

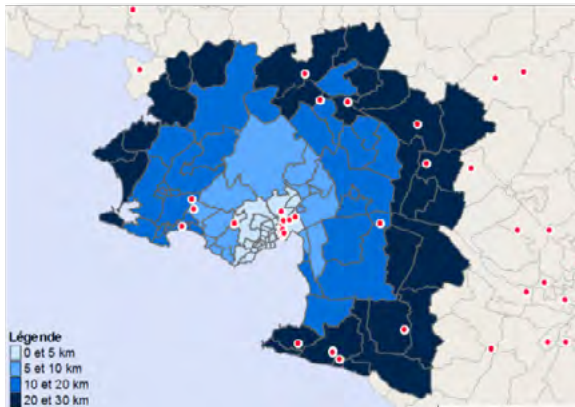
La preuve de concept

Une fois les jeux de données réunis, Randstad a démarré le travail de *data science*. Pour répondre aux objectifs de pédagogie du programme et favoriser l'apprentissage de l'équipe projet, cette étape a été rythmée par de nombreux allers et retours entre Randstad et RésO Villes. Des versions « tests » et des restitutions intermédiaires ont jalonné la phase de production. Durant cette étape, Randstad a mis son savoir-faire technique au service du programme « Data & Quartiers », formant ainsi l'équipe de RésO Villes. Inversement, RésO Villes a apporté sa connaissance des quartiers et des enjeux liés à l'emploi dans les quartiers.

À cette étape du travail, de nombreuses données ont été exploitées et présentées pour une phase de test :

- Le « top 15 » des codes ROME avec le plus grand nombre de demandeurs d'emploi habitant dans un QPV à l'échelle des régions Pays de la Loire et Bretagne (visualisation de données fournies par Pôle emploi)
- Le « top 15 » des codes ROME avec le plus grand nombre de demandeurs d'emploi des QPV de Saint-Nazaire et de Brest
- Une cartographie de la localisation des offres d'emploi correspondant aux métiers recherchés par les demandeurs d'emploi des QPV « Petit Caporal » de Saint-Nazaire et « Bellevue » de Brest (croisement des données Pôle emploi et Randstad)

Ces visualisations de données ont été accompagnées **d'un exercice de mise en récit (on parle parfois de *data storytelling*)** afin de faciliter la présentation des travaux de *data science* et d'amener les partenaires du programme à se projeter sur les utilisations possibles pour les habitants des QPV et/ou les décideurs publics.



Exemple de visualisation de la localisation des offres d'emploi correspondant aux métiers les plus recherchés par les habitants du quartier « Petit Caporal » à Saint-Nazaire

Le partage de ces travaux intermédiaires a permis de **démontrer que le croisement des données des offres d'emploi (fournies par Randstad) et des données des emplois les plus demandés par quartier (fournies par Pôle emploi) révèle des informations jusque-là inédites.** L'analyse visuelle via les cartographies présentées montre déjà certaines disparités territoriales. Et une question se pose quant aux capacités des habitants à se déplacer pour accéder à ces emplois.

La mise à l'échelle

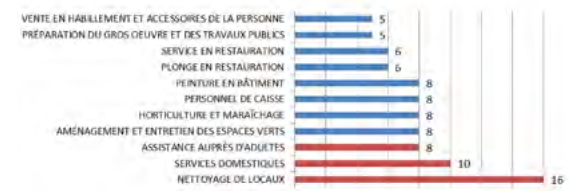
RésO Villes et ses partenaires vont alors s'atteler à généraliser (et automatiser l'analyse). Pour cela, une nouvelle étape est franchie en matière de données. Puisque le test a été concluant avec un mois « test », Pôle emploi va procéder à **l'extraction des données des demandeurs d'emploi des 78 quartiers sur la totalité de l'année 2020.** Nous disposons donc du « cumul annuel dédoublonné des demandeurs d'emploi par code ROME et par quartier ». **Avec toujours la limite protectrice du secret statistique.**

Voici un exemple des nouvelles données fournies par Pôle emploi et leur interprétation :

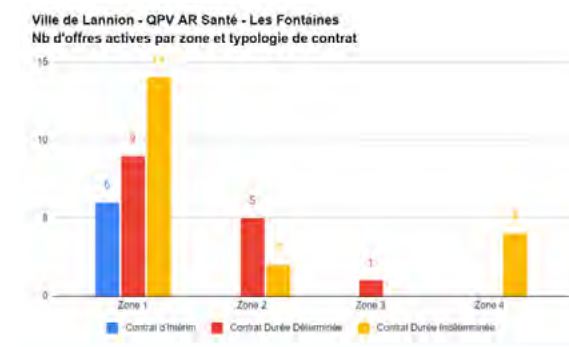
DPT : 22
QPV : QP022007
Libellés QPV : LA FONTAINE DES EAUX
Code commune : 22050
Libellé commune : Dinan
ROME : A1203
Libellés ROME : Aménagement et entretien des espaces verts
Demandeurs d'emploi inscrits en 2020 : 9

Au cours de l'année 2020, 9 personnes différentes vivant dans le quartier La Fontaine des Eaux se sont inscrites à Pôle emploi en indiquant rechercher un emploi correspondant au métier « aménagement et entretien des espaces verts ».

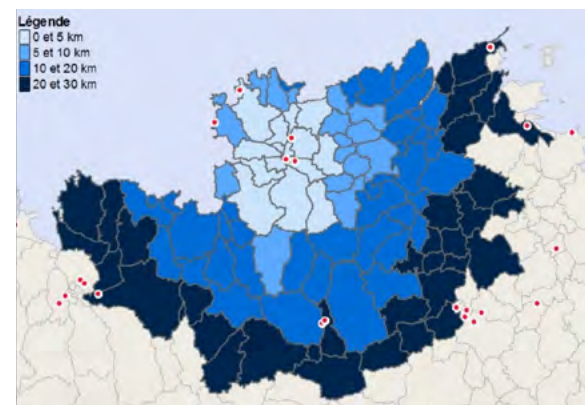
La deuxième itération avec l'équipe de *data science* de Randstad a donc abouti, quelques semaines plus tard à la livraison d'une série de visualisations, dont voici des extraits pour le quartier Ar Santé – Les Fontaines à Lannion



Exemple de visualisation des métiers les plus recherchés en 2020 pour le QPV Ar Santé - Les Fontaines à Lannion.



Exemple de visualisation du nombre total d'offres d'emploi disponibles, par typologie d'emploi (CDI, CDD, travail temporaire) et par proximité géographique avec le quartier (moins de 5 km, de 5 à 10 km, de 10 à 20 km, de 20 à 30 km) pour le QPV Ar Santé - Les Fontaines à Lannion.



Cartographie des offres d'emploi correspondant aux 5 codes ROME le plus recherchés.

La présentation des résultats de cette deuxième itération auprès de RésO Villes a permis de **confirmer le caractère innovant de l'expérimentation et l'intérêt des décideurs publics comme des acteurs des QPV pour les résultats.**

Mais **plusieurs points ont été soulevés pour améliorer encore les travaux sur les données.**

Les tableaux présentent des informations riches mais complexes (en mélangeant par exemple les distances et le statut des emplois). En outre, la localisation des emplois proposés par les agences d'intérim biaise les résultats car l'adresse disponible est très souvent celle... de l'agence d'intérim ! Enfin, le travail très fin quartier par quartier, avec des données différentes d'un quartier à l'autre en termes de métiers recherchés, rend difficile toute comparaison. **Paradoxe : après avoir tout fait pour obtenir des données précises pour chaque quartier, le groupe de travail souhaite aussi des données globales...**

Ces éléments ont été pris en compte pour réaliser le prototype data.

Le prototype data

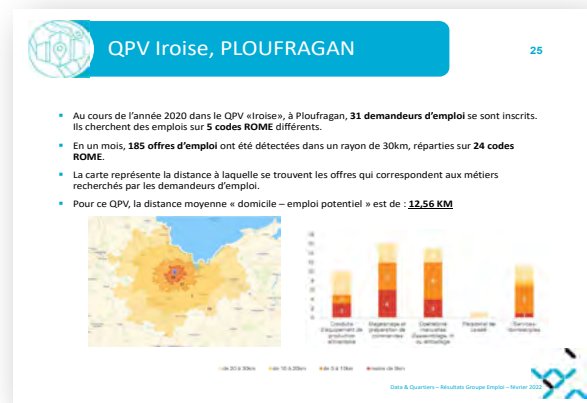
La livraison finale de ce travail, qui a donc pris plusieurs mois (avec une coupure liée au contexte sanitaire de l'année 2020) comporte deux éléments complémentaires : **un outil d'analyse inédit quartier par quartier**, et **une application visuelle interactive permettant de comparer les quartiers entre eux.**

78 fiches sur la distance emploi recherché - travail

Le terme de prototype n'est peut-être plus totalement approprié pour ce premier support étant donnée l'absence de développement d'interface informatique pour la lecture des données. Certains préféreront donc utiliser le terme de « maquette ». Nous avons développé pour chaque QPV deux visualisations graphiques inédites : une visualisation cartographique des offres d'emploi qui correspondent aux métiers les plus recherchés et une visualisation plus fine de la répartition des offres par code ROME, sous forme d'histogrammes.

Avec ces données, nous avons aussi créé **un indicateur de « distance moyenne domicile – emploi potentiel » par quartier.**

Les fiches permettent aux décideurs politiques et aux acteurs des quartiers d'avoir, en un coup d'œil, des informations clés sur l'emploi dans les QPV qui les concernent. C'est un outil d'aide à la décision, certes à ce stade non automatisé, mais qui permet une vision nouvelle de la réalité des questions d'emploi dans chacun des quartiers.



Exemple de fiche emploi par QPV reprenant certaines données brutes, des visualisations graphiques et un indicateur de distance moyenne.

➔ **LA TOTALITÉ DES FICHES EMPLOI POUR LES 78 QPV EST DISPONIBLE À L'ADRESSE SUIVANTE :**

<https://www.resovilles.com/programme-data-quartiers-distance-domicile-travail-les-habitants-des-qpv-pas-tous-loges-a-la-meme-enseigne/>

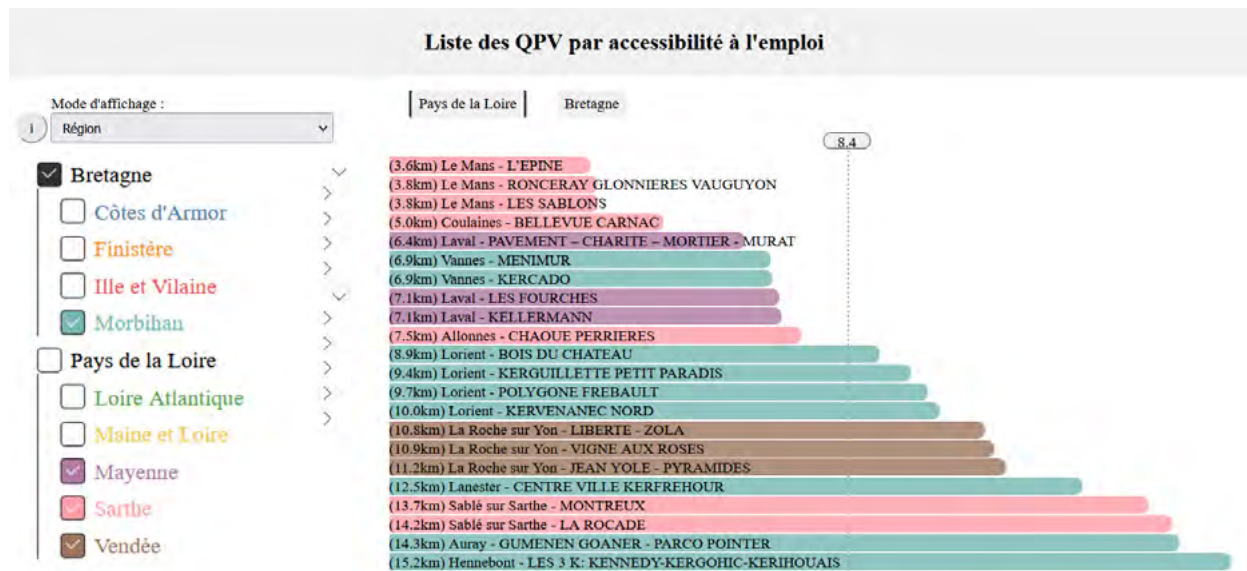


Flashez le QR code

Une « dataviz » dynamique

En plus de ces 78 fiches Emploi QPV, nous avons développé **un outil de visualisation dynamique des distances entre le domicile et les emplois potentiels.** L'objectif de la construction de cet outil était double : permettre de comparer plus facilement les différents QPV entre eux mais aussi sensibiliser et faire de la pédagogie auprès des élus et des acteurs des quartiers, pas ou peu habitués à « jouer » avec ce genre d'outils.

Le programme « Data & Quartiers » nous a démontré que ce genre d'interface permet l'appropriation des données et des modélisations. Nous pensons que de tels outils ont **un plus fort impact que des rapports statistiques classiques.**



Capture d'écran de l'outil de visualisation dynamique des distances entre le domicile et les emplois potentiels

En quelques clics, il est possible de voir que parmi nos 78 QPV, **les 10 quartiers pour lesquels les distances domicile-emploi sont les plus courtes sont situés à Rennes, Nantes ou Angers** (avec moins de 4km de distance à parcourir en moyenne). À l'inverse, si l'on observe les QPV avec les distances les plus élevées, on retrouve des quartiers situés dans de plus petites villes comme Fontenay-Le-Comte (20km), Hennebont (15km), Saumur (14,5km), etc.

En utilisant la *dataviz* plus en détail, on remarque aussi que parmi les 5 villes de plus de 100 000 habitants incluses dans l'analyse, Brest se démarque avec une distance moyenne domicile-emploi de 8,2km beaucoup plus élevée qu'à Angers (3,6km), Le Mans (3,7km), Rennes (3,8km) et Nantes (3,9km). Une différence qui s'explique sans doute par la géographie et le développement urbain.

Chacun peut librement s'essayer à d'autres lectures.

➔ **L'OUTIL EST DISPONIBLE ICI :**
<https://www.resovilles.com/emploi-qpv/>



Flashez le QR code

Cette datavisualisation (dans le jargon data on parle de *dataviz*) a été réalisée par le prestataire Valoway en utilisant des technologies web simples (HTML, CSS et javascript).

FOCUS: La datavisualisation

La représentation graphique des données, ou datavisualisation (souvent abrégée en *dataviz*) est une discipline qui consiste à faciliter la compréhension de phénomènes, et en conséquence aider à la prise de décision, en restituant des données souvent complexes de la façon la plus intelligible possible.

Les avancées technologiques et notamment l'essor d'outils en libre-service qui permettent à des utilisateurs ne disposant pas de compétences poussées en informatique de travailler eux-mêmes sur leurs données, ont démocratisé une nouvelle génération de *dataviz* dynamiques qui remplacent ou complètent les traditionnels graphiques issus de logiciels tableurs.

Les datavisualisations sont de plus en plus présentes dans tous les domaines (journalisme, sport, politique, etc.) pour livrer aux utilisateurs finaux des données non seulement compréhensibles, mais aussi agréables à « naviguer ».

Nous conseillons par exemple, pour le plaisir des yeux, les travaux de Federica Fragapane (consultables ici : <https://www.behance.net/FedericaFragapane>).

APARTÉ :

Quelques enseignements

Tirons ici quelques enseignements de ce premier « cas d'usage ».

Enseignement n°1

la valeur des données augmente avec les possibilités de croisement

C'est probablement l'enseignement principal de ce premier groupe de travail. Les **données mises à disposition par Pôle emploi** quartier par quartier étaient très intéressantes. Celles « scappées » par le **Groupe Randstad** l'étaient aussi. Mais **c'est leur croisement qui génère le plus d'informations utiles**.

Nous voyons par exemple que les enjeux d'accès à l'emploi en termes de distance et de déplacement ne sont pas du tout les mêmes dans les QPV des grandes métropoles que dans ceux des territoires péri-urbains ou ruraux. Prises isolément, ni les données de Pôle emploi, ni celle de Randstad, ne permettent de le savoir.

Enseignement n°2

le partage des données ne va pas de soi... et c'est normal !

Ce premier prototype a été riche d'enseignements sur la méthode pour l'équipe de RésO Villes. **Il ne suffit pas d'afficher un objectif d'intérêt général pour que tous les acteurs mettent spontanément leurs données à disposition**. Qu'il s'agisse de données publiques ou de données privées. **Les acteurs veulent avoir des garanties** : garanties que le RGPD sera respecté, garantie que des données ne vont pas être utilisées à d'autres fins, garanties qu'elles seront traitées en conformité avec des règles juridiques (et même éthiques) adéquates.

Et c'est très bien ainsi, même s'il faut du temps pour construire un cadre de confiance et le traduire dans des conventions *ad hoc*.

Enseignement n°3

l'importance de la preuve de concept (ou POC)

Dans un projet Data, la **première phase de travaux réalisée sur un échantillon volontairement faible de données** (une période courte, un territoire réduit) est communément appelée « preuve de concept » ou « POC » pour l'acronyme en anglais de *Proof Of Concept*. Il s'agit d'explorer les données, d'en valider la qualité et de vérifier certaines hypothèses de départ. Cela permet aussi de réduire les potentielles incompréhensions entre les spécialistes de la data et les experts en charge des projets. Dans certains cas, un POC non concluant peut mener à l'abandon du projet. Son intérêt est de minimiser l'incertitude inhérente aux projets d'innovation, à moindre coût.

Enseignement n°4

les algorithmes au service de la politique de la ville ?

« Data & Quartiers » visait à utiliser les outils de la *data science* au service des quartiers. L'innovation en matière de recours aux données repose sur deux axes principaux : l'accès à des données nouvelles (permettant d'accéder à des informations inédites) et/ou l'exploitation des données par des procédés innovants (des algorithmes, simples ou très sophistiqués notamment en cas de recours à l'intelligence artificielle). Pour ce prototype emploi, les algorithmes ont eu leur importance... en amont. Notamment les algorithmes de Randstad qui collectent quotidiennement, dédoublonnent, trient et organisent les données issues de 11 000 sites internet, et les algorithmes de Tom-Tom qui ont servi à calculer automatiquement les distances entre les emplois identifiés et le centre de nos quartiers.

Mais le travail mené ensuite a été sobre en algorithmie. Pour donner à voir ces données innovantes, nous avons préféré des outils moins sophistiqués mais plus impactants.

Data & Santé dans les quartiers : une étude sur la fragilité des habitants face au Covid-19

Le cadrage du projet

La crise sanitaire du Covid-19 a marqué les partenaires de « Data & Quartiers » de plusieurs manières. D'abord, comme pour tout le monde, il a fallu adapter notre méthode de travail en privilégiant les sessions à distance et en prenant en compte le fait que les priorités de chacun étaient centrées sur d'autres missions, plus essentielles. Ensuite parce que les membres du Conseil d'administration de RésO Villes ont souhaité **s'appuyer sur « Data & Quartiers » pour proposer un travail utile aux acteurs publics dans la période.**

Très vite des études nationales et internationales ont mis en avant le fait que les habitants de certains quartiers pourraient être plus fragiles que d'autres face au Covid-19⁵. En parallèle naissaient aussi des polémiques sur une indisciplinisme supposée dans certains quartiers face aux règles de confinement. Puisque différents travaux semblaient montrer que **toutes les populations ne sont pas égales face au Covid-19, que certaines présentent des risques plus importants que d'autres d'être contaminées ou de développer des formes graves**, et puisque des facteurs sanitaires ou sociaux étaient annoncés comme pouvant être à l'origine de cette inégalité, **RésO Villes a souhaité objectiver la réalité des risques en Bretagne et Pays de la Loire.**

Cette étude a été conduite pour toutes les communes de la Région Bretagne ayant des quartiers prioritaires de la politique de la ville : Auray, Brest, Concarneau, Lannion, Lorient, Quimper, Rennes, Saint-Brieuc, Saint-Malo et Vannes. Au départ, la Région Pays de la Loire était bien sûr dans le périmètre du groupe de travail. Mais nous n'avons pas pu accéder aux données utiles, nous y reviendrons.

Le groupe de travail sur la santé rassemblait : l'équipe de **RésO Villes, Harmonie Mutuelle, l'Agence Régionale de la Santé Pays de la Loire, Valoway** ainsi que **Brest Métropole, Le Mans, Saint-Nazaire** et le bureau de **RésO Villes.**

Son premier travail a été de **sélectionner des indicateurs explicatifs d'une certaine fragilité**

face au virus : des indicateurs sanitaires et des indicateurs sociaux. Les indicateurs sanitaires ont été choisis en s'appuyant sur des travaux internationaux. Ainsi par exemple, nous savions que le taux de personnes sous **traitement antidiabétique** ou le taux de personnes sous **traitement antiasthmatique** nous seraient utiles. L'hypothèse des équipes de « Data & Quartiers », fondée sur leur connaissance des quartiers, était que les habitants des quartiers prioritaires souffraient plus que les habitants des autres quartiers, de ces pathologies. Mais encore fallait-il disposer de données pour le vérifier. D'autres types de comorbidité existent. Le taux de personnes déclarées comme souffrant d'une **affection de longue durée** pouvait aussi permettre de vérifier la présence d'habitants souffrant d'autres maladies que le diabète ou l'asthme les exposant fortement au Covid-19. L'affection longue durée (souvent citée par son acronyme « ALD ») détermine un régime de prise en charge totale par la Sécurité sociale pour les personnes qui sont victimes d'une maladie répertoriée. On en compte une trentaine, telles que les AVC, les insuffisances cardiaques, la maladie d'Alzheimer, la maladie de Parkinson, la sclérose en plaques ou encore les suites d'une transplantation d'organe.

Les **indicateurs sociaux** renvoient aux conditions de vie ou sociales des habitants des quartiers et permettent de cibler géographiquement ces habitants. L'avantage de ces données c'est qu'elles sont facilement accessibles, du moins à l'échelle des communes ou des IRIS⁶. Des données existent sur le **taux de personnes bénéficiaires de la CMU-C** (devenue C2S), ou encore sur le **taux de population vivant en logement HLM**, ces deux éléments combinés révélant un niveau de ressources et un mode d'habitat. Nous avons aussi utilisé le **taux de population immigrée** qui est une donnée fournie par l'INSEE et que nous avons utilisé avec précaution. En effet, en France comme ailleurs en Europe, la mise en œuvre de statistiques ethniques n'est pas autorisée. Mais la donnée utilisée apportait un éclairage complémentaire intéressant pour identifier des zones de vulnérabilité.

⁵ Citons par exemple les études SAPRIS (<https://anr.fr/Projet-ANR-20-COVI-0009>) et EpiCoV (<https://www.epicov.fr/publications>) menées par l'Inserm.

⁶ Les IRIS sont des unités statistiques spatiales plus petites que les communes utilisées notamment par l'INSEE. L'acronyme signifie : Ilots Regroupés pour l'Information Statistique.

✓ Les étapes de travail

L'obtention des données

L'ensemble des données sociales a été récupéré sur le portail *open data* de l'INSEE. Le taux de population bénéficiaire de la CMU-C a été extrait par IRIS au sein d'un jeu de données plus global sur les bénéficiaires de l'assurance maladie, le taux de population vivant dans des logements HLM a été extrait par IRIS d'un jeu de données sur le logement et le taux de population immigrée a été extrait, par IRIS également, au sein d'un jeu de données sur la structure de la population. L'accès à ces données a été extrêmement simple : il a suffi de se rendre sur le site internet de l'INSEE pour les sélectionner et les télécharger.

L'accès aux données sanitaires n'a pas été aussi facile... L'objectif était de pouvoir accéder aux données des personnes souffrant de diabète, d'asthme et d'une ou plusieurs affections de longue durée. Nous recherchions ces données par zone IRIS de sorte à pouvoir les croiser avec les données sociales et de bénéficier d'une granularité fine pour pouvoir mieux observer les territoires. Nous avons donc sollicité plusieurs acteurs. Des échanges ont eu lieu avec le **Health Data Hub**, une organisation nationale créée en 2019 comme tiers de confiance gestionnaire des bases de données du système de santé Français. Il est notamment chargé de mettre des données à disposition de la recherche médicale, suivant un protocole très rigoureux impliquant un contrôle de la Commission nationale informatique et libertés (CNIL).

En parallèle, une demande a été formulée auprès des **Agences régionales de santé (ARS) des Régions Bretagne et Pays de la Loire**. Aucune réponse immédiate n'a été obtenue. Au cœur de la crise sanitaire, les deux organisations répondaient à des urgences autrement plus prioritaires que notre projet.

Au final, les seules données que nous avons pu récupérer nous ont été transmises par un autre acteur : **l'Observatoire régional de la santé de Bretagne**. Cet observatoire disposait des données que nous souhaitions analyser à l'échelle des IRIS. Pourquoi ? Comment ? Nous n'en n'aurons pas l'explication, mais probablement cet Observatoire (co-financé par le Conseil régional de Bretagne) avait-il déjà travaillé sur ces pathologies à l'échelle des IRIS. Les données étaient donc disponibles. Nous apprendrons plus tard qu'en Pays de la Loire ces jeux de données n'existaient tout simplement pas. Ce sera aussi un enseignement utile : **il n'est pas possible de disposer de jeux de données qui n'ont tout simplement jamais été constitués**. Nous saurons nous en souvenir et anticiperons cette difficulté pour un autre prototype conduit avec la Caisse nationale d'assurance maladie en obtenant de nos interlocuteurs qu'ils procèdent à des extractions de données complexes à l'échelle de certains quartiers.

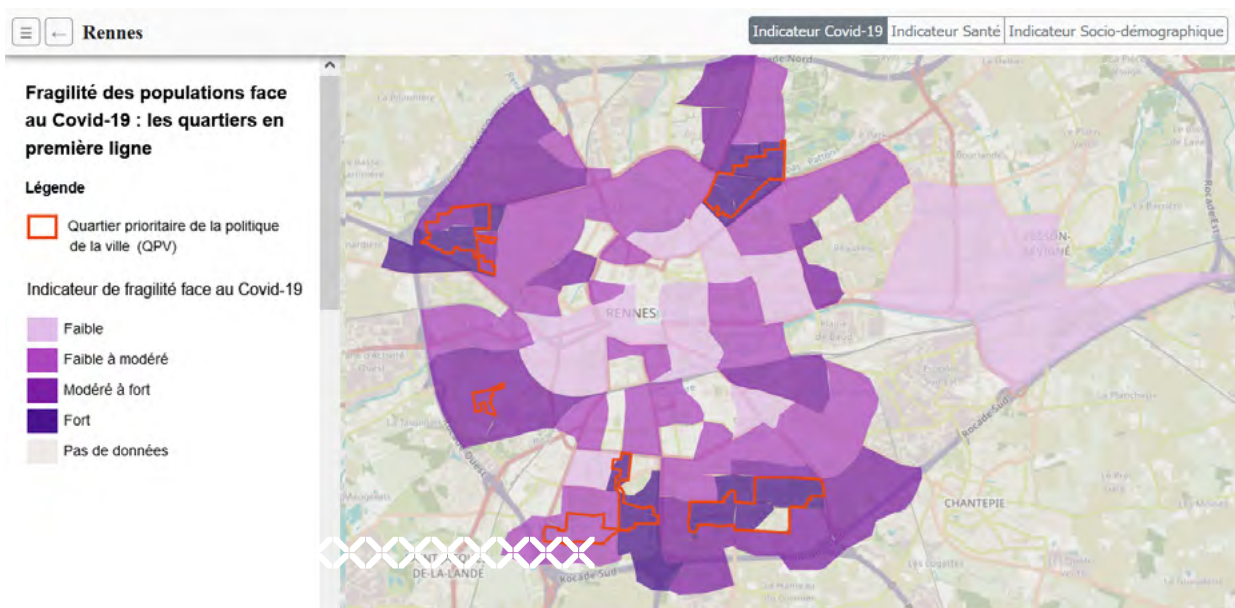
FOCUS: L'ouverture des données publiques

L'ouverture des données publiques ou *open data*, est un mouvement né en Amérique du Nord lorsque des villes comme New York, Chicago ou Montréal ont commencé à publier beaucoup de données sur leur gestion en 2008 et 2009. En France, quelques communes ont aussi publié très tôt des données, notamment Rennes, Bordeaux, Paris, Nantes ou Strasbourg entre 2010 et 2013. En 2016, la loi pour une République numérique a rendu cet exercice obligatoire, à la fois pour les administrations d'État et pour toutes les collectivités locales de plus de 3 500 habitants (ayant au moins 50 agents en équivalent temps plein).

Les objectifs de l'*open data* sont divers. Il s'agit d'abord de faire preuve de transparence démocratique. En publiant leurs données, les acteurs publics respectent un principe de redevabilité de leur action vis-à-vis des citoyens. Il s'agit aussi d'améliorer les coopérations entre administrations publiques en facilitant la circulation des données pour de meilleures prises de décision. Il s'agit enfin de permettre à des entreprises de créer de nouveaux services à partir de ces données.

Pour « Data & Quartiers », nous avons utilisé de nombreuses données ouvertes mises à disposition par des services publics nationaux (INSEE notamment) ou locaux. Ces données sont mises à disposition dans les formats directement exploitables (par exemple des tableaux de données en format .csv), c'est l'une des règles de l'*open data*.

Durant la pandémie plusieurs exemples d'applications célèbres ont montré comment des données ouvertes pouvaient être réutilisées : ce fut le cas de CovidTracker ou encore de Vite Ma Dose pour les vaccins.



Capture d'écran de l'outil de visualisation géographique de la fragilité face au Covid-19

La réalisation cartographique et le prototype data

Dès le début du travail sur le Covid-19, il a été souhaité que les résultats de l'étude puissent être lus par des élus, des agents des collectivités territoriales et des experts de la politique de la ville. Nous souhaitions disposer de cartes identifiant les zones de fragilité face à la maladie.

Le travail d'analyse a reposé sur le croisement des facteurs de risques : les facteurs sociaux d'un côté, les facteurs de santé de l'autre. Il n'y a pas ici de technique élaborée de *data science*. L'innovation réside dans le fait d'avoir réussi à accéder aux données !

La démarche a été itérative. L'équipe en charge des travaux de géomatique a livré dans un premier temps des cartographies statiques sur un nombre restreint de territoires. Puis un outil plus sophistiqué a été développé.

➡ **ACCÉDER AU PROTOTYPE DATA ICI :**

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



Flashez le QR code

Les résultats des travaux confirment les intuitions de départ de RésO Villes et valident à l'échelle locale les études nationales et internationales les ayant inspirés : toutes les populations ne sont pas égales face au Covid-19 et certaines présentent des risques plus importants d'être contaminés et/ou de développer des formes graves.

Pour la Bretagne, les résultats sont éloquentes : **les QPV sont systématiquement situés au cœur des zones dans lesquelles le risque de fragilité face au Covid-19 est le plus fort** (même si toutes les zones fragiles ne sont pas des quartiers de la politique de la ville). Dit d'une autre façon, **les habitants des QPV sont en général plus vulnérables que les habitants d'autres quartiers car ils cumulent des difficultés qui les fragilisent** (maladies longues et/ou pathologies telles que l'asthme ou le diabète) ou des conditions sociales qui les exposent plus au virus.

Data & Santé dans les quartiers : une étude sur le non-recours à la Complémentaire Santé Solidaire

Le cadrage du projet

En complément des travaux sur le Covid-19, un groupe de travail dédié à la question de la santé s'est formé. Il rassemblait les partenaires suivants : la **CARENE** (Communauté d'agglomération de Saint-Nazaire), **Brest Métropole**, **Harmonie mutuelle**, l'**ARS des Pays de la Loire** et **Valoway**. La **Caisse Nationale d'Assurance Maladie** a rejoint cette liste lors d'une deuxième phase des travaux.

La première réunion a eu lieu en décembre 2019. De nombreuses pistes de travail ont été explorées : les pathologies qui affectent le plus les QPV par rapport aux autres quartiers, les habitudes alimentaires des habitants des QPV et leur santé, les pratiques sportives des enfants dans les QPV et leur santé, la réalité des déserts médicaux dans les QPV, l'accès aux soins...

À l'issue de cet atelier, **trois thématiques principales ont été sélectionnées**.

Peut-on, grâce à des données existantes, **identifier des difficultés d'accès aux soins qui seraient spécifiques aux quartiers** de la politique de la ville (exemples : les habitants des QPV recourent-ils aux dispositifs de soins qui leur sont proposés ? Bénéficient-ils des aides prévues à leur intention ?). Peut-on **identifier des spécificités médicales propres aux QPV** (exemple : la santé mentale dans les QPV, la proportion de mortalité évitable par rapport aux autres quartiers...) ? Peut-on **identifier des caractéristiques propres aux QPV qui auraient un fort impact sur la santé** (offre alimentaire dans les QPV, la réalité de la consommation de stupéfiants, pratiques sportives...)?



Réunion du groupe de travail Santé en décembre 2019

Les membres du groupe de travail se sont réunis de nouveau en février 2020 à Saint-Nazaire. Cette réunion a été l'occasion de rencontrer les équipes de **Valoway**, le prestataire qui va réaliser les travaux de *data science*. **Et une formation sur la méthodologie et les bonnes pratiques pour la conduite de projets data a été dispensée à tous.**

Le groupe de travail a finalement retenu un objectif précis : essayer de **construire un outil qui permette de lutter contre le non-recours à la Complémentaire santé solidarité (CSS)**. L'hypothèse de départ des membres du groupe était la suivante : le taux de non-recours à la CSS est sans doute plus élevé dans les QPV que dans d'autres quartiers, ce qui est d'autant plus grave que le nombre de bénéficiaires potentiels y est certainement bien plus important. **Dispose-t-on de données (et d'outils) pour estimer les bénéficiaires potentiels de la CSS dans les quartiers prioritaires, et pour calculer ensuite le taux de non-recours ?**

Une première méthode de travail assez simple à mettre en œuvre a été proposée : puisque l'accès à la CSS dépend uniquement des ressources et de la composition du foyer, **les jeux de données concernant les revenus et les familles disponibles en open data sur le portail de l'INSEE** doivent permettre d'estimer le nombre des ayants droit à différentes échelles : métropole ou EPCI, communes, IRIS et quartier.

Une seconde étape a été proposée dès cette réunion de cadrage. Si l'on est capable d'identifier les bénéficiaires dans un quartier, est-on capable d'en déduire qui sont les « non-recourants » ? Il ne s'agirait plus de se limiter à un ou deux indicateurs mais de **faire un « portrait-robot » des bénéficiaires** grâce à **des outils de modélisation** intégrant de nombreuses variables socio-économiques ou socio-démographiques (revenu, âge, niveau de diplôme, emploi, immigration, etc.). Si cette démarche débouche sur un résultat probant, on en voit le bénéfice immédiat : des politiques de prévention du non-recours, par exemple par de la communication ciblée ou un accompagnement individuel, seront possibles.

L'équipe de *data science* a été emballée par cette perspective car les outils de modélisation sont des outils complexes très intéressants à actionner. Mais ils ont besoin, pour produire des résultats utiles, de beaucoup de données... L'étape de l'exploration des données ou *data mining*, va donc être cruciale.

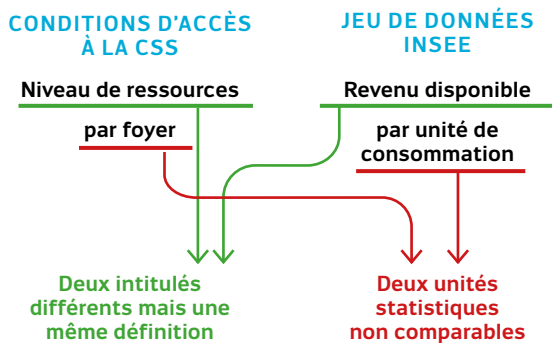
✓ Les étapes de travail (Phase 1 : estimation du non-recours)

L'obtention des données

L'expérience a d'abord été lancée pour les QPV situés dans les départements du Finistère et de la Loire-Atlantique puis rapidement étendue à l'ensemble des QPV de Bretagne et des Pays de la Loire.

La première étape visait à **déterminer le nombre de bénéficiaires potentiels de la Complémentaire Santé Solidaire (CSS) par IRIS** en application d'un critère de revenu par foyer. **Exercice a priori simple** puisqu'à cette échelle les données INSEE sont en *open data*.

Nous avons eu tort d'être optimistes. Des jeux de données se ressemblent mais ils ne racontent pas tous la même chose. Entrons un peu dans la technique. L'INSEE met à disposition des données de « revenu disponible par Unité de Consommation ». Et les conditions d'accès à la CSS dépendent du niveau de ressources rapporté au nombre de personnes composant le foyer. Les termes « revenu disponible » employé par l'INSEE et « niveau de ressources » utilisé pour la CSS désignent la même chose. En revanche, les termes « Unité de Consommation » et « Foyer » sont différents.



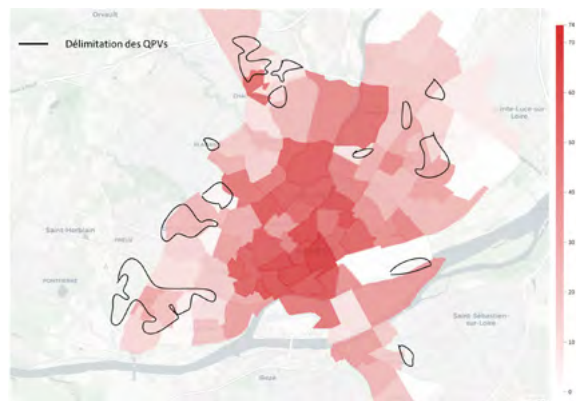
Il a donc fallu mettre en place une méthode qui rapproche les unités de consommation (UC) et les foyers (ou ménages) en intégrant d'autres données INSEE sur le détail de la composition des foyers par IRIS. Et ceci pour chaque IRIS.

Les modalités de ce calcul sont détaillées en annexe à partir de **l'exemple de l'IRIS Graslin Commerce à Nantes**. Sur ce quartier nous avons pu déterminer qu'il y a **493 bénéficiaires potentiels** sur une population de 4765 habitants. Et il va maintenant être possible de croiser ce nombre avec le chiffre réel des bénéficiaires,

une donnée fournie par la Caisse Nationale d'Assurance Maladie et disponible sur le portail de l'INSEE. Étonnant : dans ce quartier du centre-ville de Nantes, il n'y a que 110 bénéficiaires sur une estimation de 493 bénéficiaires potentiels, soit un **taux de non-recours de 78 % !**

La preuve de concept

La description de ces premières étapes peut donner l'impression d'un processus artisanal et laborieux mais l'importation des données de tous les IRIS de Bretagne et des Pays de la Loire et l'utilisation d'outils de traitement automatisés (ici le logiciel DATAIKU) permettent d'**obtenir rapidement une idée précise des zones dans lesquelles le non-recours est important**. Et les premières cartes qui en découlent sont riches d'enseignements.

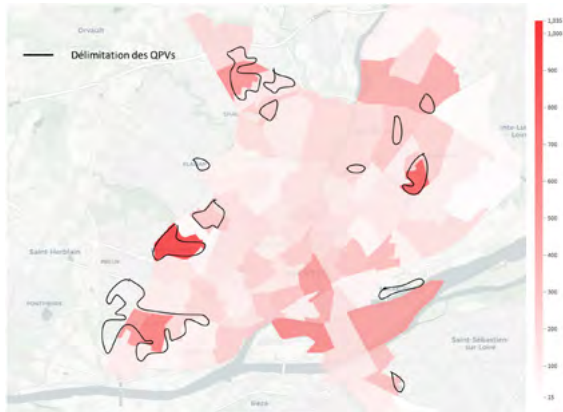


Cartographie de l'estimation du taux de non-recours à la CSS à Nantes (en % par zone IRIS)

La représentation des estimations du taux de non-recours appliquée à la ville de Nantes nous réserve une surprise : de façon schématique le taux de non-recours est très élevé... dans les quartiers les plus riches. A l'inverse, dans les QPV les personnes éligibles font valoir leurs droits dans des proportions bien plus importantes. Ce constat est, de premier abord, contre-intuitif. Il contredit l'hypothèse du groupe de travail selon laquelle « *le taux de non-recours à la CSS serait plus élevé dans les quartiers politique de la ville que dans les autres quartiers* ». Et il démontre peut-être que **dans les quartiers les personnes en difficulté sont plus et mieux accompagnées qu'ailleurs**. Il reviendra aux experts de la santé publique et aux acteurs de la politique de la ville de valider ou non cette interprétation.

Mais nous n'en sommes pas restés là.

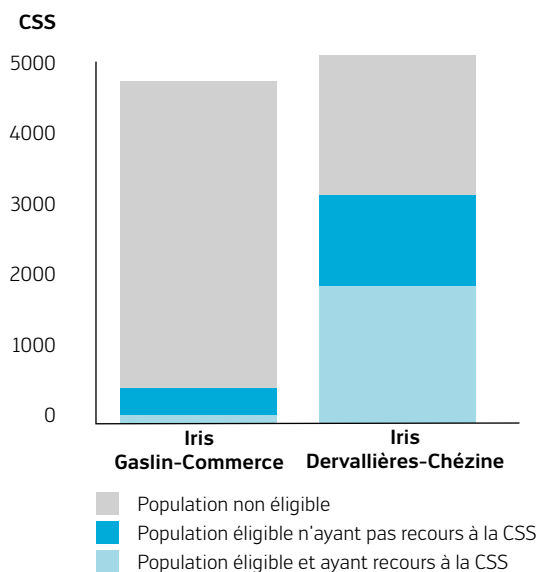
Une seconde carte représente une estimation des non-recourant non plus en pourcentage mais en valeur absolue, c'est-à-dire en nombre de personnes.



Cartographie de l'estimation du non-recours à la CSS à Nantes (en valeur absolue)

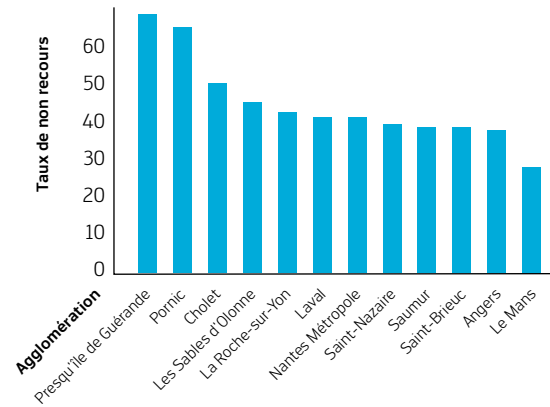
Et c'est bien dans les QPV que se trouve le nombre le plus élevé de bénéficiaires potentiels n'ayant pas recours à la CSS... Il faut donc nuancer notre hypothèse précédente : si les habitants des QPV sont mieux informés de leurs droits, le nombre de bénéficiaires potentiels étant plus élevé, c'est bien dans ces quartiers que l'on retrouve aujourd'hui encore le plus gros contingent de non-recourants.

Pour bien comprendre la manière de lire ces données, voici une autre illustration concernant deux quartiers nantais. Dans l'IRIS Gaslin-Commerce, quartier du centre-ville avec un haut niveau de revenu, il y a un faible pourcentage de bénéficiaires de la CSS au regard de ce qui devrait être. Mais au final, le non-recours concerne assez peu de personnes (plus de 300 quand même !). Dans la zone Dervallières-Chézine, quartier à faible niveau de revenu, le pourcentage de recours est bien meilleur. Mais près de 1000 personnes ne font pas valoir leurs droits !



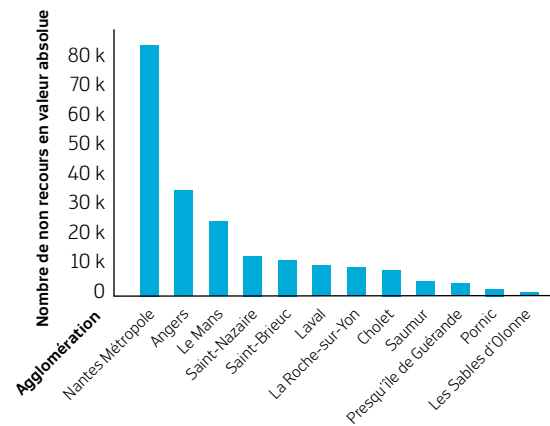
Comparaison du recours à la CSS dans deux quartiers de la ville de Nantes

En agréant les données fournies à l'échelle des codes IRIS, il devient possible de réaliser d'autres types de graphiques et de comparaisons, aux échelles des villes, des agglomérations, des départements, des régions, etc. L'équipe de RésO Villes a ainsi pu comparer des taux de non-recours par agglomération et EPCI.



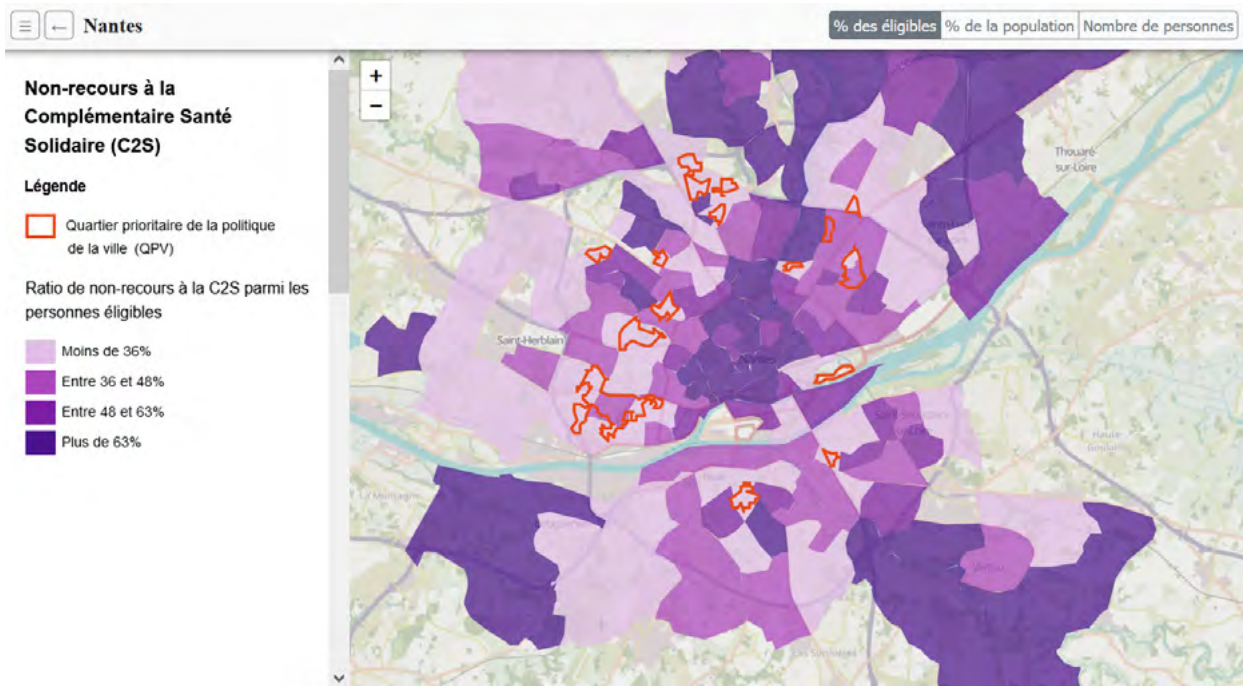
Comparaison des taux de non-recours à la CSS par agglomération sur les régions Bretagne et Pays de la Loire

Et le phénomène entrevu précédemment à l'échelle des IRIS à l'intérieur de la ville de Nantes semble se répéter à d'autres échelles. Ainsi par exemple, la ville de Pornic qui compte parmi celles avec le plus fort niveau de revenu en Pays de la Loire présente le deuxième taux de non-recours le plus élevé parmi les agglomérations comparées, mais un nombre de non-recourant en valeur absolue très limité.



Comparaison du non-recours à la CSS en valeur absolue par agglomération sur les régions Bretagne et Pays de la Loire

Ces premières analyses sont intéressantes pour les acteurs locaux. Elles donnent des premières indications sur des stratégies possibles de lutte contre le non-recours ou sur des moyens de communication à renforcer.



Capture d'écran de l'outil de visualisation géographique du non-recours à la CSS

Le prototype Data

À partir des traitements de données réalisés par Valoway, une cartographie web interactive a été réalisée⁷. Elle a été mise à disposition des acteurs de tous les quartiers de Bretagne et des Pays de la Loire et rendue publique.

➡ **ELLE EST DISPONIBLE SUR LE SITE DE RÉSO VILLES :**

<https://www.resovilles.com/programme-sante-non-recours-a-la-complementaire-sante-solidaire-c2s/>



Flashez le QR code

✓ Les étapes de travail (Phase 2 : modélisation des profils)

La première phase de travaux a déjà été riche d'enseignements. Il a donc été décidé de poursuivre et d'engager la seconde phase : celle de l'identification des profils des personnes qui n'ont pas recours à la CSS. Pour cela, la première action vise à connaître les bénéficiaires actuels. Si possible quartier par quartier.

L'obtention des données

Les caractéristiques individuelles des bénéficiaires de la CSS ne sont pas disponibles, et c'est normal, en *open data*. Nous recherchons des données précises et les plus nombreuses possible pour savoir qui sont les bénéficiaires : leur âge, leur genre, la composition de leur foyer, leurs revenus, leur catégorie socio-professionnelle, leur niveau de diplôme... Et nous voulons être en mesure de les croiser entre elles. En effet, s'il est possible de définir le profil type des bénéficiaires de la CSS dans un quartier et que nous savons qu'il y a quelques centaines de bénéficiaires potentiels qui n'y ont pas recours, alors il sera possible de « cibler » ceux qui ont des profils similaires.

Les *data scientists* ont l'habitude de ces méthodes qui visent à rechercher des « semblables » (en anglais des *look alike*). C'est ainsi que procèdent des

⁷ Les cartes du programme « Data & Quartiers » ont été réalisées par Adrien Pavie, géomaticien.

APARTÉ :

La difficulté à obtenir des données à l'échelle des QPV

La présentation de ce prototype est l'occasion de **s'arrêter sur une difficulté importante rencontrée tout au long du programme « Data & Quartiers » : l'impossibilité d'accéder à des données par QPV ou de redécouper les IRIS en fonction des périmètres des QPV.**

Certes, il existe des données statistiques par quartier. Elles sont notamment produites par l'INSEE à la demande de l'ANCT ou de l'Observatoire national de la politique de la ville. Mais elles ne concernent pas toutes les données disponibles à l'échelle des IRIS, loin s'en faut. Et encore moins les données nouvelles que nous avons souhaité utiliser.

Durant les deux dernières années du programme, nous avons recherché un outil permettant de convertir des données ou des indicateurs fournis par IRIS en données et indicateurs à l'échelle des QPV. Chaque QPV est en effet composé d'un pourcentage de 1, 2, 3 ou plus IRIS. Nous avons officiellement saisi le ministère de la Cohésion des territoires et notamment son administrateur ministériel des données au titre des obligations d'*open data*. Peine perdue, non pas par absence de volonté, mais parce que cet outil semble ne pas exister. Ne pas... ou ne plus exister ? Car il a bien fallu construire des indicateurs par quartier lors

des précédents redécoupages. Et nous comprenons bien, à travers ce constat, pourquoi un récent rapport officiel dénonce « *le voile d'ignorance* » qui pèse sur la connaissance des quartiers¹.

L'équipe de RésO Villes ainsi que ses partenaires CIVITEO et Valoway n'ont pas baissé les bras. Avec l'aide d'un nouveau partenaire, la société Poligma, nous avons essayé de reconstruire une grille permettant de définir un pourcentage de chaque IRIS à affecter à chaque quartier. Géographiquement les pourtours des quartiers sont connus et des outils permettent des approximations, quitte à opérer ce découpage manuellement pour les QPV de Bretagne et de Pays de la Loire (après tout, il n'y a que 78 quartiers !). Mais l'équation est plus complexe : les répartitions doivent prendre en compte par exemple le type d'habitat car, c'est une évidence, il y a plus d'habitants dans une tour que dans les pavillons voisins qui sont dans le même IRIS. Après plusieurs tentatives intégrant des données multiples (y compris des données fiscales et foncières disponibles en *open data*) nous avons renoncé².

Nos travaux se poursuivent à l'échelle des IRIS et nos cartes et nos conclusions, comme pour le COVID ou la C2S englobent les quartiers sans toujours pouvoir les isoler aussi parfaitement que nous l'aurions voulu.

¹ Rapport de la Commission nationale chargée de la réflexion sur les prochains contrats de ville, avril 2022

² Cet encadré résume un sujet important lié au manque d'outils de connaissance des quartiers. Le détail de la méthodologie et des tentatives faites durant le programme pour y remédier figure en annexe 2.

grandes enseignes en matière de marketing : elles cherchent à avoir un maximum d'information sur leurs clients pour ensuite communiquer vers tous ceux qui leurs ressemblent... Ici, « **Data & Quartiers** » **détourne des outils inventés par des entreprises commerciales pour les mettre au service de l'intérêt général.**

RésO Villes a engagé un dialogue avec son partenaire **Harmonie mutuelle** (qui est l'un des gestionnaires de la CSS) et s'est ensuite tourné vers la **Caisse Nationale d'Assurance Maladie (CNAM)**, qui dispose des données sur les bénéficiaires. Mais attention, il s'agit là de données personnelles ! Et il était exclu

que ces données qui concernent des habitants des quartiers, qui plus est bénéficiant d'une aide sociale essentielle, puissent être utilisées.

Il a fallu près de **six mois d'échanges et de contrôles (notamment juridiques) pour finaliser une convention de mise à disposition de données** signée par les deux parties. Cette convention prévoyait la mise à disposition de données sur les bénéficiaires de la CSS agrégées quartier par quartier (ce qui rend impossible leur identification). Mais les Caisses primaires d'assurance maladie qui disposent localement de ces données, ne les classent pas par QPV... Un important travail a été engagé pour **développer des**

requêtes informatiques permettant d'extraire les données par QPV à partir de l'adresse des bénéficiaires. Nouvelle difficulté : il a été impossible d'obtenir la liste des voies postales constitutives des QPV. Cette donnée indispensable pour réaliser l'extraction n'existe ni en *open data*, ni auprès de l'Observatoire National de la Politique de la Ville. Il a fallu reconstituer, pour 10 quartiers test, ces listes à la main. Comme quoi, **la data science c'est aussi parfois du système D !**

La preuve de concept

Préparation des données et choix des algorithmes

L'équipe de Valoway, partenaire en charge de la *data science*, a commencé par explorer quatre types de données fournies pour chacun des 10 QPV ciblés et représentant un total de 11 648 foyers. Il s'agissait du nombre de personnes par foyer, du type de contrat de complémentaire santé (CSS gratuite ou participative), des tranches d'âges et de la répartition des ressources du foyer par déciles.

La construction d'un modèle a amené l'équipe à transformer certaines données fournies par classes en

moyenne, par exemple : calcul d'un revenu moyen par unité de consommation à partir de la répartition des ressources du foyer par déciles, calcul d'un âge moyen du foyer à partir de la répartition des bénéficiaires par tranches d'âges, etc... L'objectif est de disposer de variables pertinentes pour faire une « segmentation ».

Pour essayer d'identifier des « profils types », nous avons eu recours à des algorithmes, c'est-à-dire des suites logiques de calcul capables d'extraire des modèles à partir de données massives. C'est ainsi par exemple qu'Amazon dispose de plusieurs dizaines de « clients types » pour « cibler » les visiteurs de son site internet.

« Data & Quartiers » ce n'est pas Amazon, mais nous avons aussi des algorithmes à notre disposition. Encore faut-il choisir le bon. C'est-à-dire **un algorithme performant et interprétable**. L'interprétabilité d'un modèle désigne le degré de facilité avec lequel un humain est en mesure de comprendre les décisions prises. Il s'agit d'éviter « l'effet boîte noire » observé lorsqu'un modèle ne fournit pas clairement d'explications sur le processus de transformation des données d'entrées, ce qui arrive par exemple lorsque l'on a recours à de l'intelligence artificielle.

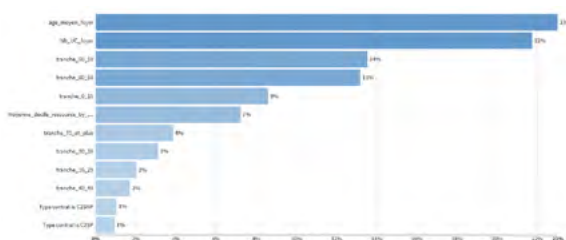


ENCADRÉ TECHNIQUE : L'importance des variables pour la segmentation

Quelques explications pour les curieux et les spécialistes.

Pour préparer les jeux de données, il a fallu prendre en compte plusieurs éléments.

L'importance des variables : quel poids pèse la variable dans la création des différentes classes de la segmentation ? Par exemple la variable « type de contrat » avait peu d'incidence et son « impact » a été estimé à moins de 1 % .



Analyse de l'importance des variables pour la segmentation des bénéficiaires de la CSS

La corrélation des variables : quelle est la proximité entre deux variables envisagées pour la segmentation ? Des variables avec une corrélation trop importante peuvent apporter de l'information redondante et ainsi fausser l'interprétation du modèle. Par exemple, les variables « ressources par foyer » et « ressources par unités de consommation », directement corrélées, ne peuvent pas être toutes les deux conservées.



Analyse de la corrélation des variables pour la segmentation des bénéficiaires de la CSS



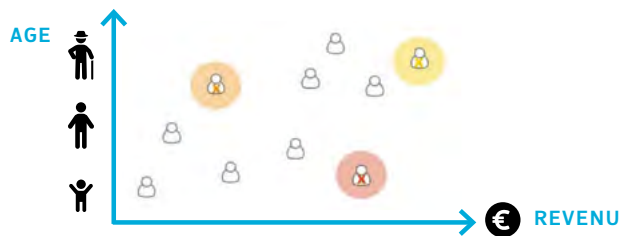


ENCADRÉ TECHNIQUE : Fonctionnement de l'algorithme K-Means

L'algorithme K-Means est un algorithme de classification qui permet d'identifier des familles (appelées des « classes ») au sein d'une population importante caractérisée par de nombreuses données.

Prenons l'exemple d'une segmentation en 3 classes réalisée sur un ensemble de 10 individus et basée sur deux critères seulement (l'âge et le revenu). La première étape dite d'« initialisation » consiste à choisir de façon aléatoire 3 individus qui vont représenter nos 3 classes.

Étape 1 : initialisation



Pour chacun des 7 individus restants, on calcule la proximité par rapport aux 3 représentants des 3 classes choisis aléatoirement lors de l'étape précédente.

Étape 2 : calcul des distances



Chaque individu est ensuite affecté à la classe dont il est le plus proche.

Étape 3 : affectation aux classes



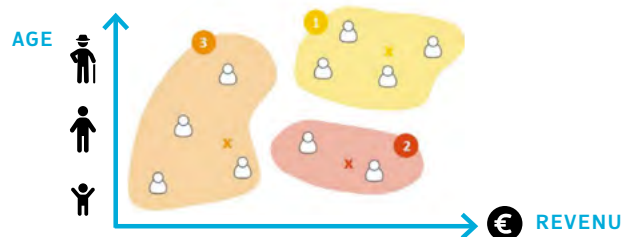
Une fois que tous les individus ont été affectés à l'une des trois classes de départ, on calcule le centre de gravité de chaque classe :

Étape 4 : centre de gravité



On répète alors l'étape de calcul des distances pour chaque individu, cette fois par rapport au centre de gravité de chacune des 3 classes. Les individus sont éventuellement réaffectés à un nouveau groupe. Le centre de gravité de chaque classe s'en trouve modifié. Ces 3 étapes (calcul des distances, affectation aux centres, calcul des nouveaux centres) sont répétées autant de fois que nécessaire jusqu'à ce qu'aucun individu ne soit plus réaffecté.

Étape 5 : itérations



La performance du modèle quant à elle, désigne la précision avec laquelle l'objectif de départ est atteint. Dans le cas d'une segmentation, cette performance est estimée à travers un indicateur appelé la **silhouette**, qui mesure lui-même la **cohésion** (proximité entre les différentes variables d'un même groupe, ou classe) et la **séparation** (distance moyenne entre les variables d'une même classe et celles des classes voisines).

Après plusieurs itérations, les *data scientists* de Valoway ont conservé quatre variables (moyenne des déciles de ressources, nombre d'unités de consommation, âge moyen du foyer, nombre de personnes dans chaque tranche d'âge) et ont fait le **choix de l'algorithme K-Means**, dont le fonctionnement est expliqué dans l'encadré de façon simplifiée.

Résultats

Les différentes itérations menées ont abouti à une segmentation en quatre classes dont les principales caractéristiques sont présentées de manière succincte dans le tableau.

AU GLOBAL en moyenne	Nb personnes foyer 2,46	Âge moyen du foyer 33,8 ans	Revenu moyen du foyer 7 538 €
Foyers avec un quinquagénaire avec peu de ressources (16%)	1,85 -24%	57% de 50-59 ans vs 8% au global	6 651 € <small>(une répartition similaire au global)</small>
Personnes âgées avec ressources moyennes (15%)	1,15 -53%	91% de plus de 60 ans vs 8% au global	65% dans les déciles 7 et 8 vs 20% au global
Familles nombreuses avec ressources les plus élevées (24%)	4,98 +102%	58% de moins de 16 ans vs 40% au global	51% dans les déciles 9 et 10 vs 20% au global
Jeunes avec ressources les plus faibles (45%)	1,78 -27%	32% de 16-29 ans vs 20% au global	51% avec 0 ressources vs 40% au global

Résultats de la segmentation des bénéficiaires de la CSS pour les 10 QPV traités

Comment lire ce tableau ? 45 % des bénéficiaires de la CSS dans les 10 QPV étudiés appartiennent à la classe 4 (dernière ligne). La composition moyenne des foyers dans cette classe 1 est de 1,78 personnes, 27 % en dessous de la moyenne de la population totale étudiée (2,46 personnes par foyer). 32 % ont entre 16 et 29 ans alors que seulement 20 % de la population totale étudiée se situe dans cette même tranche d'âge. 51 % d'entre eux n'ont pas de ressources.

Pour chacune des « classes » nous avons un intitulé possible. Les 4 grandes catégories de bénéficiaires de la CSS sont les suivantes :

- « Foyers avec un quinquagénaire avec peu de ressources »
- « Personnes âgées vivant seules avec des ressources moyennes »
- « Familles nombreuses avec souvent deux parents trentenaires et entre deux et trois enfants en bas âge (moins de 16 ans) »

- « Jeunes de moins de 30 ans vivant seuls ou en couple avec les niveaux de ressources les plus faibles dont une majorité de revenus nuls »

Cette segmentation expérimentale mériterait d'être améliorée en obtenant des données à un niveau de granularité plus fin ou en y ajoutant de nouvelles données. C'est cependant **un très bon point de départ pour mieux connaître et donc potentiellement aller à la rencontre des « non-recourant » à la CSS.**

Se préparer à l'échec...

Nous souhaitons franchir une étape suivante en identifiant quartier par quartier le nombre de bénéficiaires potentiels par catégorie.

L'impact de cette démarche est potentiellement très fort. Imaginons dans un quartier comment les acteurs du territoire (la CAF, les mutuelles, le CCAS...) pourraient agir s'ils savaient que par exemple dans tel quartier il y a probablement XX foyers de jeunes de moins de 30 ans en situation de non-recours. Une campagne d'information spécifique peut leur être destinée !

Très vite, nous avons collectivement pris conscience de l'impossibilité de réaliser cette projection, du moins à court terme. Pour une raison simple : les données ne sont pas disponibles. Une explication s'impose. Il aurait fallu avoir accès aux mêmes croisements de données que ceux ayant été à l'origine de la segmentation, c'est-à-dire connaître pour chaque QPV le nombre de personnes présentes parmi 294 combinaisons (7 tranches d'âge x 7 tranches de compositions de foyer x 6 tranches de revenus)... Nos interlocuteurs de l'Agence Nationale de la Cohésion des Territoires (ANCT) nous ont précisé que ces données étaient sans doute techniquement accessibles. Mais il aurait fallu anticiper notre demande pour organiser, non pas la mise à disposition des jeux de données, mais leur production à partir des données sources de l'INSEE.

Nous ne désespérons pas, donc, de pouvoir donner suite à ces travaux dans le futur, mais nous avons dû nous résigner à abandonner au stade de la segmentation cette deuxième phase de travaux sur le non-recours à la CSS.

APARTÉ :

L'importance d'anticiper l'accès aux données

La mise à disposition de données par certaines organisations (publiques ou privées) relève parfois d'un véritable parcours du combattant. La bonne volonté et l'envie de collaborer ne suffisent pas. Tous les partenaires du programme « Data & Quartiers » avaient la volonté de contribuer à ce projet expérimental. Mais les processus sont parfois longs. Ils doivent être anticipés et pris en compte au moment de réfléchir au planning. Les *data scientists* en charge de « faire parler » les données risquent de se retrouver sans matière première... Or, ce sont des compétences coûteuses (un *data scientist* expérimenté est facturé au minimum 800 € HT la journée) et il convient de les faire intervenir à bon escient.

Voici quelques enseignements du programme « Data & Quartiers » pour préparer et accélérer l'étape d'obtention des données.

1. Identifier les décideurs et les convaincre en amont du bien fondé de votre démarche et des apports potentiels pour leur organisation.
2. Se renseigner en amont sur les processus internes de validation de la mise à disposition des données.
3. Proposer un cadre juridique robuste (par exemple un modèle de convention de mise à disposition de données) qui anticipe les difficultés et tout particulièrement la protection des données personnelles
4. Lister les dispositions mises en œuvre pour assurer un traitement correct des données (hébergement, personnes habilitées, usages, limitation dans le temps et éventuelle destruction, etc.)
5. Mettre en place une certification éthique de votre démarche data pour offrir des garanties, et donc rassurer, les partenaires (exemple : label Ekitia¹)

¹ Voir ci-après le processus de labélisation éthique du programme « Data & Quartiers »

Data & Mobilités dans les quartiers : une étude sur les modalités de déplacement Domicile-Emploi

[+] Le cadrage du projet

Les premiers travaux sur l'emploi avaient laissé la porte ouverte à une suite. Identifier des distances entre domicile et emploi est une chose. Savoir comment s'y rendre en est une autre.

Un groupe de travail a été constitué sur la mobilité dans les quartiers en intégrant un groupe d'étudiants du Master SIGAT (Système d'Information Géographique et Analyse des Territoires) de l'Université de Rennes 2. L'Université de Rennes 2 est aussi devenue partenaire du programme.

Lors d'une première réunion de travail, une liste de problématiques potentielles a été dressée :

- Pour chaque quartier et catégorie de métiers, quelles possibilités de transport s'offrent aux habitants (à pied, à vélo, en voiture, en transports en commun...)?
- Y a-t-il des différences d'accessibilité à l'emploi en fonction des horaires ?
- Existe-t-il des métiers plus accessibles que d'autres pour les habitants des QPV ?
- L'accessibilité à l'emploi est-elle différente entre les QPV et les autres quartiers ? Et entre les différents QPV eux-mêmes ?
- À Rennes de façon spécifique : l'ouverture de la deuxième ligne de métro va-t-elle améliorer l'accessibilité à l'emploi des habitants des QPV ?

Pour différentes raisons pratiques, et tout d'abord la certitude d'accéder facilement à de nombreuses données⁸, il a été décidé de centrer les travaux sur le territoire de Rennes Métropole.

⊙ Les étapes de travail

L'obtention des données

Une convention de mise à disposition de données a d'abord été signée entre RésO Villes et l'Université de Rennes 2 pour permettre la réutilisation des données fournies par le Groupe Randstad et Pôle emploi lors des travaux du groupe Emploi. Des échanges ont aussi

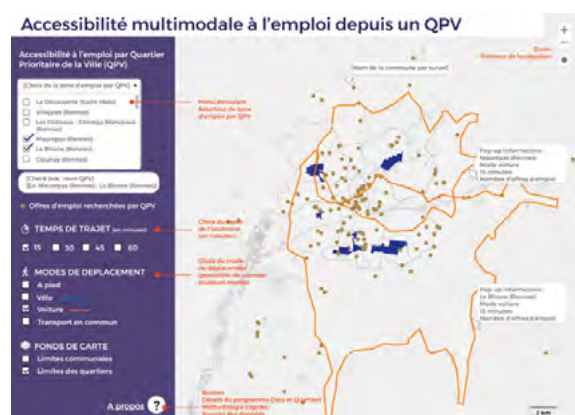
eu lieu entre les représentants du Groupe Randstad et les étudiants du Master SIGAT pour faciliter la compréhension et la prise en main des différents fichiers et des bases de données.

Les données disponibles en *open data* sur les infrastructures (par exemple les pistes et bandes cyclables) et les transports en commun (emplacements des arrêts, tracé des lignes, horaires, etc.) ont été utilisées. Ce sont des données standardisées sous un format standardisé dit « GTFS » pour *General Transit Feed Specification*.

Nous avons sollicité les équipes de Rennes Métropole pour obtenir aussi de façon anticipée les données prévisionnelles de la deuxième ligne de métro : tracé et horaires de la nouvelle ligne, mais aussi les ouvertures, fermetures et modifications des lignes de bus impactées dès l'inauguration. Cela nous aurait permis de réaliser une analyse d'impact sur certaines mobilités domicile – emploi des habitants des QPV : quel gain de temps ? Pour quel type d'offres d'emploi ? Malheureusement, ces données n'étaient pas encore disponibles au moment de la réalisation des travaux.

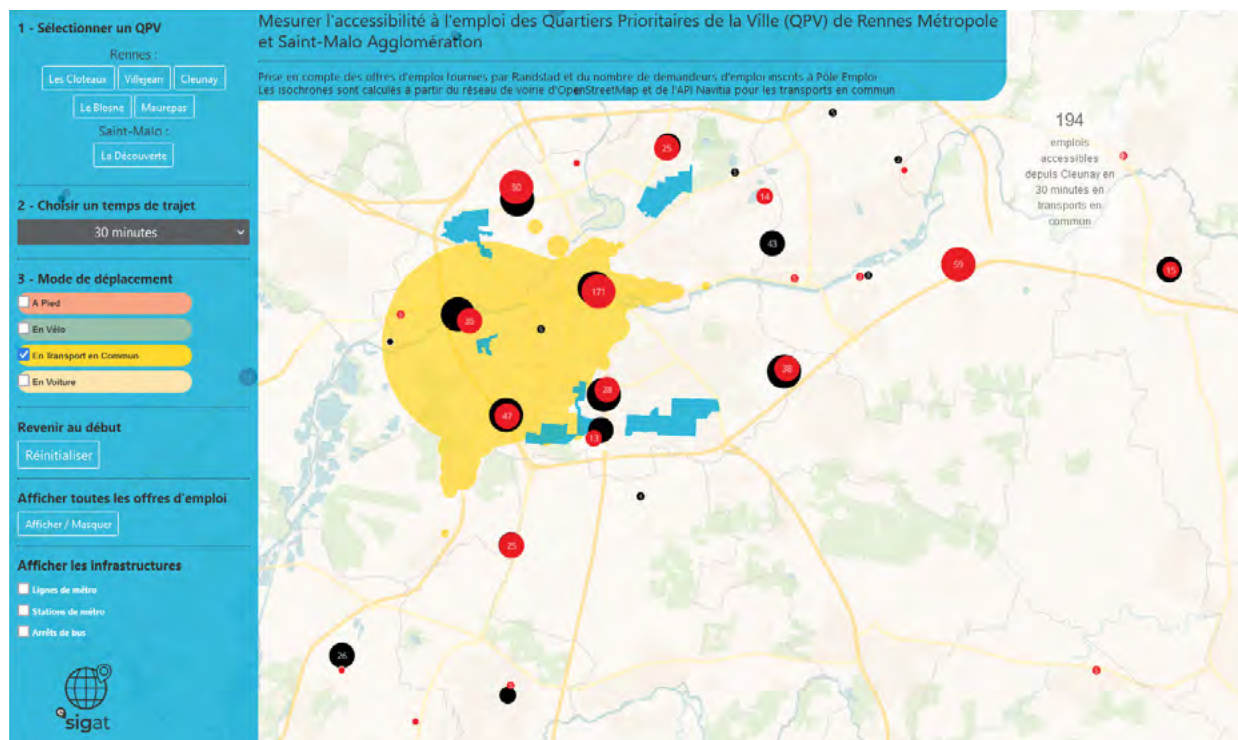
La preuve de concept

L'équipe de Rennes 2 a dans un premier temps, exploré l'ensemble des données fournies sur l'emploi disponible et les quartiers et construit une maquette d'un outil permettant de simuler les différentes manières d'accéder d'un QPV à un emploi donné.



Maquette du livrable Mobilités envisagé

⁸ Rennes est un territoire réputé pour sa longue tradition d'*open data*. La Ville de Rennes a été la première ville de France à « ouvrir ses données » en 2010.



Capture d'écran de l'outil de visualisation géographique des mobilités domicile-emploi des habitants des QPV en Ile-et-Vilaine

La première proposition reposait sur la création d'une carte dynamique permettant de sélectionner des offres d'emploi à l'intérieur d'isochrones. C'est-à-dire des tracés représentant les zones accessibles en un temps donné selon le mode de transport disponible.

Très utile pour représenter la donnée disponible, cette carte présentait malgré tout un inconvénient : elle ne permettait pas de faire des analyses sur la réalité des mobilités domicile – emploi, et notamment de faire des comparaisons entre quartiers. Il a donc été décidé de conserver l'aspect ludique et la facilité de prise en main de la première maquette et de l'enrichir d'une couche de statistiques afin de faciliter l'analyse et la prise de décision.

Le prototype Data

Le prototype livré par les étudiants du Master SIGAT est un véritable site web à l'ergonomie intuitive et conviviale comprenant un outil de géovisualisation associé à plusieurs *dataviz*. L'utilisateur a la possibilité d'interagir et de « jouer avec les données ».

VOUS POUVEZ ACCÉDER AU PROTOTYPE MOBILITÉS VIA CE LIEN :

<https://www.resovilles.com/mobilite/index.html>

Il est ainsi possible, quartier par quartier, de sélectionner des modalités de transport préférées, ou inversement de fixer une limite de temps. Les différentes infrastructures (lignes, arrêts de métro et de bus), sont représentées de façon optionnelle pour permettre à l'utilisateur de se repérer dans l'espace et objectiver la présence d'infrastructures favorisant la mobilité à proximité du QPV observé.

Il est possible de comparer les différents modes de transport depuis un même QPV. Il est aussi possible de comparer les QPV entre eux grâce à des *dataviz* complémentaires disponibles sur le site. Bref : un exemple réussi de prototype rapide susceptible de donner de nombreuses idées aux acteurs de la politique de la ville.

Mentionnons ici que les solutions techniques choisies sont *open source*, ce qui permet d'assurer la pérennité du prototype en s'affranchissant de toute dépendance à une licence tout en n'imposant pas de coût supplémentaire pour le projet⁹.



Flashez le QR code

⁹ Pour les spécialistes : librairie Javascript Chart.js10 pour la réalisation des graphiques interactifs, bibliothèque javascript MapLibre pour les cartes, HTML et CSS pour l'habillage.

Conclusion :

politique de la ville, éthique et data

À parcourir le livre blanc, le lecteur aura constaté que cette expérimentation est riche d'enseignements variés. De nombreux acteurs locaux, en Bretagne et en Pays de la Loire, dans d'autres régions également, ont sollicité les équipes de RésO Villes pour comprendre et découvrir le projet. Des villes, des associations, des conseillers citoyens ont ainsi pu prendre part à des temps de présentation et de travail sur la data et les quartiers.

Des acteurs nationaux s'en sont aussi saisi : l'ANCT, qui a été très présente durant 3 années, et Pôle emploi, le Health Data Hub, la CNAM, l'INSEE...

Généraliser « Data & Quartiers » ?

La commission nationale chargée de la réflexion sur les prochains contrats de ville nous a également sollicités. Dans son rapport publié en avril 2022¹⁰, elle écrit quelques vérités crues sur la manière dont sont observés les quartiers. Elle dénonce l'absence de données nécessaires à la territorialisation des diagnostics. Il s'agit du même enjeu que celui auquel nous avons été confrontés en cherchant à reventiler des données à l'échelle des quartiers. La commission évoque même la nécessité de « lever la voile d'ignorance » (sic !) qui pèse sur la connaissance des quartiers.

Elle salue l'expérimentation « Data & Quartiers », notamment parce que nous avons pu montrer que des données privées qui représentent un intérêt général peuvent utilement venir compléter les données publiques disponibles. Et la commission va plus loin en proposant de « faire de la data science un outil de connaissance territoriale et de territorialisation des politiques publiques au bénéfice des quartiers » et en recommandant « d'intégrer dans les prochains contrats de ville des clauses data ».

Il s'agirait alors d'expérimenter à plus grande échelle « Data & Quartiers ». Acceptons-en l'augure !

Des enjeux nombreux

Passer d'une expérimentation comme celle-ci à une utilisation efficace et généralisée des données au service des acteurs de la politique de la ville ne se fera pas si facilement. L'expérience des métropoles pionnières en matière de gestion des données (pensons à Rennes et Nantes par exemple) peut nous être utile. Des nombreux « prototypes » restent parfois dans les cartons. Il faut réunir des conditions bien particulières pour installer dans la durée un pilotage avec des données.

Il faudra d'abord s'assurer que les données nécessaires qui ont été mises à disposition de façon expérimentale pourront l'être de façon pérenne. Cela peut prendre du temps. Et il y aura un coût. Ceci vaut aussi bien pour des acteurs publics (pensons par exemple aux extractions qui ont été faites par la CNAM) que des acteurs privés (le groupe Randstad n'a sans doute pas vocation à offrir ses données pour tous les QPV de France sans contrepartie).

Il faudra aussi que des acteurs soient formés et puissent les utiliser. L'étape d'acculturation, si importante dans le programme « Data & Quartiers » devra être organisée partout, auprès des élus, des agents territoriaux ou des acteurs associatifs. Et devra parfois être répétée.

Il faudra construire des outils techniques pour que la data science soit rendue accessibles aux acteurs de terrain : des interfaces, des flux et des tableaux de bord automatisés, des cartes interactives, des outils d'analyse « clef en main »... Tout ceci a un coût. L'intégration de ces coûts dans les prochains contrats de ville semble une nécessité.

Il faudra aussi prendre en compte des enjeux démocratiques et politiques. Une collectivité territoriale, un ministère, un centre de ressources... ce n'est ni Google, ni Facebook, ni Amazon ! L'utilisation massive des données mise au service de l'intérêt général se doit de respecter des règles qui vont bien au-delà du simple respect de la loi, notamment en matière de protection des données. Des acteurs engagés dans des politiques qui concernent des quartiers et des publics sensibles doivent se doter d'une éthique de la donnée.

¹⁰ Pour un acte II de la politique de la ville, Ministère chargé de la ville, avril 2022.

Une indispensable éthique de la donnée

Nous avons gardé ce point pour la conclusion. Et pourtant il était présent dès les premiers moments du programme expérimental. Quelles données allons-nous utiliser ? Y a-t-il des risques d'atteinte à la vie privée ? Ou de stigmatisation des quartiers et de leurs habitants ?

L'équipe de RésO Villes, CIVITEO, Valoway, les partenaires publics et privés ont été « obsédés » par ces enjeux. Il y a tant d'exemples de discriminations au quotidien, comment éviter que le recours aux nouveaux outils numériques en rajoute ? Les données de nos travaux présentent-elles des risques ? On parle de santé, d'emploi, d'origines, d'aides sociales... Tous les professionnels savent que ces données sont sensibles.

Nous avons veillé à mettre en place les meilleures pratiques possibles tout au long du cycle de vie des données utilisées. Les processus ont été détaillés dans le livre blanc : signature de conventions, contrôle du respect du RGPD, secret statistique, délimitation des usages autorisés, personnes habilitées...

Mais les enjeux sont plus larges : il faut porter une attention particulière au stockage des données, aux étapes de mise en qualité des données, au choix des algorithmes (leur explicabilité, leur robustesse, l'absence de biais ou autres dérives), à la sobriété énergétique des diverses solutions retenues, à la transparence et l'inclusion des citoyens, etc.

Au fil du programme nous avons mis en œuvre, de façon empirique d'abord puis de façon systématique, des décisions qui au final constituent un véritable corpus éthique.

Une démarche reconnue et labellisée

Très rapidement nous avons souhaité formaliser la démarche. La première étape a été de s'approprier et de signer **la Charte métropolitaine de la donnée de Nantes métropole**. Cette charte a été, en 2019, la première charte éthique relative à la gestion des données mise en place en France par une collectivité territoriale¹¹. Elle fait aujourd'hui autorité et elle a inspiré de nombreux autres territoires, y compris à l'étranger. Elle fixe des principes sur les garanties apportées quant à l'accès à des données publiques ou d'intérêt général, la protection des données personnelles ou la transparence par exemple. Elle est ouverte aux acteurs publics mais aussi privés et associatifs.

¹¹ <https://metropole.nantes.fr/charte-donnee>

¹² www.ekitia.fr/accueil/



Mais nous avons envie d'aller plus loin. **RésO Villes s'est rapproché de l'association EKITIA¹²** qui œuvre à la création d'un cadre de confiance éthique et souverain pour le partage de données. Et **qui délivre un label éthique à l'issue d'un processus de contrôle et d'audit rigoureux.**



Le programme « Data & Quartiers » a été audité par un évaluateur indépendant, sur plus d'une centaine de points concernant nos règles, nos objectifs, nos outils, nos méthodes, nos erreurs et leurs corrections...

À l'heure où ces lignes sont rédigées, nous venons d'apprendre que le label EKITIA est décerné au programme « Data & Quartiers ».

Voici une conclusion très satisfaisante...

Paroles d'acteurs

 **PARTIE 3**

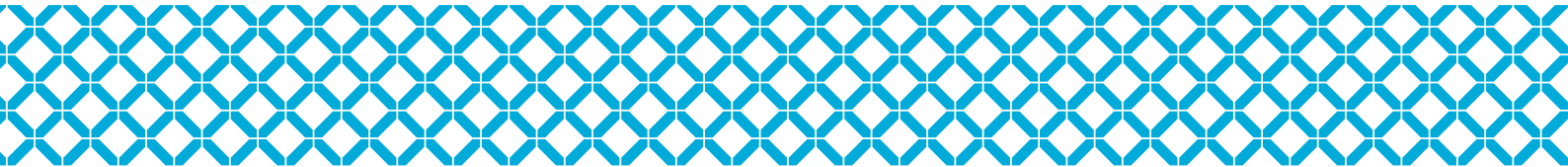




Aminata Keita

Christophe Girard

**Grande équipe de la réussite républicaine,
Direction déléguée à la Politique de la ville,
Agence nationale de la cohésion des territoires**



L'Agence nationale de cohésion des territoires a soutenu le programme « Data & Quartiers » dès son lancement. Quelles sont les raisons de cet engagement fort aux côtés de RésO Villes ?

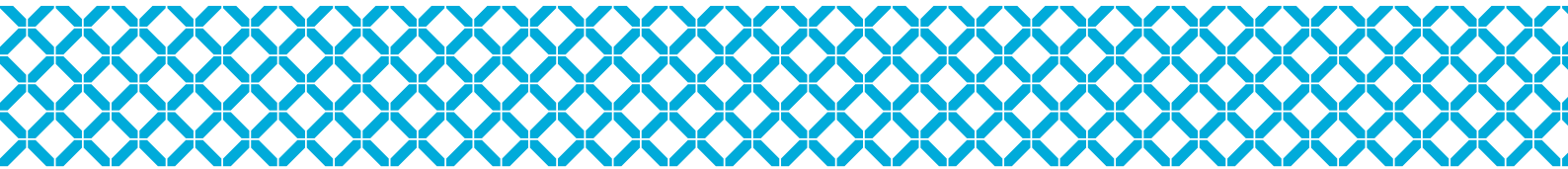
Les centres de ressources de la politique de la ville, dont RésO Villes, sont des acteurs essentiels de la mise en œuvre des politiques publiques à destination des habitants des quartiers. Ce sont aussi des acteurs de la compréhension de la vie des quartiers et de leur évolution. L'ANCT, dans le fil des administrations nationales chargées de la politique de la ville qui l'ont précédée, entretient donc des liens étroits avec ces structures.

Le CGET (devenu depuis l'ANCT) avait pris part au forum « Cités Cap' » dont l'histoire est retracée dans le livre blanc, et avait compris dès 2018 l'intérêt de conduire une expérimentation pour montrer comment l'exploitation des data peut avoir un impact positif sur la conduite des politiques publiques. L'ANCT s'est donc engagée dès le début du programme en 2019 avec le souhait de modéliser une démarche et de produire des outils transférables pour leur essaimage sur l'ensemble du territoire national.

Pouvez-vous nous expliquer en quoi la gestion des données est devenue une priorité de l'ANCT ? Et comment les data peuvent avoir un impact sur les politiques menées par le ministère ?

Au sein de l'Agence, nous sommes persuadés que les innovations nées des nouveaux usages de la donnée peuvent apporter des réponses à des problématiques spécifiques vécues au sein des quartiers de la politique de la ville. Nous pensons aussi que ces données et ces outils transforment l'action publique ainsi que nos pratiques professionnelles.

L'ANCT accompagne les collectivités et leurs partenaires dans la mise en place de « communs » au service de la transformation numérique des territoires ;



avec l'ambition d'assurer un rôle de « fabrique à projets » pour permettre aux collectivités de mener à bien leurs projets.

Quels enseignements principaux tirez-vous de ces trois années de programme « Data & Quartiers » ?

Nous avons eu la confirmation que la transformation numérique des territoires, la gouvernance par la donnée ou l'innovation territoriale sont des enjeux importants qui nécessitent de franchir collectivement certaines étapes. Il est ainsi essentiel que les acteurs de l'urbain et de la cohésion sociale développent des compétences spécifiques et une culture commune. Il faut également que les organisations s'adaptent pour que l'ensemble des parties prenantes coopère, notamment en partageant leurs données.

D'un point de vue très pratique, ce programme conforte un modèle « ascendant » de la politique de la ville que l'on promeut. Il faut partir des problématiques des territoires et des usagers pour mieux adapter et territorialiser les politiques publiques. Et pour cela il faut disposer de données.



Jean-Aimé Rakotoarisoa

Conseiller citoyen du quartier Malakoff (Nantes)



En tant que conseiller citoyen, vivant au cœur des réalités du quartier, de quel œil voyez-vous l'arrivée de ces nouveaux outils et de ces nouvelles méthodes qui utilisent massivement des données ?

Je considère que c'est une des manières possibles pour améliorer la compréhension de la dynamique sociale d'un quartier comme Malakoff. Ces nouvelles approches vont aider à aboutir à une vision beaucoup plus claire des réalités, et permettre une meilleure coordination des actions. Dans un quartier, il n'y a jamais de situation figée, c'est donc important d'avoir des outils qui intègrent ces changements en temps réel pour faire des analyses et des projections plus faciles et rapides. Encore faut-il savoir ce que l'on cherche. Il faut choisir des sujets précis. Il faut aussi surmonter les blocages, conscients ou inconscients, sur ces questions. Et il faut bien sûr encadrer les démarches.

Concrètement, qu'est-ce que les data pourraient apporter au quotidien ?

Cela pourrait permettre d'avoir une cartographie précise des strates sociales du quartier, pour identifier les problématiques et se focaliser sur les gens les plus à risques et/ou fragiles. Ce qui est important c'est de comprendre l'impact direct sur les gens. Dans quelle mesure, nos actions améliorent vraiment le quotidien des habitants et des habitantes de nos quartiers. Au-delà des statistiques, le vécu et le ressenti constituent des facteurs pas toujours mesurables mais qui sont aussi en prendre en considération.

Pensez-vous que les citoyens d'une manière générale sont suffisamment informés sur les sujets autour de la data ?

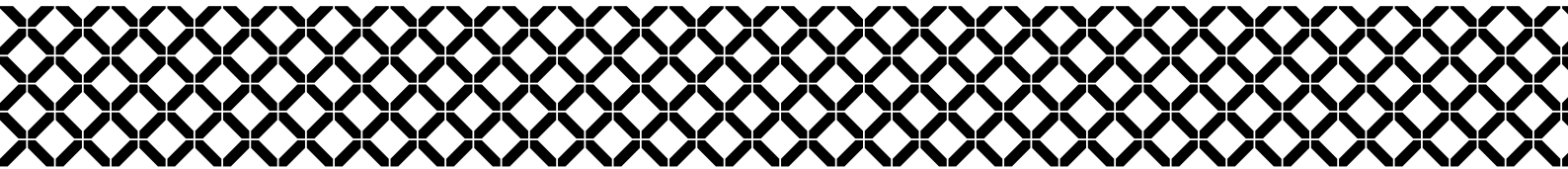
Tous les citoyens ne sont pas sur le même plan. Certains cherchent à s'informer et veulent pouvoir s'en servir. En revanche, les classes sociales défavorisées, qui devraient le plus bénéficier de nos actions sociales, restent souvent en dehors des circuits d'information. Préoccupés avant tout par les besoins du quotidien ils restent en marge de ces questions. Ils vivent au jour le jour et ne perçoivent pas les changements qui pourraient leur être bénéfiques. Ils ne sont plus « dans la course ».

Il est urgent de remettre tout le monde au même niveau afin d'opérer une prise de conscience que toutes les actions, même les plus petites, peuvent améliorer les quotidiens.



Benoît Blondet

Directeur Inclusion protection Care,
Harmonie Mutuelle



Harmonie Mutuelle est devenu partenaire du programme « Data & Quartiers » dès son démarrage. Pouvez-vous nous expliquer les raisons de cet engagement ?

Lorsque la proposition nous a été faite de rejoindre le programme « Data & Quartiers », nous avons pensé que cela pouvait être mis en perspective avec notre engagement d'entreprise mutualiste à mission et nos valeurs de solidarité et de contribution à l'intérêt général.

Nous avons aussi rejoint le programme car il répondait à un enjeu pour Harmonie Mutuelle : avoir une connaissance plus fine et ancrée dans les territoires du phénomène du non-recours à la Complémentaire Santé Solidaire, dont nous accompagnons le déploiement.

Concrètement, quelle a été votre contribution au programme ?

Nous avons apporté notre connaissance de la CSS et soumis le sujet du non-recours au groupe de travail. Nous avons apporté notre œil d'expert sur le sujet pour aider à cadrer les travaux de *data science*. Enfin, nous avons contribué à embarquer la Caisse Nationale d'Assurance Maladie dans le programme, via la mise à disposition de données complémentaires.

Quels enseignements tirez-vous de cette coopération avec les acteurs de la politique de la ville ?

Un des principaux enseignements de ce programme réside dans la diversité des acteurs réunis pour mettre en commun des connaissances, des savoir-faire, des données afin de répondre à une problématique choisie. Cette notion de collectif avec la participation de pouvoirs publics, de collectivités, d'associations, d'entreprises, etc. correspond aussi à l'ADN d'Harmonie Mutuelle.



Francky Trichet

**Vice-Président délégué au numérique,
à l'innovation et à l'international,
Nantes Métropole**



En tant qu'élu nantais, vous portez de nombreux projets numériques innovants, quelle place occupe le programme « Data & Quartiers » dans la stratégie métropolitaine ?

La stratégie numérique métropolitaine comprend de nombreux volets. Elle repose sur un historique pionnier. Nantes métropole a été l'un des premiers territoires à publier des données en *open data*, bien avant que la loi n'y oblige. Elle repose aussi sur une volonté politique de garder la maîtrise des outils déployés par nos services mais aussi par les entreprises qui travaillent avec la collectivité, et une volonté politique d'agir dans un cadre éthique. Nantes Métropole a adopté en 2019 la première charte éthique territoriale de la donnée en France.

Comme beaucoup de territoires, nous développons des usages innovants de la donnée dans des domaines techniques : pour faire des économies d'énergie, pour gérer la distribution de l'eau, pour améliorer les services de mobilité. Mais nous essayons aussi d'encourager des projets qui sortent de ces domaines traditionnels des territoires que l'on nomme « intelligents ». Nous avons piloté une expérience très innovante de recours à l'intelligence artificielle pour réduire le gaspillage alimentaire dans les cantines. Nous travaillons aussi avec de nombreux acteurs nantais sur une utilisation des données liées à notre environnement pour construire des politiques de santé.

Le programme « Data & Quartiers » entre dans cette famille de projets : mettre la donnée au service de politiques de l'emploi ou de politiques sociales. Bien sûr dans un cadre éthique car la gestion sociale implique l'utilisation de données qui sont souvent plus sensibles que des données techniques. Protéger ces données est une exigence non négociable.

Les méthodes utilisées par RésO Villes et ses partenaires, détaillées dans ce livre blanc, sont très novatrices pour la politique de la ville. Mais qu'en est-il de leur adoption par les métropoles pour la conduite de leurs propres politiques publiques ?

Pour ma part, je retiens un point essentiel de ce livre blanc : la complexité et l'importance de la contractualisation pour permettre un partage et des échanges de données au service de l'intérêt général. Qu'il s'agisse des données de Pôle emploi, des données de santé ou des données d'un groupe privé comme Randstad, le livre blanc montre l'importance du dialogue et la structuration nécessaire d'un cadre de confiance.

La charte métropolitaine de la donnée de Nantes Métropole parle de « données d'intérêt métropolitain », c'est-à-dire des données produites par des acteurs privés



qui n'ont pas d'obligation de les transmettre (ou parfois des données produites par des acteurs publics mais qui ne sont pas en *open data*). La démarche de RésO Villes est très intéressante et très pragmatique. Elle pourrait être très utile, par exemple, pour permettre à tous les acteurs en charge des politiques environnementales et de la santé, d'échanger des données essentielles pour mesurer l'impact de notre environnement sur la santé de tous. Et pourquoi pas quartier par quartier. La méthode utilisée pour les QPV est très inspirante !

Durant cette première étape, « Data & Quartiers » a concerné 3 problématiques prioritaires propres aux quartiers (emploi, santé, mobilité). Quels autres sujets devraient selon vous figurer en dans la suite du programme ?

Nous sommes au tout début de l'exploitation du potentiel des données. Sous réserve de solides garanties méthodologiques et éthiques, il est certainement possible de travailler sur de nombreux autres sujets. Les collectivités connaissent l'offre culturelle qui est offerte dans les quartiers, mais connaît-on les pratiques culturelles réelles des habitants ? Ou leurs loisirs en dehors des quartiers ? Ou leurs habitudes alimentaires ? Nous projetons des représentations qui manquent souvent d'objectivité, y compris de la part des professionnels et des acteurs de la politique de la ville, faute de données. C'est en tout cas ce que dit aussi ce livre blanc mais aussi les récents rapports officiels qui évoquent « Data & Quartiers ».

Nos partenaires canadiens de Montréal ont publié une étude très intéressante sur les îlots de chaleur dans la ville. Beaucoup de métropoles françaises disposent de ce type de données, et sont en train de construire des « jumeaux numériques » pour les présenter. Mais à Montréal, ils ont croisé ces données avec celles de la pauvreté. Le résultat est imparable : ce sont les personnes au plus faible revenu et les personnes d'origine étrangère qui vivent dans les pires îlots de chaleur... Voilà un beau sujet pour la suite de « Data & Quartiers ».



Émilie Sarrazin

Directrice, RésO Villes



RésO Villes vient de vivre 3 ans d'un programme qui met la data au centre de la politique de la ville, pouvez-vous nous dire quel était votre niveau de connaissance sur le sujet avant le lancement de « Data & Quartiers » ?

Il était proche du néant. A titre personnel, j'avais suivi les démarches d'*open data* dans les territoires, qui me semblaient répondre à des enjeux démocratiques et être particulièrement nécessaires mais sans réellement approfondir. À titre professionnel, je n'avais pas été confrontée à ces questions, ni dans le public, ni dans le privé.

J'ai donc moi-même eu un besoin d'acculturation. J'ai eu besoin de développer une forme d'affinité à la donnée, d'en comprendre la force, le rôle et les enjeux, d'en saisir la valeur et la richesse pour une association comme la nôtre.

Pouvez-vous nous parler de la façon dont vos équipes et vos interlocuteurs se sont saisis du sujet ?

Les équipes de RésO Villes ont accompagné rapidement la dynamique du programme sans crainte ni appréhension. Il y avait une forme d'évidence. Nous avons envie de comprendre ensemble tout ce que les données pouvaient apporter, pour mieux le partager ensuite. C'est une sorte « d'esprit » de la donnée qu'il faut mettre en place dans l'organisation, sans peur et en toute sincérité pour accompagner ensuite les acteurs.

Pour les interlocuteurs de l'association, les niveaux d'accueil ont été variés : de l'enthousiasme à l'indifférence, de la réticence à la bienveillance. Il n'y a pas de profil « type », juste des rythmes de compréhension différents et le fait d'avoir choisi des sujets variés dans un rythme qui laissait du temps pour de la pédagogie a permis de lever les doutes des plus éloignés.

Comment envisagez-vous le futur de la politique de la ville à l'aune des nouvelles méthodes autour de la data décrites dans ce livre blanc ?

Nous ne pourrions pas faire sans une démarche data dans le futur de la politique de la ville.

Parce que ce que nous a enseigné ce programme c'est que les data peuvent être tout à la fois des outils d'aide à la décision mais elles peuvent aussi accompagner une forme d'objectivation des choix, des analyses et des situations pour mieux proposer, modifier ou évaluer. La politique de la ville doit se saisir de ces nouvelles méthodes pour continuer à faire ce qu'elle fait depuis son origine : lutter contre les inégalités, résorber les écarts, être attentive aux fragilités et sensible aux expérimentations.



Patrick Vanoli

Directeur du pôle intelligence marché,
groupe Randstad France



Le Groupe Randstad France a été présent aux côtés de RésO Villes très en amont dans l'élaboration du programme « Data & Quartiers ». Quelles sont les raisons de cet engagement ?

Chez Randstad France, nous pensons qu'une meilleure connaissance et compréhension du marché de l'emploi local est à même de faciliter l'accès à l'emploi. L'approche innovante de ce programme, qui marie compétences des secteurs public et privé, doit permettre d'apporter cet éclairage, trop souvent absent des quartiers. Permettre à leurs habitants d'adapter leurs recherches d'emploi pour maximiser leurs chances de réussite, voilà la meilleure raison de s'engager dans le programme « Data & Quartiers ».

Concrètement, quelle a été votre contribution au programme ?

Notre groupe a développé un outil, d'abord pour ses propres usages, qui s'appelle « Randstad *Smart data* ». Il apporte une vision claire et précise de la situation du marché de l'emploi, en croisant différents critères : secteur d'activité, métier, localisation, période d'analyse...

Avec *Smart data*, nous avons pu identifier le volume d'offres d'emploi publiées correspondant aux métiers recherchés par les habitants des quartiers prioritaires de la politique de la ville (QPV), dans les régions Pays-de-la-Loire et Bretagne. Une modélisation réalisée par nos équipes de *data science* a permis d'apporter un niveau d'information très fin quant au nombre d'offres d'emploi publiées par métier, aux typologies de contrats proposées et à l'éloignement en kilomètres des centres de chaque quartier.

Quels enseignements tirez-vous de cette coopération avec les acteurs de la politique de la ville ?

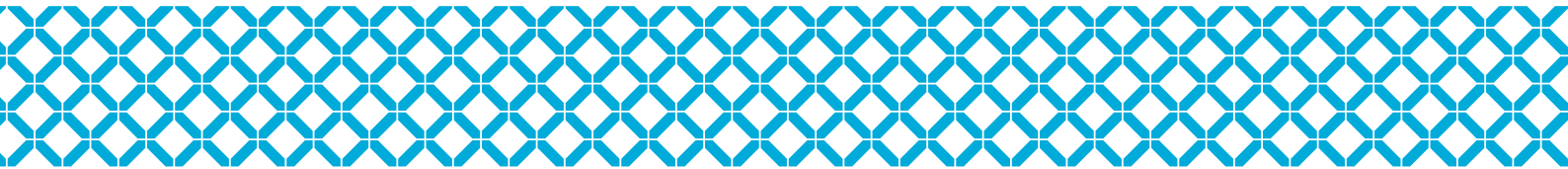
Le programme « Data & Quartiers » démontre deux choses essentielles. Tout d'abord, qu'une fine connaissance du marché de l'emploi local pourrait permettre aux habitants d'adapter leurs démarches de manière très opérationnelle. Ensuite, que la collaboration public-privé est nécessaire et essentielle pour bien appréhender les problématiques de l'emploi sur les territoires.

Par ailleurs, ce programme expérimental nous a aussi été utile en interne : il a nourri nos réflexions pour continuer à faire évoluer notre solution Randstad *Smart data*.



Louis-Vincent Potura

Dirigeant, Valoway



Votre entreprise a rejoint le programme « Data & Quartiers » en tant que partenaire technique, pouvez-vous nous préciser en quoi ce rôle a consisté ?

Valoway et son équipe d'experts ont accompagné le programme à la fois dans l'approche méthodologique et dans la mise en œuvre concrète des cas d'études de *data science*. Nous avons ainsi produit des analyses, des visualisations et des cartographies, par exemple à partir des corrélations entre les données démographiques des Quartiers et les données *open data* sur les thématiques de la Santé et de l'Emploi.

En dehors de ce programme, avez-vous d'autres missions de ce type auprès d'acteurs publics ?

Nous n'avons pas encore eu l'opportunité de mener directement d'autres missions de ce type auprès d'acteurs publics. La démarche de « Data & Quartiers » reste rare et pionnière. Mais en revanche les équipes de Valoway travaillent très régulièrement avec des données en *open data* mises à dispositions par des acteurs publics (Assemblée Nationale, Insee, Cerema...)

Est-ce que ce programme, dédié à l'intérêt général, a apporté des choses nouvelles à vos équipes ?

C'est une vraie satisfaction personnelle pour chacun des collaborateurs qui a pu participer à ces travaux et cela a été une très belle collaboration pour les équipes Valoway. Nous avons pu accompagner les collaborateurs de l'association RésO Villes et leurs partenaires dans une découverte et dans leur compréhension des sciences de la donnée.

Mais nous sommes allés au-delà de cette première étape. Nous avons mis en œuvre des outils d'analyse et de pilotage qui pourront avoir un impact réel sur les politiques des quartiers et sur des enjeux très concrets et mobilisateurs !



Cécile Nonin

Directrice, Villes et Territoires Occitanie



Vous avez assisté à une présentation de retour d'expérience de RésO Villes sur le programme « Data & Quartiers ». Quelles réflexions cela a-t-il déclenché ?

Nous avons organisé une présentation du programme « Data & Quartiers » en Occitanie pour répondre aux interrogations d'acteurs du territoire qui avaient la curiosité de comprendre ce qu'est le *big data*. Ils voulaient comprendre la manière dont ces nouveaux outils pouvaient être mis au service des politiques sociales, notamment en Occitanie où l'État puis le Conseil Régional se sont saisis du sujet et proposent un accompagnement des collectivités volontaires.

Les participants ont fait part de leur crainte que suscite *a priori* l'exploitation des données, notamment les enjeux éthiques (atteintes aux libertés publiques en réponse à des choix sécuritaires par exemple). Pour autant, les participants estimaient nécessaire d'anticiper la réflexion pour que les données soient mises au service de l'intérêt général dans le respect des personnes, notamment des habitants des QPV.

Villes et Territoires a proposé à son CA, composé d'élus, de s'engager dans une expérimentation similaire afin d'avoir, en partenariat avec RésO Villes, un second territoire d'expérimentation. Cette démarche aurait pu se faire en lien avec le Conseil régional qui est très engagé dans la gestion des données, avec l'appui de l'association Occitanie Data. Mais en 2019, le timing n'était pas le bon, la plupart des élus étaient en fin de mandat.

Comment envisagez-vous le futur de la politique de la ville à l'aune des nouvelles méthodes autour de la data présentées par RésO Ville ?

Le *big data* va s'inscrire logiquement dans les démarches d'observation des territoires. Il s'agit d'anticiper cette évolution et d'outiller la politique de la ville pour faire des analyses prospectives et permettre un pilotage réactif des politiques publiques du territoire. La garantie d'un cadre de confiance éthique et souverain apporté par Occitanie Data permettrait de rassurer élus et techniciens, et d'apporter une connaissance plus fine du vécu des habitants. Les quartiers prioritaires sont, au même titre que les territoires ruraux, des territoires intelligents !



Renaud Epstein

**Sociologue spécialiste des politiques urbaines,
Maître de conférence en science politique**



Le programme « Data & Quartiers » a montré de façon expérimentale que de nouvelles méthodes et de nouvelles données peuvent permettre un regard différent sur certaines problématiques au cœur des quartiers. En quoi est-ce utile selon vous ? Les outils d'observation classiques ne sont-ils pas suffisants ?

Les outils classiques d'observation des quartiers reposent essentiellement sur la statistique publique. Or, celle-ci est soumise à des contraintes, notamment en termes de rythme de publication (souvent à l'année, parfois même au gré des recensements) et d'échelle géographique (toutes les données publiées ne sont pas disponibles à la maille QPV par exemple). De plus, la catégorisation des données, pensée à l'échelle nationale, ne répond pas forcément aux besoins locaux. Il est donc parfois difficile de réaliser des analyses fines permettant d'éclairer les problématiques spécifiques des quartiers.

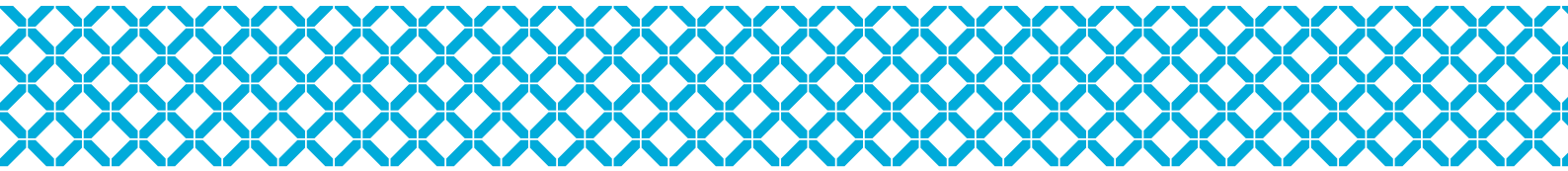
Malgré les progrès réalisés par la statistique publique ces vingt dernières années et malgré la mise à disposition d'un nombre croissant de données produites par de multiples institutions (CAF, Pôle emploi, Impôts, bailleurs sociaux, transporteurs, etc.), il y a des besoins de connaissances auquel la sphère publique n'est pas en mesure de répondre.

La volonté de RésO Villes d'aller à la recherche de sources de données complémentaires, notamment de données en possession d'acteurs privés comme le Groupe Randstad, est à saluer. Ce type de démarche est vraiment nécessaire afin d'améliorer la connaissance des quartiers.

Les utilisations innovantes des données concernent des sujets variés (santé, emploi, mobilité, sécurité, logement, ...). Y a-t-il selon vous des sujets pour lesquels il est prioritaire de renforcer l'usage des données ? Pour quel impact concernant les habitants des QPV à court ou moyen terme ?

Plutôt que des sujets prioritaires, je pense qu'il faut raisonner en termes de données prioritaires. Il faut combler les insuffisances de la statistique publique et cela passe par l'obtention de données permettant des suivis longitudinaux des individus ou des ménages. À l'ère du *big data*, cela est tout à fait possible, à condition bien entendu de prendre toutes les précautions d'usage en termes de protection de la vie privée et en contrôlant les biais.

L'utilisation des données massives de téléphonie permet par exemple de réaliser des analyses sur la mixité sociale dans la ville, au-delà de la mixité purement



résidentielle. On peut mieux comprendre les lieux et les temporalités de la mixité et avoir une vision plus fine des pratiques différenciées des espaces urbains en fonction du sexe, de l'âge, des revenus, du lieu de résidence, etc.

Autre exemple : l'analyse des cartes de fidélité des grandes surfaces permettrait de combler une vision à ce jour assez pauvre des habitudes de consommation des habitants des quartiers. Cela pourrait avoir des impacts notables en termes de politiques de santé via la mise en place d'actions de prévention personnalisées.

Citoyens, centres de ressources, collectivités, ministères, entreprises privées : qui doit-on inciter à mettre plus de données à disposition pour définir et mener la politique de la ville de demain ?

Pour commencer, il faudrait que les acteurs publics remplissent leurs obligations de mise à disposition de données, ce qui n'est pas toujours fait, ou pas toujours fait de façon satisfaisante.

Et il y a un type de données publiques dont la disponibilité fait cruellement défaut : les données budgétaires territorialisées. Si on veut évaluer les politiques de la ville, il faut en premier lieu se pencher sur les moyens financiers alloués aux quartiers au titre de chaque politique et de chaque programme. Il faudrait être en mesure de présenter les budgets publics de façon détaillée et géolocalisée, comme la cour des comptes le préconise depuis une quinzaine d'années. Il y a peut-être un défi d'ordre technique à surmonter mais les gains en termes d'analyse et d'évaluation des politiques de la ville seraient énormes.

Annexes



Annexe 1

Exemple de calcul des « unités de consommation » et du « Plafond de ressources » pour l'IRIS GRASLIN COMMERCE (Nantes)

Données disponibles : nombre de personnes et de ménages, nombre de familles, nombre de familles monoparentales, nombre de familles biparentales, nombre de familles avec 1, 2, 3, 4 enfants et plus, nombre de personnes de plus de 15 ans.



Graphique représentant les compositions familiales à l'intérieur de l'IRIS Graslin-Commerce (source INSEE)

Calcul du nombre d'unités de consommation sur chaque code d'IRIS en partant des règles définies : 1 UC pour le premier adulte du ménage, 0,5 UC pour les autres personnes de 14 ans ou plus, 0,3 UC pour les enfants de moins de 14 ans.

Ce qui donne pour Graslin-Commerce :

Pour les adultes

- 105 familles monoparentales x 1UC = 105 UC
- 825 familles biparentales x (1+0,5)UC = 1237,5 UC
- 1818 personnes de plus de 15 ans vivant seules x 1 UC = 1818 UC

Pour les enfants

- 458 enfants de moins de 15 ans x 0,3 UC = 137,4 UC
- 734 enfants de plus de 15 ans ne vivant pas seul x 0,5 UC = 367 UC

Soit un total de 3664,9 UC sur l'IRIS Graslin-Commerce.

Attention : les règles de comptage des UC prennent en compte les enfants de plus ou moins de 14 ans et les données INSEE fournissent l'information du nombre d'enfants de plus ou moins de 15 ans... Nous avons donc dû faire une approximation statistique.



Conversion des compositions familiales fournies par l'INSEE en Unités de Consommation à l'intérieur de l'IRIS Graslin-Commerce

Nous sommes donc en mesure d'établir pour chaque code IRIS une conversion ménage-UC. Par exemple pour Graslin-Commerce, il y a 2973 ménages et 3664,9 UC ce qui équivaut donc à 1,23 UC par ménage.

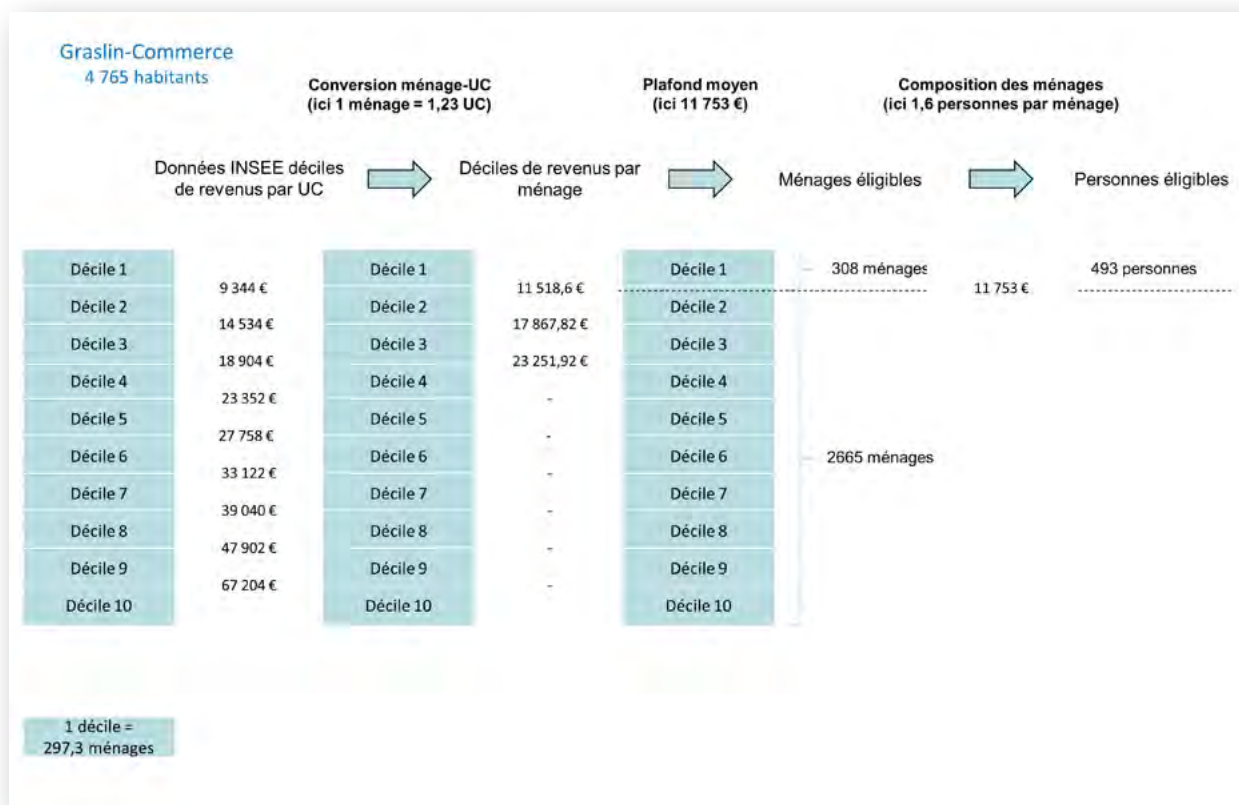
Il faut ensuite calculer le plafond moyen de ressources nécessaires par ménage pour avoir accès à la CSS. Nous possédons via l'INSEE des informations concernant le nombre de ménages et le nombre de personnes à l'intérieur de chaque code IRIS. Nous en déduisons une composition moyenne de chaque ménage. Par exemple pour Graslin-Commerce, nous avons 2973 ménages et 4765 habitants, donc une composition moyenne de 1,6 personne par ménage.

En projetant cette composition moyenne des ménages sur la grille des conditions de ressources, nous calculons pour chaque code IRIS le plafond moyen de ressources en dessous duquel l'accès à la CSS est possible. Dans le cas de notre exemple nous obtenons un plafond de 11753 euros en réalisant le calcul suivant :

Nombre de personnes composant le foyer	Plafond annuel Complémentaire santé solidaire sans participation financière
1 personne	9 041 €
2 personnes	13 561 €
3 personnes	16 273 €
4 personnes	18 985 €
Au-delà de 4 personnes	+ 3 616,24 € par personne supplémentaire

I.R. 13033 (Graslin-Commerce) : 13561*9041 + 4520*4520*0,6 + 2712*9041+2712 = 11 753
 Résultat : 11 753 €

Illustration de la méthodologie de calcul du plafond moyen d'accès à la CSS sur le code IRIS Graslin-Commerce



Méthodologie de calcul du nombre de bénéficiaires potentiels à la CSS par code IRIS (exemple sur l'IRIS Graslin-Commerce)

Il ne reste plus qu'à appliquer ce plafond à la répartition réelle des revenus à l'intérieur de chaque code IRIS pour estimer le nombre de bénéficiaires potentiels à la CSS. Pour ce faire, nous utilisons le jeu de données de l'INSEE sur la répartition des revenus par déciles et par unités de consommation et nous y appliquons la conversion ménage-UC et le plafond moyen calculés lors des étapes antérieures.

Dans l'exemple du quartier Graslin-Commerce, le nombre de bénéficiaires potentiels de la CSS estimé est de 493 personnes.

ANNEXE 2

QPV et IRIS : à la recherche de données introuvables

Posons ici le problème de façon brute

Après de nombreux débats ayant conduit notamment à une simplification des critères et à un zonage plus fin, le périmètre actuel des QPV a été entièrement refondu en 2014¹. Le découpage des zones est depuis lors connu et disponible en *open data*².

Pour les besoins de différents travaux ministériels, des données ont été produites par différents acteurs sur le périmètre de ces quartiers. Il s'agissait le plus souvent d'extraire les données de leurs bases à l'échelle des quartiers en fonction d'adresses connues. L'INSEE bien sûr³, Pôle emploi, la CNAM... travaillent régulièrement à l'échelle des quartiers prioritaires.

Mais pour conduire nos projets, nous avons besoin de croiser d'autres données et des données plus fines à l'échelle des quartiers. Et ces données existent mais elles sont publiées à l'échelle des IRIS.

Toutes ces données à portée de main mais non exploitables à l'échelle des quartiers... pas de quoi arrêter un *data scientist* !

Nous avons cherché durant de nombreux mois une « table de conversion » permettant de convertir de façon aussi juste que possible les données des IRIS en données pour les QPV. Nous étions convaincus qu'un tel outil devait exister, et nombre de nos interlocuteurs ministériels ou encore dans le monde académique aussi. En effet si cet outil de conversion n'existe pas... cela signifie que les acteurs de la politique de la ville sont privés de très nombreuses données utiles.

Les découpages

Le programme « Data & Quartiers » utilise prioritairement de l'*open data* et la maille géographique de mise à disposition la plus fine est généralement celle des IRIS (acronyme pour Îlots Regroupés pour l'Information Statistique). En usage depuis 1999, les IRIS ont été définis conjointement par l'Institut National de la Statistique et des Études Économiques (INSEE)

et les mairies. Ils épousent en général les contours des communes lorsque celle-ci ne dépassent pas 5000 habitants ; au-delà, les communes sont divisées en plusieurs IRIS en fonction de critères géographiques et démographiques comme le nombre d'habitants, l'homogénéité du type d'habitat, le type d'activité principale (résidence, entreprises, zones spécifiques comme les ports ou les forêts, etc.).

Les QPV quant à eux ont été définis par le Commissariat général à l'égalité des territoires (CGET, ancêtre de l'ANCT) en se basant sur un critère de concentration de personnes à bas revenus. Ce découpage a été réalisé à partir de données d'études effectuées sur un carroyage de 200 mètres (à l'exception de l'Outre-mer sauf la Martinique et la Réunion).

Les IRIS et les QPV sont donc deux réalités géographiques distinctes qui se superposent, comme nous pouvons l'apprécier sur les exemples ci-dessous :



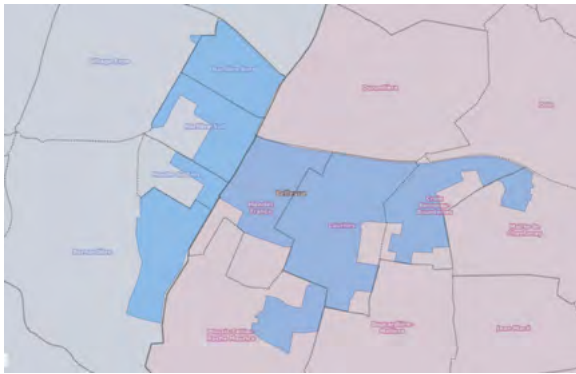
Délimitations géographiques du QPV «Maurepas» à Rennes, source géoportail IGN

Exemple 1 : Le QPV « Maurepas », à Rennes, s'étend sur quatre IRIS différents. Il englobe la totalité de l'IRIS « Brno », ainsi qu'une partie des IRIS « Le Gast Ouest », « Le Gast Est » et « Emmanuel Mounier ».

¹ La refonte de la géographie prioritaire de la politique de la ville figure dans la loi Lamy du 21 février 2014 (article 5) et les contours des quartiers ont été fixés par décret : décrets n° 2014-1750 et 1751 du 30 décembre 2014.

² <https://www.data.gouv.fr/fr/datasets/quartiers-prioritaires-de-la-politique-de-la-ville-qpv/>

³ De nombreuses données par QPV sont disponibles ici : <https://www.insee.fr/fr/statistiques/2500477>



Délimitations géographiques du QPV «Bellevue» entre Nantes et Saint-Herblain, source géoportail IGN

Exemple 2 : Le QPV « Bellevue », s'étend sur dix IRIS différents, situés dans deux communes : « Harlière Nord », « Harlière Sud », « Moulin du bois », « Village Expo » et « Bernardière » à Saint-Herblain et « Mendes France », « Lauriers », « Croix Bonneau-Bourderies », « Mairie de Chantenay » et « Plessis Cellier-Roche Maurice » à Nantes.

Notre quête

Afin d'illustrer l'intérêt que pourrait revêtir la mise à disposition d'une table de conversion IRIS ↔ QPV, prenons l'exemple de l'analyse du non-recours à la Complémentaire Santé Solidaire (C2S) réalisée dans le cadre des travaux du groupe Santé.

Nous nous sommes attelés à créer un indicateur de « taux de non-recours » à la C2S, qui représente la part des inscrits à la C2S sur l'ensemble des bénéficiaires potentiels⁴. Voici une extraction des résultats pour les IRIS composant le QPV « Maurepas » à Rennes.

IRIS	Inscrits à la CSS ¹⁷	Eligibles à la CSS ¹⁸	Taux de non-recours
Le Gast Ouest	309 personnes	401 personnes	23%
Le Gast Est	319 personnes	609 personnes	48%
Emmanuel Mounier	520 personnes	1127 personnes	54%
Brno	716 personnes	1414 personnes	49%

Taux de non-recours à la C2S pour les IRIS composant le QPV Maurepas à Rennes

On voit que parmi les IRIS composant un QPV, notre indicateur peut varier dans des proportions importantes (le non-recours est 2,3 fois plus élevé dans l'IRIS Emmanuel Mounier que dans Le Gast Ouest). Il est impossible sans la mise à disposition d'une table de répartition de calculer notre indicateur pour le QPV Maurepas.

La première méthode néanmoins intéressante consiste à faire des cartes et à superposer les périmètres. C'est le choix qui a été fait pour « Data & Quartiers ». Voici par exemple une représentation du tableau des taux de non-recours à la C2S pour les IRIS composant le QPV Maurepas à Rennes au format cartographique :



Cartographie du taux de non-recours à la C2S dans les IRIS à proximité du QPV Maurepas à Rennes



Cartographie du taux de non-recours à la C2S à Rennes proximité du QPV Maurepas à Rennes

Il est possible grâce à ce type de visualisations d'apprécier certains faits et de tirer de conclusions à l'échelle des QPV malgré l'absence d'indicateur propre à chaque QPV. On se rend compte dans ce cas précis que le taux de non-recours est moins élevé dans la périphérie de Rennes, là où sont situés les QPV. Ce type de visualisation comporte néanmoins des limites. Il est par exemple impossible de comparer les QPV entre eux ou de comparer différentes métropoles les unes avec les autres.

D'où la **nécessité de convertir les données disponibles grâce à une table de correspondance** qui nous permettrait de générer des indicateurs à l'échelle QPV à partir de données fournies à l'échelle IRIS.

⁴ Le détail de la méthodologie figure dans la partie 2 du livre blanc.

Des démarches infructueuses

Nous étions convaincus au départ du programme qu'une telle table de correspondance devait exister. Encouragés par différents partenaires nous avons donc entrepris de nombreuses démarches pour la trouver.

1. Explorer les portails *open data*. Ce fut bien sûr le premier réflexe des différents consultants, *data scientists* et autres acteurs des données intervenant sur le programme. Si une table de répartition ou des instructions logiques permettant de redistribuer les habitants entre IRIS et QPV existaient, elles devaient forcément être mises à disposition en *open data* quelque part. Des jeux de données sur les QPV sont en effet disponibles dont une table d'appartenance géographique qui précise, pour chaque quartier, les territoires auquel il est rattaché (EPCI, unité urbaine, département, région) ou un autre jeu de données sur les périmètres géographiques des quartiers. Rien de cela ne permet cependant de résoudre notre problème.

2. Solliciter les acteurs en charge de la politique de la ville et de l'Observation des quartiers. Si la table convoitée n'était pas disponible en *open data*, il s'agissait certainement d'un oubli, ou il existait une bonne raison pour ne pas la diffuser... mais elle devait exister. Le Pôle Analyse et Diagnostic Territoriaux de l'Observatoire National de la Politique de la Ville nous a rapidement dirigés vers le site du Système d'Information Géographique de la Politique de la Ville⁵. Un formulaire en ligne permet de renseigner une adresse afin de savoir si elle est située ou non dans un QPV. Cet outil, très utile pour d'autres usages sans doute, ne correspondait pas à notre besoin.



Capture d'écran de l'outil de localisation des adresses fourni sur le site du système d'information géographique de la ville

⁵ <https://sig.ville.gouv.fr/>

⁶ Comme d'autres avant elles, la société Poligma a d'abord vendu ses analyses aux élus à des fins électorales avant de commercialiser des outils des connaissances des territoires pour le pilotage de l'action publique.

Nous avons sollicité par la suite d'autres services clés au sein de l'Agence Nationale de la Cohésion des Territoires (ANCT) et sollicité l'Administrateur ministériel des données, pour arriver ensemble à la conclusion que cette table de correspondance n'était pas disponible au ministère.

Contacté en parallèle, l'INSEE, nous a fourni les mêmes réponses négatives.

Construire notre propre outil

Puisque l'outil convoité ne semblait pas exister, et étant donné le fort intérêt à pouvoir générer des indicateurs nouveaux à l'échelle QPV, nous avons estimé qu'il valait la peine d'essayer de le construire par nous-même. En effet, les efforts consentis pour mettre en place une table de correspondance seraient largement compensés par l'intérêt des utilisations possibles par la suite pour RésO Villes. De manière générale un tel outil, automatisé, serait très utile pour l'ensemble des centres de ressources, des collectivités et des acteurs des quartiers partout en France. Tout jeu de données fourni à l'échelle des IRIS pourrait être, de façon automatique, converti pour analyse par QPV, facilitant ainsi les prises de décision.

Nous avons fait appel à l'entreprise Poligma, qui a rejoint le programme « Data & Quartiers » en tant que partenaire, pour essayer de construire cette table de correspondance. Poligma est une entreprise montpelliéraine spécialisée dans la collecte et l'analyse des données des territoires. Poligma croise un nombre très important de données de tous types en intégrant à l'échelle la plus fine au service des acteurs publics⁶.

Le travail s'est déroulé en plusieurs temps.

1^{ère} itération : analyse par recouvrement des aires des IRIS et d'un QPV

La première tentative de construction de la table de correspondance a consisté à analyser uniquement les aires géographiques et à calculer un « ratio de recouvrement » IRIS/QPV. Par exemple, pour le QPV Maurepas, le tableau suivant est possible.

QPV	IRIS	RATIO
Maurepas	Emmanuel Mounier	67,11%
Maurepas	Brno	99,43%
Maurepas	Le Gast Ouest	34,52%
Maurepas	Le Gast Est	57,86%
Maurepas	La Motte Brulon	0,03%
Maurepas	Morbihan Est	0,00%
Maurepas	Saint-Laurent	0,09%
Maurepas	Les Gayeulles	0,01%
Maurepas	Le Gallet-les Longs Champs Nord	0,04%

Recouvrement des aires géographiques du QPV Maurepas et des IRIS le constituant

Le tableau doit se lire ainsi : 67,11 % de l'IRIS Emmanuel Mounier est à l'intérieur du QPV Maurepas, 99,43 % de l'IRIS Brno est à l'intérieur du QPV Maurepas, 0,01 % de l'IRIS Les Gayeulles est à l'intérieur du QPV Maurepas, etc. De nouveaux IRIS apparaissent par rapport à notre première lecture visuelle (voir supra). Il s'agit d'IRIS ayant une délimitation géographique avec le QPV : des rues et des axes urbains utilisés comme frontières du QPV. Le ratio de recouvrement est proche de zéro et peut être considéré en pratique comme nul.

Afin de vérifier la pertinence de la méthodologie mise en place, nous avons recherché des statistiques fournies aussi bien par IRIS que par QPV. Il serait ainsi possible d'appliquer le système de répartition proposé aux données fournies par IRIS et de comparer le résultat obtenu aux données réelles par QPV. Nous avons choisi pour cela les données de population 2018. Concrètement, si l'on continue sur notre exemple de Maurepas, à Rennes, voici comment nous avons procédé : population estimée QPV Maurepas 2018 = 67,11 % de la population réelle IRIS Emmanuel Mounier + 99,43 % de la population réelle IRIS Brno + etc.

Nous obtenons une estimation de la population du QPV Maurepas en 2018 de 6035 habitants, à comparer à la population réelle fournie par l'INSEE à la même date de 6206 habitants, soit un écart de 171 habitants ou 2,75 %, ce qui n'est pas si mal. Malheureusement cette marge d'erreur très faible n'est pas la norme dans les autres QPV de Bretagne et de Pays de la Loire. Avec la même méthode, nous constatons des écarts importants qui biaiseraient significativement la suite de nos travaux...

QPV	Commune	Population estimée	Population réelle	Différence	Différence en %
Bellevue	Brest	3578	7120	-3542	-49,75%
Keredern	Brest	1897	2008	-111	-5,53%
Lambazellec Bourg	Brest	342	1210	-868	-71,76%
Maurepas	Rennes	6035	6206	-171	-2,75%
Le Blossne	Rennes	9705	11550	-1845	-15,98%
Cleunay	Rennes	301	1317	-1016	-77,18%
Malakoff	Nantes	747	3486	-2739	-78,57%
Bottière Pin Sec	Nantes	5264	5757	-493	-8,57%
Roseraie	Angers	6850	9589	-2739	-28,57%
Savary	Angers	532	1079	-547	-50,72%

Comparaison des populations réelles et estimées de quelques QPV avec la première table de correspondance

Un simple coup d'œil sur la carte aérienne de certains QPV permet d'en comprendre la raison. Par exemple à Brest, si nous comparons visuellement le QPV Lambazellec Bourg (en bleu ci-dessous) et l'IRIS Lambazellec dont il fait partie (pointillés noir), nous voyons que le QPV ne représente à vue d'œil que 20 % de l'IRIS, mais qu'il regroupe probablement plus de 70 % de sa population. En effet, le reste de l'IRIS est composé d'un cimetière, d'un hôpital, d'un parc, etc... La population n'est donc pas répartie à l'intérieur de l'IRIS. L'approche géographique est insuffisante.



Vue aérienne de l'IRIS Lambazellec et du QPV Lambazellec Bourg

2^{ème} itération : analyse via le recouvrement des aires construites (foncier cadastral) des IRIS et QPV

L'équipe de Poligma a proposé une deuxième approche intégrant les données du foncier cadastral permettant de croiser les IRIS et les QPV en intégrant la proportion de foncier bâti (données cadastrales).

Nous avons appliqué la même méthode de contrôle en utilisant des données connues (population 2018 par QPV et par IRIS). Les résultats sont à nouveau décevants...

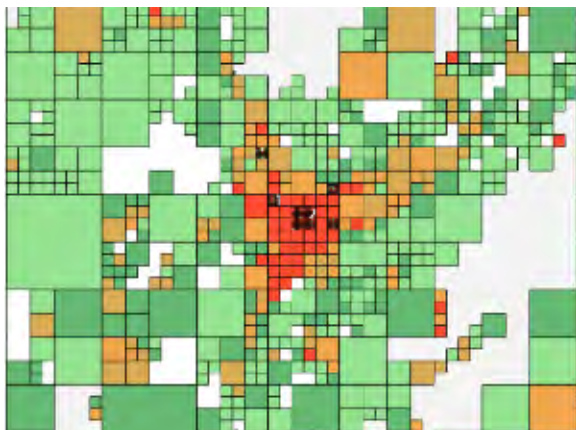
QPV	Commune	Population estimée	Population réelle	Différence	Différence en %
Bellevue	Brest	4706	7120	-2414	-33,91%
Keredern	Brest	1697	2008	-311	-15,49%
Lambazellec Bourg	Brest	378	1210	-832	-68,74%
Maurepas	Rennes	5548	6206	-658	-10,60%
Le Blossne	Rennes	10227	11550	-1323	-11,45%
Cleunay	Rennes	466	1317	-851	-64,63%
Malakoff	Nantes	1845	3486	-1641	-47,07%
Bottière Pin Sec	Nantes	5511	5757	-246	-4,28%
Roseraie	Angers	7307	9589	-2282	-23,80%
Savary	Angers	579	1079	-500	-46,37%

Comparaison des populations réelles et estimées de quelques QPV avec la deuxième table de correspondance

L'explication tient très probablement au fait que seule l'emprise foncière a été prise en compte. Et pas la hauteur des bâtiments. A l'évidence une maison individuelle n'a pas le même poids dans les calculs qu'un immeuble de plusieurs étages. Même si son emprise foncière est importante. Faut de données disponibles simplement concernant la hauteur des bâtiments, ou leur population, nous devons changer de méthode.

3^{ème} itération : tentative à partir des données carroyées et des QPV

La maille géographique la plus fine de mise à disposition des données en *open data* est généralement celle des IRIS. Mais certains jeux de données sont disponibles à une échelle encore plus précise : celle des données carroyées. Poligma a travaillé lors de cette troisième itération sur les données carroyées au niveau naturel, avec un pavage du territoire par carrés de 200 mètres de côté (ou plus en cas de non-respect du seuil de confidentialité de 11 foyers). La méthodologie employée est la même que lors de la première itération, à savoir une prise en compte des aires de recouvrement géographiques.



Exemple de pavage du territoire obtenu avec les carreaux du niveau naturel (source <https://www.insee.fr/fr/statistiques/4176281>)

Nous avons eu recours à nouveau aux données de population réelles afin d'estimer la fiabilité de la table de correspondance ainsi obtenue :

QPV	Commune	Population estimée	Population réelle	Différence	Différence en %
Bellevue	Brest	10802	7120	-3682	-51,71%
Keredern	Brest	7904	2008	-5896	-293,63%
Lambezellec Bourg	Brest	3324	1210	-2114	-174,71%
Maurepas	Rennes	18526	6206	-12320	-198,52%
Le Blossne	Rennes	24575	11550	-13025	-112,77%
Cleunay	Rennes	9470	1317	-8153	-619,06%
Malakoff	Nantes	7235	3486	-3749	-107,54%
Bottière Pin Sec	Nantes	12382	5757	-6625	-115,08%
Roseraie	Angers	12923	9589	-3334	-34,77%
Savary	Angers	8130	1079	-7051	-653,48%

Comparaison des populations réelles et estimées de quelques QPV avec la troisième table de correspondance

Les estimations avec cette troisième méthode s'avèrent plus éloignées encore de la réalité que lors des deux itérations précédentes... Faut de temps, le programme expérimental touchant à sa fin, nous avons du renoncer.

Quelques enseignements pour l'avenir

Au-delà de la déception après les efforts menés en vain pour essayer de construire une table de correspondance IRIS ↔ QPV, c'est d'abord un sentiment d'incrédulité qui domine au sein des partenaires et des acteurs de « Data & Quartiers ». Comment mesurer l'efficacité des actions mises en place et piloter la politique de la ville par quartier lorsque de nombreux indicateurs ne sont pas disponibles à l'échelle de ces quartiers ? Et si ces indicateurs sont produits à d'autres échelles, comment se fait-il que rien n'ait été mis en place depuis 2014 pour pouvoir les suivre dans les limites de la géographie prioritaire ?

La recherche d'une table de correspondance n'est qu'un pis-aller, mais elle aurait au moins permis de compléter nos travaux. Notamment pour achever le prototype sur le non-recours à la C2S. Nous avons réussi à définir les profils types des bénéficiaires dans certains quartiers mais nous sommes incapables de savoir si de nombreux habitants présentent des profils similaires et sont exclus du dispositif.

Le présent livre blanc n'a rien d'un rapport officiel et n'a pas été écrit pour formuler des recommandations, mais si le programme « Data & Quartiers » devait être prolongé, nul doute qu'il faudrait débiter par ce chantier : **faire en sorte que des données massivement disponibles à l'échelle des IRIS soient dorénavant disponibles à l'échelle des QPV.**

Remerciements



Ils et elles ont participé à différentes étapes du programme « Data & Quartiers », nous tenions à les remercier.

Agence Nationale de la cohésion des territoires

Anne Beauchesne
Frédéric Bourthoumieu
Christophe Girard
Léa Gislais
Aminata Keita
Francois-Antoine Mariani
Noémie Oswald
Pierre-Louis Rolle
Perrine Simian

ARS Pays de la Loire

Chantal Boudet
Bernard Geffroy

Brest Métropole

Maël Le Berre
Arnaud Willaime

Caisse Nationale d'Assurance Maladie

Jean-Marie Blanchoz
Agathe Finas
Corinne Lamarque
Claire Marcade
Christelle Pierre-Emile
Fanny Richard
Catherine Vibert

CARENE Saint-Nazaire Agglomération

Benoît Ferrandon
Sandrine Hachon
Catherine Marchand
Maëlle Russier

Les Centres de ressources de la politique de la ville

Mustapha Berra, Cité Ressources
Cécile Dublanche, Ville au Carré
Cécile Nonin, Villes et Territoires Occitanie

CIVITEO

Jordan Esnault
Jacques Priol
Romain Vallée

Groupe Randstad France

Mohamed Morouche
Patrick Vanoli
Anne-Sophie Zanini

Harmonie Mutuelle

Benoît Blondet
Olivier Vaillant

Le Mans Métropole

Virginie Saugier

Nantes Métropole

Mathieu Janicki
Jean-Michel Jaouen
Sandrine Logaridis
Claire Sacheaud
Francky Trichet

Rennes Métropole

Hugues Aubin
Marion Glatron
Cécile Tamoudi

RésO Villes

Mathias Cadot
Frédéric Frenard
Nathalie Gosselin
Isabelle Melscoët
Emilie Pigeon
Adelaide Richard
Emilie Sarrazin
Maximilien Steindorsson

Pôle emploi

Stéphane Héraud
Kathleen Marie-Joseph
David Vera

Poligma

Stéphane Boisson
Tom Somerville Roberts

Université de Rennes 2 (Master SIGAT)

Charlie Deshayes
Clémence François
Boris Mericksay
Carl Nosland
Julie Poitevin

Valoway

Julien Cantin
Alice Gallais
Baptiste Hervé
Pernilla Mercier
Adrien Pavie
Louis-Vincent Potura

RésO
Villes

CIVITEO



RésO Villes et le cabinet CIVITEO mettent à disposition l'entièreté du contenu de ce livre blanc sous licence CC BY-NC 2.0 (partage et adaptation autorisés, pas d'utilisation commerciale)

ISBN : 979-10-92670-16-5

Septembre 2022 - Réalisation : pesberg.com

Photo de couverture : © Frédéric Le Moullour - Brest Métropole