



ÉCOLE NATIONALE
DES SCIENCES
GÉOGRAPHIQUES

Ecole Nationale des
Sciences Géographiques



Autorité de régulation des
communications
électroniques, des postes
et de la distribution de la
presse

Mémoire - Rapport de stage
Cycle des Ingénieurs de l'ENSG 3^{ème} année

Analyse cartographique et statistique de déploiements mobiles



ÉCOLE NATIONALE
DES SCIENCES
GÉOGRAPHIQUES

Charles Laverdure

Septembre 2022

☒ Non confidentiel ☐ Confidentiel IGN ☐ Confidentiel Industrie ☐ Jusqu'au ...

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6-8 Avenue Blaise Pascal - Cité Descartes - 77420 Champs-sur-Marne
Téléphone 01 64 15 31 00 Télécopie 01 64 15 31 07

Jury

Président de jury :

Pierre-Yves HARDOUIN

Commanditaire :

Guillaume Decorzent

Encadrement de stage :

Guillaume Decorzent

Enseignant référent :

Malika Grim-Yefsah

Responsable pédagogique du cycle Ingénieur :

Jean-François Hangouët, IGN/ENSG/PEGI

Responsable pédagogique du cycle 3^{ème} année TSI :

Victor Coindet, IGN/ENSG

Gestion du stage :

Delphine Genes, Claire Driessens, IGN/ENSG/DFI

© ENSG

Stage de troisième année du 10/05/2022 au 14/11/2022

Diffusion web : ☒ Internet ☒ Intranet

Situation du document :

Rapport de stage de fin d'études présenté en fin de 3^{ème} année du cycle des Ingénieurs

Nombres de pages : 55 pages dont 17 d'annexes

Système hôte : L^AT_EX

Modifications :

EDITION	REVISION	DATE	PAGES MODIFIEES
1	0	09/2022	Création

Avant-propos

Je tiens tout d'abord à remercier mon tuteur et le responsable de mon accueil au sein de l'ARCEP lors de ce stage, Guillaume DECORZENT. M. DECORZENT m'a permis de rapidement m'intégrer à l'environnement de travail de l'ARCEP. Il a dès le début du stage su m'orienter dans les consignes en gardant un œil avisé sur mon travail. Je le remercie aussi pour la grande autonomie qu'il m'a laissée dans la manière d'aborder les études que j'ai réalisées.

Au-delà de M. DECORZENT, je veux ici remercier toute l'ARCEP et plus particulièrement la DMI pour son accueil chaleureux et l'encadrement général que j'ai pu recevoir. Mes collègues proches ont fait de ce stage une très bonne expérience professionnelle mais aussi humaine.

Concernant l'encadrement académique de ce stage, je tiens à remercier Mme Malika GRIM-YEFSAH, son apport concernant la méthodologie de travail en analyse de données en début de stage a été essentiel pour que tout se passe bien. Elle m'a aussi permis de trouver les bonnes terminologies afin de faire passer les informations nécessaires à la restitution de mes travaux.

Ce stage étant la conclusion de mon parcours de formation ingénieur à l'ENSG, je tiens bien sûr à remercier toute l'équipe pédagogique qui m'a accompagnée jusqu'à ce point lors de mes trois années de scolarité. Une mention toute particulière pour Victor COINDET, responsable du cycle TSI qui m'a permis de développer mes connaissances en science de l'information, ce vers quoi je veux maintenant orienter ma carrière.

Note au lecteur : Une version révisée des codes Python évoqués dans ce rapport sera accessible via un répertoire Gitlab à une date ultérieure à la publication de ce rapport. Le lien vers ce répertoire sera communiqué au président du jury avant la soutenance et sera inclus dans les prochaines versions de ce rapport.

Résumé

La Direction Mobile et Innovation (DMI) se concentre sur la **régulation** des technologies mobiles au sein de l'Autorité de Régulation des Communications Electroniques, des Postes et de la distribution de la Presse (**ARCEP**). Dans cette direction, l'Unité Couverture et Investissements Mobiles (**UCIM**) s'occupe de la régulation technique des obligations des opérateurs de **téléphonie mobile** en France (métropole et outre-mer). Les services des réseaux mobiles se décomposent en un service « **voix** » (appels, SMS...) et un service « **données** » (internet mobile : web, streaming, jeux...) avec une importance de plus en plus élevée pour le deuxième.

L'**UCIM** est chargée de produire de la donnée ou d'analyser les données des **opérateurs**. Cela permet de statuer sur le respect des engagements des opérateurs vis-à-vis des **législations** mises en place par l'**Etat français** et de proposer de nouveaux textes.

Mon stage s'inscrit dans une phase d'**automatisation, de standardisation et d'interprétation** des données **géoréférencées** liées aux **déploiements mobiles** et à leur utilisation en France. Ce stage se décompose en plusieurs projets courts rendus de manière continue. Ce rapport présentera quatre projets choisis pour leurs diversités de sujets et de technologies employées.

Dans un premier temps, la création d'outils de « **data visualisation** ». Ces outils permettent aux agents de l'ARCEP d'accéder rapidement aux informations dont ils ont besoin d'une manière visuelle et paramétrable. Ils sont utilisés afin de s'informer ou d'effectuer un travail de suivi. On retrouve un outil de suivi des enquêtes de **qualité de service** et un outil de suivi de l'avancée d'un programme d'état destiné à mieux couvrir en 4G la population.

Dans un second temps, mon stage s'est composé d'une phase plus exploratoire. En travaillant sur des projets transverses, j'ai pu recouper différents types de données afin de mettre en lumière des pistes de travail pouvant être réutilisées dans le futur. On y retrouve une étude prédictive de déploiements optimaux de sites mobiles dans le cadre d'un programme d'état et une étude sur le rapport entre qualité de service pour l'utilisateur et capacité de transfert de données disponibles.

Mots clés : régulation, ARCEP, UCIM, téléphonie mobile, voix, donnée, opérateurs, législations, Etat français, automatisation, standardisation, interprétation, données géoréférencées, déploiements mobiles, data visualisation, qualité de service.

Résumé

The mobile and innovation department is responsible for the **regulation** of mobile phones technologies as part of the French body for the regulation of electronic communications, press and postal services (**ARCEP**). Within those, one can find the unit in charge of coverage and mobile investments (**UCIM**) which regulates the technical obligations that **wireless carriers** operating inside the French territory (Metropolitan France and overseas territories) have to obey. This regulation work can be divided into two parts; **voice and data** technologies.

The **UCIM** unit oversees the production of its own data and exploitation of data communicated by the **carriers**. This data gathering makes ruling on carriers' commitments toward the different **legislations** possible. It also allows for the proposition of new texts that could be added in the law by the **French government**.

My internship includes itself in the **automatization, standardization and interpretation** processes of **georeferenced data** linked to **mobile deployments** and its uses in France. This internship is fractioned into multiple small projects with a continuous handoff of the work. For the sake of clarity, I have chosen to present four of those projects in this report. These projects have been chosen for the range of subjects they cover as well as the different technologies they employ.

Firstly, I've coded three software of **data visualization**. Those tools allow for ARCEP agents to quickly access and visualize the information they need and with the parameters they want. The tools both have a goal of project follow-up or simply informatory. There is a **quality of service** campaigns following tool and a follow-up tool for a national project with the goal of extending the population 4G coverage in France.

Secondly, my internship was centered around exploratory work. I have worked on transverse project for the whole department and was able to cross different types of data to extract valuable insight and potential considerations for the future. I did a study on predictive and most optimum site deployments as part of a national project and a study on the relation between the quality of service versus the available broadband.

Key words : regulation, ARCEP, UCIM, wireless, voice, data, carriers, legislations, French government, automatization, standardization, interpretation, georeferenced data, mobile deployments, data visualization, quality of service.

Table des matières

Glossaire et sigles utiles	3
Introduction	4
1 Contexte et objectifs du stage	5
1.1 L'ARCEP	5
1.2 Les besoins et les données	6
1.3 Les objectifs	8
1.4 Les outils	9
2 Outils de suivi et de contrôle	10
2.1 La qualité de service	10
2.2 Les sites mobiles et le programme de dispositif de couverture ciblée	16
3 Analyse des déploiements	19
3.1 Etude exploratoire sur l'implantation des sites	20
3.2 Capacité/Largeur de bande et qualité de service	26
4 Retour d'expérience et continuité	31
4.1 Difficultés rencontrées	31
4.2 Améliorations des outils	32
4.3 Fin du stage, suite et utilisations futures des projets	32
Conclusion	35
A Contexte	41
B Architecture des applications	43
C Application QOS	45
D Application DCC	48
E Etude de déploiements	51
F Data visualisation sur la largeur de bande	53

Glossaire et sigles utiles

ENSG École Nationale des Sciences Géographiques

ARCEP Autorité de Régulation des Communications Electroniques, des Postes et de la distribution de la Presse

DMI Direction Mobile et Innovation

UCIM Unité Couverture et Investissements Mobiles

URD Unité de Régulation par la Donnée

ANFR Agence Nationale des Fréquences

QOS Quality Of Service (Qualité de service)

DCC Dispositif de Couverture Ciblée

IDE Integrated Development Environment (Environnement de développement intégré)

KPI Key Performance Indicator

MOS Mean Opinion Score

Introduction

"Les réseaux comme bien commun."

Quelle phrase pourrait mieux ouvrir ce rapport que la devise de l'Autorité de Régulation des Communications Electroniques, des Postes et de la distribution de la Presse (ARCEP). Cette devise concentre en quelques mots ce que je vais détailler dans la suite de ce rapport et permet de comprendre le travail quotidien des plus de cent soixante-dix agents qui composent l'autorité.

Le rôle des organismes de régulation dans le système public français est parfois mal connu ou mal compris. Il est pourtant essentiel pour assurer, sur les marchés qu'il régule, une concurrence efficace et un service de qualité et au meilleur prix pour les citoyens en s'assurant qu'aucune entité ne puisse profiter d'eux. On appelle souvent l'ARCOM (ex CSA) le gendarme des médias, dans ce cas, l'ARCEP pourrait être considérée comme le gendarme des communications françaises.

Cette mission de régulation prend une place croissante dans un monde de plus en plus interconnecté et demandeur d'échanges rapides, sûrs et aux coûts justes. Au quotidien, l'ARCEP surveille et informe les pouvoirs publics concernant le respect et la mise en place des législations.

Le terme de 'communications' est large, ainsi, pour la suite de ce rapport et pour centrer le sujet, nous allons nous concentrer sur la partie qui nous intéresse dans le cadre de ce stage, les communications via la téléphonie mobile.

La mission de régulation, qui pourrait donc se composer de deux volets : surveillance et information, se fait aujourd'hui grâce à l'agrégation de données multiples, qui, une fois croisées, permettent de dresser une carte de l'état des lieux des communications mobiles en France. Cette carte permet de donner une direction pour les futures régulations afin de s'assurer que le citoyen français dispose des meilleurs services de téléphonie mobile possibles. Le terme 'meilleur' est lui aussi vague. Parle-t-on d'égalité d'accès ? De performance ? De coûts appropriés ? D'empreinte écologique faible ? Nous préciserons cela au cours de ce rapport.

Ce rapport se concentre sur la question de la mise en place d'outils et d'indicateurs à partir des données disponibles au sein de l'ARCEP. Ces derniers permettent de faire ressortir des axes de réflexion et des informations s'ajoutant à l'arsenal existant des régulateurs de l'unité mobile et innovation afin de faire leur travail avec le maximum de clairvoyance possible.

On retrouvera au début de ce rapport une partie de contexte et de présentation de l'ARCEP puis quatre parties relatant de quelques projets mis en place durant ce stage avant de prendre du recul sur les travaux accomplis et de conclure.

CONTEXTE ET OBJECTIFS DU STAGE

CHAPITRE 1

1.1 L'ARCEP

Le premier chapitre de ce rapport vise à introduire le contexte du stage en présentant les différents acteurs et la place de mon stage au sein de la structure qui m'a accueillie.

Commençons par présenter l'Autorité de Régulation des Communications Electroniques, des Postes et de la distribution de la Presse (ARCEP).

Comme son nom l'indique, il s'agit d'une autorité de régulation étatique autrement appelé Autorité Administrative Indépendante (AAI). Le rôle de l'Etat à travers ces AAI est «[...] de « préserver, garantir, stabiliser le libre jeu du marché » ou de « corriger, équilibrer, orienter » en assurant, dans le même mouvement, le respect du principe de libre concurrence et la protection d'intérêts publics non économiques, tels que la protection des libertés fondamentales, les droits des consommateurs, la sécurité des produits ou l'accès universel à des services d'intérêt général». [10]

Au sein de l'ARCEP, on décrit l'autorité comme « gardienne et architecte des réseaux d'échanges comme biens communs » [3]

Les missions de l'ARCEP sont multiples. Comme son nom l'indique, ses missions vont de la régulation de la distribution de la presse et celle des postes jusqu'à la mission qui prend le plus de place au sein de la structure, la régulation des communications électroniques.[1]

L'organisation est de fait très segmentée. On retrouve sept directions chacune chargée d'un secteur spécifique et regroupant près de 180 personnes. Toutes ces directions sont sous le contrôle d'une entité de direction spécifique : le Collège.

L'autorité étant indépendante, elle ne peut pas être soumise aux pouvoirs politiques en place au moment des décisions. La responsabilité des décisions diverses de l'ARCEP repose donc sur le Collège. Ce Collège est composé de sept membres expérimentés et venant d'horizons divers qui doivent statuer sur le travail effectué par les agents de l'ARCEP pour prendre les sanctions, mesures ou planifications adaptées. Ces membres sont nommés par le président de la République (3 membres), le président de l'Assemblée nationale (2 membres) et le président du Sénat (2 membres) pour une durée de 6 ans.

L'ARCEP dispose d'un crédit de fonctionnement alloué par la loi des finances et cette allocation de fonds de fonctionnement est contrôlée par la Cour des comptes.

Comme présenté sur la figure 1 disponible en taille complète en Annexe A, parmi les directions de l'ARCEP se trouve la Direction Mobile et Innovation (DMI). C'est au sein de cette dernière que j'ai effectué mon stage. Cette direction est chargée « de la gestion du spectre hertzien, de la régulation concurrentielle des marchés mobiles et des relations avec l'écosystème de l'innovation ».[2]. On y retrouve donc une association de missions de gestion technique, économique et administrative.

La DMI se divise en 5 unités sous la houlette de chefs d'unités chapotés par le directeur Mobile et Innovation, M. Franck Tarrier. Mon stage s'est effectué sur des projets transverses à toute la DMI mais j'étais cependant rattaché à l'Unité Couverture et Investissement Mobiles (UCIM) dont mon maître de stage, Guillaume Decorzent est le chef d'unité.

Pour la bonne compréhension du reste de ce rapport, on doit aussi mentionner des unités spécifiques dans leur rôles et leurs tâches avec lesquelles j'ai eu à travailler.

On retrouve l'unité des services informatiques (SI), en charge de la partie logicielle et logistique à l'ARCEP. J'ai aussi beaucoup travaillé avec l'unité de régulation par la donnée (URD). Cette dernière a un rôle de suivi de la relation avec les associations de consommateurs, et du lien avec les particuliers mais surtout un rôle d'automatisation des processus, de traitement et d'organisation de la donnée au sein de l'ARCEP. On compte dans leurs rangs deux géomaticiens qui assureront la suite de mon travail, comme il sera expliqué dans le dernier chapitre de ce rapport.

Après cette présentation exhaustive du fonctionnement et de la place de mon stage au sein de l'ARCEP, nous allons maintenant pouvoir étudier les besoins qui ont poussé l'UCIM à recruter un stagiaire.



FIGURE 1.1 – Logo de l'ARCEP

1.2 Les besoins et les données

L'UCIM entreprend depuis 3 ans un travail de restructuration et de pérennisation de l'organisation et des traitements de ses données. Au quotidien, les six membres de l'unité travaillent sur des sujets qui sont à renouveler trimestriellement ou annuellement. Ces travaux se ressemblent d'une campagne à l'autre.

Dans les travaux récurrents, on peut distinguer quatre types de tâches différentes. Tout d'abord, la mise en place des cartes de couverture (données fournies par les opérateurs). Puis, les enquêtes de qualités de service (données fournies par des prestataires). Ensuite vient la relation avec les collectivités et les citoyens via notamment la plateforme *j'alerte l'ARCEP* [5] qui permet aux utilisateurs (particuliers, collectivités. . .) de signaler les problèmes qu'ils rencontrent sur les marchés régulés par l'Arcep. Enfin, toutes ces données sont agrégées sur le site monreseau-mobile.arcep.fr qui centralise

les données traitées par l'unité et les rend accessibles au grand public. L'UCIM est aussi responsable de travaux de plus longue durée et moins récurrents. Il y a la mise en place des obligations des opérateurs lors des attributions ou renouvellements d'Attribution d'Utilisation des Fréquences (AUF). Parmi ces obligations passées on retrouve le programme de Dispositif de Couverture Ciblée (DCC), mis en place en 2018 à travers le « New deal [6] » et qui est suivi au quotidien par l'unité. On se rend compte que chaque membre de l'équipe est très occupé au quotidien dans des tâches qui se répètent avec une fréquence plus ou moins élevée.

L'unité de régulation par la donnée (URD) présentée précédemment est chargée de mettre en place des processus permettant d'automatiser les traitements et les analyses pour que le travail des agents se concentre sur l'évolution et l'amélioration des missions déjà en place et qu'ils puissent se débarrasser des tâches répétitives. Dans un autre volet, l'URD a une mission d'harmonisation de la donnée et de son stockage afin de pouvoir avoir une traçabilité dans le temps.

Comme indiqué, ce travail d'automatisation a toujours été un souci au sein de l'unité, mais il a pris une dimension nouvelle avec la création de l'unité de régulation par la donnée, avec une montée en puissance régulière tout particulièrement depuis 2019-2020 et l'arrivée de deux géomaticiens. De plus, l'URD n'est pas au fait de tous les enjeux du quotidien de l'UCIM puisqu'il s'agit d'une unité transverse à toute l'ARCEP.

Ensuite, chaque unité de la DMI produit et analyse des données mais les recroisements inter-unités ne sont pas optimisés. Il y a pourtant de la matière à faire des analyses croisées puisque toutes les unités traitent de sujets liés au mobile.

Etant donné la diversité des types de données gérées par la DMI, la grande majorité des données produites et utilisées par les unités se présente sous la forme de fichiers Excel. L'ARCEP pousse via l'URD vers la mise en place de bases de données SQL standardisées, mais leur utilisation n'est pas encore répandue. Pour l'instant, ces bases de données sont utilisées en back office pour les passages entre les données brutes et les KPIs (Key Performance Indicator) produits automatiquement.

Au-delà des données tabulaires, certaines unités dont l'UCIM utilisent des données aux formats géographiques. On peut donc retrouver des fichiers shapefile, de la suite ESRI ou encore de l'environnement Mapbox.

Ces données sont accessibles par les membres des unités pour leur travaux et l'ARCEP s'engage aussi dans une démarche grandissante d'open source avec la publication régulière sur data.gouv de données traitées et standardisées afin qu'elles puissent être accessibles pour le grand public.

Pour finir cette présentation des données, on peut préciser la sensibilité relative de certaines données. En tant que stagiaire, je n'ai pas eu accès à l'entièreté des données utilisables au sein de la DMI. Certaines données sont classées sensibles. Elles peuvent l'être pour plusieurs raisons parmi lesquelles on retrouve le secret défense pour la gestion des fréquences qui rentre en collision avec celles utilisées par les armées ou autres corps sensibles de l'Etat. On trouve aussi des données relatives à la gestion de la concurrence entre les opérateurs et sont donc sous le couvert du secret des affaires.

1.3 Les objectifs

Mon stage s'insère dans une double démarche. D'une part, identifier avec les agents de la DMI des thématiques prioritaires pouvant bénéficier de croisements de données et d'un suivi stable dans le temps liés aux grandes tendances de déploiement des réseaux mobiles, cela incluant également la définition des indicateurs associés. Dans le même temps, il s'agit de se servir de ces analyses pour mettre en place des processus réutilisables dans le futur par les agents pour sortir les mêmes indicateurs avec des données futures.

Concrètement, mon stage sera transverse à la DMI avec la mise en place d'outils et d'indicateurs intégrés aux logiciels déjà existants permettant de donner une vision d'ensemble sur des sujets communs aux unités concernées.

On voit que le but de stage est donc aussi de prendre du recul sur les travaux du quotidien et d'apporter un regard extérieur sur les données disponibles afin de proposer de nouvelle façon de les exploiter.

D'un point de vue de l'organisation, les six mois de stage sont subdivisés en différentes phases. On compte trois phases principales. Le diagramme de GANTT du stage est disponible en Annexe [A](#).

Durant les 5 premières semaines, j'ai effectué une phase de familiarisation avec les données et de recueil des besoins. Durant cette phase j'ai pu avoir des entretiens avec toutes les unités de la DMI afin de me familiariser avec leur fonctionnement et voir en quoi mes travaux pourraient leur être utile. Une fois cette phase accomplie, un plan de travail est mis en place avec mon maître de stage afin de se concentrer sur certains projets, mettre en place des objectifs précis et une timeline. Un ordre de priorité dans la réalisation des tâches a aussi été mis en place, les projets étant réalisés les uns à la suite des autres et certains étant plus annexes qu'essentiels. Enfin, un des points importants était de définir pour chaque projet comment on le note comme accompli. Il s'agit de mettre au clair dès le début le point d'arrêt de chaque projet, la plupart étant exploratoires et pouvant donc durer plus longtemps que prévu.

Ensuite vient la phase d'implémentation. Cette phase dure près de 3 mois, soit la majeure partie du stage. Durant cette phase, je prends le temps de mettre en place les projets et études qui seront détaillés dans les parties 2 et 3 de ce rapport.

Enfin une période est prévue en fin de stage pour pouvoir revoir les rendus, de faire de potentielles modifications et de s'assurer de la pérennité de leur utilisation.

Les objectifs sont conciliés dans un document Powerpoint que je présente à l'ensemble de la DMI. Ce Powerpoint est accompagné de fichiers Word qui détaillent le contenu des études retenues par mon maître de stage afin de centraliser l'ensemble des objectifs du stage.

Les objectifs particuliers de chaque projet ou étude seront présentés au cours des points de ce rapport.

1.4 Les outils

Une grande liberté m'a été accordée sur le choix des outils afin de réaliser mon stage. Je pouvais en effet choisir le langage de programmation que je souhaitais ainsi que les logiciels adaptés aux différents projets sur lesquels j'ai pu travailler. Cette liberté de choix était tout de même encadrée par l'unité Systèmes d'Information (SI) qui a un catalogue assez large de logiciels adaptés à chaque travail potentiel. L'autre facteur à prendre en compte dans le choix des outils était la pérennisation des projets mis en place. En effet, pour chaque nouvelle technologie employée, je devais m'assurer qu'un profil technique au sein de l'ARCEP puisse entretenir l'outil ou prendre le relais d'une étude. Il s'agissait donc de faire la part entre l'apport de nouvelles technologies et d'innovation et le fait qu'elle entre dans le cadre technologique utilisé à l'ARCEP.

Pour commencer la présentation des outils, les systèmes d'exploitation présent à l'ARCEP sont Windows pour les ordinateurs du personnel et les machines virtuelles et linux pour les serveurs de déploiement des applications.

Chaque agent de l'ARCEP a à sa disposition un PC avec une suite de logiciels déjà préinstallés liés au profil de l'agent recruté. Dans mon cas, des outils cartographiques comme QGIS étaient installés et aussi une instance d'Anaconda afin de me permettre d'accéder aux IDE et aux bases Python et R.

Au quotidien, j'ai fait beaucoup de gestion de données exploitées via le logiciel Excel ou encore PgAdmin, le logiciel de gestion de bases de données géographiques (BDDG) PostgreSQL.

Pour les traitements et les analyses de données, j'ai utilisé principalement le langage Python et l'IDE VsCode. S'il s'agissait d'une étude/analyse, j'utilisais le notebook Jupyter [8] afin de mieux présenter la démarche et les résultats. En cas de mise en place d'une application WEB j'ai utilisé le micro framework Flask pour Python afin de connecter mes codes Python (backend) avec une instance web et d'ensuite utiliser les technologies web pour la partie front-end.

Comme précisé en préambule de ce rapport, nous détaillons ici les projets que j'ai pu réaliser durant mon stage en commençant par les outils de suivi et de contrôle. Ils représentent deux applications WEB que j'ai pu implémenter et déployer sur le réseau interne de l'ARCEP. J'ai moi-même pris en charge l'intégralité de leur réalisation, du traitement des données initiales au design des plateformes, du backend au frontend, de l'intégration des données jusqu'au déploiement de l'application. Pour chacun des deux projets que j'ai choisi de mettre dans ce rapport, je préciserai les besoins que j'ai pu recueillir, les éléments qui ont fait que ce projet a été retenu par mon maître de stage et les objectifs liés à ce projet. Puis, pour chacun des projets, j'ai décidé de focaliser le rapport sur un point spécifique lié à ce projet qui en montre la complexité et permet de mettre en avant la résolution d'un problème. Les architectures des applications étant similaires (voir détails des architectures en annexe B), je pense que ce focus sur des points spécifiques de chaque application est plus adapté. Enfin, chaque partie sera conclue par une présentation du produit déployé et une discussion des résultats.

La partie suivante présente dans un premier temps, un outil de suivi des données de qualité de service, puis un outil de visualisation du déploiement d'un programme mobile.

2.1 La qualité de service

Pour commencer, la première application que j'ai pu mettre en place et qui m'a permis de me familiariser avec les outils nécessaires aux déploiements de telles applications était une application de suivi en temps réel des campagnes de qualité de service (QOS).

2.1.1 Présentation des besoins et objectifs

Chaque année, lors de campagnes de QOS financées par les opérateurs, l'ARCEP organise une collecte de données mobiles sur l'ensemble du territoire métropolitain et ultramarin. Des protocoles de collecte différents pour la métropole et les outre-mer sont mis en place par deux agents dédiés au sein de l'unité couverture et investissements mobiles. Ces protocoles incluent la mise en place de KPI (Key Performance Indicator) adaptés aux enjeux des territoires et qui permettent d'illustrer la réalité du terrain pour les utilisateurs. Les tests sont effectués par un prestataire externe à l'ARCEP et les données des campagnes sont envoyées au fur et à mesure (sur une base hebdomadaire) aux agents de l'ARCEP qui vérifient leur conformité. Ces données se présentent sous la forme de tableaux Excel avec une ligne par mesure réalisée.

Concrètement, ces tests consistent en une caisse remplie de différents modèles répandus de téléphones portables, avec une carte SIM correspondant aux opérateurs disponibles dans le territoire de l'enquête, comme le montre la figure 2.1 ci-dessous. Ces mobiles sont placés dans des conditions imitant le quotidien des utilisateurs. On y retrouve des relevés OUTDOOR (dans la rue), INDOOR (dans un bâtiment), INCAR (dans un véhicule en mouvement). Les données INDOOR sont obtenues

grâce à des filtres qui atténuent le signal reçu par les appareils mobiles afin de simuler plusieurs positions à l'intérieur d'un bâtiment. Le tracé d'une campagne est déterminé par les agents de l'ARCEP en début de campagne, tout le territoire n'est pas couvert chaque année par les enquêtes. Un autre aspect quant à lui réalisé dans son entièreté chaque année est l'enquête de qualité de service sur les axes de transport principaux. On y retrouve les axes TGV, les autoroutes, certaines nationales et lignes de TER.



FIGURE 2.1 – Image de promotion d'un prestataire en enquêtes de qualité de services mobiles, [https://www.connect-testlab.com/the-netherlands-2020-methodology]

En fin de campagne, une session d'échange avec les opérateurs est organisée. Ces derniers ont la possibilité de pointer les résultats de mesures qui pourraient ne pas être représentatifs des performances des réseaux des opérateurs, en apportant des éléments objectifs pour le justifier (par exemple, retirer du « panel » de pages web testées une panne qui présenterait un dysfonctionnement lors de la période de mesures, retirer les mesures liées à un téléphone mobile défectueux, etc.). Une fois ces échanges passés, toutes les données sont agrégées et on en retire des indicateurs à l'échelle du territoire. Ces données sont rendues disponibles pour le grand public via la plateforme monreseau-mobile.arcep.fr.

Les processus d'intégration des données de QOS dans la plateforme monreseau-mobile.arcep.fr sont déjà automatisés. En revanche, durant les campagnes qui durent un peu plus de trois mois pour la France métropolitaine, les agents qui reçoivent hebdomadairement les informations des prestataires n'ont pas de moyens simples et rapides de se rendre compte de l'avancée de la campagne en cours. J'ai identifié ce besoin lors d'un entretien avec les membres de mon équipe et j'ai monté une proposition d'application web permettant d'uploader sur une plateforme les données partielles reçues au fur et à mesure des campagnes. Il s'agit ensuite d'en dégager des informations intéressantes qui permettent de s'assurer du bon déroulement de la campagne.

Les agents en charge des missions de contrôle de QOS ont exprimé le besoin de pouvoir avoir une lecture statistique simple de la campagne en cours avec des indicateurs sur la QOS voix/sms et data. De plus, j'ai proposé de pouvoir avoir une vision géographique simple de l'avancement de la campagne via une carte interactive qui présente les territoires déjà couverts par les prestataires. Cette application permet d'automatiser des manipulations que les agents de l'ARCEP faisaient à chaque réception de nouveaux fichiers directement sur Excel et de garder une traçabilité dans le temps. Cette application est aussi déployée sur les serveurs internes de l'ARCEP et est donc accessible par tous via un lien intranet qui permet un meilleur échange d'informations.

L'application compte donc trois volets, un premier volet d'upload des données qui sont ensuite envoyées au serveur afin de réaliser les pré-traitements et de ne garder que les informations importantes. Dans un second volet, l'utilisateur de la plateforme doit pouvoir choisir entre les différents

fichiers présents dans le système afin de pouvoir les consulter. Et donc le dernier volet est une partie de consultation des informations une fois que l'on a choisi les données. Cette partie de présentation des informations sera détaillée dans le dernier point de la présentation du projet. Nous allons tout d'abord nous concentrer sur la mise en place du pré-traitement des données intermédiaires rendue difficile par la grande diversité des données des enquêtes de QOS.

2.1.2 La diversité des données d'enquêtes QOS

Les données de collecte sont à plusieurs dimensions. Tout d'abord, se présentent deux volets principaux mentionnés précédemment, la voix/sms et la donnée (qui sera mentionnée dans la suite de cette partie comme « data »). La voix (appels téléphoniques classiques) et les sms sont assurés par des technologies plus anciennes (2G/3G, parfois 4G) et n'ont pas les mêmes spécificités de fonctionnement que la data. La data est en essor constant en France au cours de ces dernières années et il est donc très important pour les utilisateurs de pouvoir mesurer la qualité de service associée, d'où les mesures spécifiques mises en place par l'ARCEP.

Ensuite, pour parler plus en détail des tests effectués, ils sont de plusieurs types. Dans un premier temps les tests binaires (succès/échec). On retrouve dans cette catégorie le fait de pouvoir tenir un appel, envoyer un sms en moins de 10 secondes et pour la partie data le chargement d'une page web ou d'une vidéo en moins de 5 ou 10 secondes.

Dans un second temps, les tests quantitatifs permettent de donner une indication chiffrée sur la qualité de la mesure. Il peut s'agir d'échelles standardisées comme le MOS (Mean Opinion Score), qui permet de noter la qualité d'un échange vocal ou tout simplement de mesures classiques comme un débit en mégabits par seconde (Mbps). Nous allons revenir plus en détail sur les unités de mesure au moment de présenter les pré-traitements.

Pour résumer concrètement les données reçues par les agents, elles se présentent sous la forme de trois fichiers Excel distincts. Un premier fichier comporte les informations concernant les enquêtes de voix/sms, un deuxième concerne les enquêtes data et un troisième traite des enquêtes sur les grands axes de transport. En fin de campagne, ces fichiers regroupent l'ensemble des mesures faites et sont donc assez volumineux (2/3G pour l'enquête 2021 si l'on compte data, voix et transports par exemple).

Ces fichiers comportent plusieurs feuilles Excel avec en commun à tous les fichiers une feuille « BDD » contenant l'ensemble des mesures et une feuille « Lexique » expliquant les différents champs utilisés dans la feuille BDD. Accessoirement, le prestataire peut aussi fournir des feuilles avec des informations ciblées comme certains KPIs déjà mis en place, un focus sur tel ou tel type de test ou technologie (exemple : focus sur la 5G).

Pour la mise en place de l'application, je me suis concentré sur la feuille « BDD ». Ce tableau Excel regroupe l'ensemble des mesures de la campagne, tous types de tests et opérateurs confondus.

Une des premières étapes de l'implémentation WEB est donc d'alléger le fichier et de le subdiviser afin de pouvoir faire des appels simples et rapides depuis la plateforme WEB. Un des objectifs étant d'avoir une plateforme interactive, on ne peut pas charger l'ensemble de la base de données vers le navigateur à chaque requête.

Afin d'alléger la base de données interrogée par l'application web, la première chose à faire est de séparer les mesures par opérateur. On se concentre sur les mesures d'un des opérateurs du fait que l'utilisateur de la plateforme ne souhaite visualiser que les résultats d'un opérateur à la fois. Dans un second temps, on peut garder seulement les colonnes intéressantes pour la présentation des

résultats que l'on souhaite. En effet, pour chaque mesure réalisée (qui correspond à une ligne dans le fichier), des centaines d'informations sont retirées mais seulement une poignée nous est vraiment utile. Les informations en question sont détaillées dans les tableaux ci-dessous. La figure 2.2 recense les données communes aux relevés voix/sms et data afin d'identifier les caractéristiques notamment géographiques des mesures. Les figures 2.3 et 2.4 montrent quant à elles les données permettant d'évaluer les mesures de voix/sms avec le MOS moyen et les débits ascendants et descendants des mesures de data.

Nom de la donnée	Type	Traitement(s) effectué(s)	Description
No de la mesure	Int	--	Donné par le prestataire
Usage	String	Filtre none	INDOOR, OUTDOOR, INCAR, ...
X_lambert_93	Float	Changer le système de coordonnées de WGS84 vers Lambert 93 et isoler les x et y dans des colonnes séparées	Coordonnées de la mesure
Y_lambert_93	Float		
INSEE_COM	Int	Jointure spatiale entre shapefile des communes et coordonnées mesures	Code Insee de la commune de prise de la mesure
Density_code	Int	Jointure entre la colonne INSEE_COM du fichier original et celui de l'INSEE	Code de densité INSEE allant de 1 à 7 (1 étant très dense et 7 peu dense)
Niveau de couverture mobile	String	Jointure spatiale entre les cartes de couvertures et les coordonnées mesures	Niveau allant de TBC (très bonne couv), BC (bonne couv), à CL (couverture limitée)
Département	Int	Lien avec la colonne INSEE_COM	Code du département de prise de la mesure

FIGURE 2.2 – Liste des colonnes d'informations retenues pour l'application post pré-traitements et enregistrées dans la BDD serveur pour qualifier les mesures voix/sms et data

Nom de la donnée	Type	Traitement(s) effectué(s)	Description
Couple	String	Uniformisation des strings	Opérateur appel émis et reçu. (ex : BOUY-SFR)
Result	String	--	Résultat du test (succès, échec)
MOS Moyen	Float	--	Note de l'appel
Service	String	--	Type de test : voix (appel) ou envoi de sms

FIGURE 2.3 – Liste des colonnes d'informations retenues pour l'application post pré-traitements et enregistrées dans la BDD serveur pour évaluer les mesures voix/sms

Nom de la donnée	Type	Traitement(s) effectué(s)	Description
Appel	String	--	Opérateur demande émise et techno (BYT_4G)
Type Test	String	--	Si il s'agit d'un upload ou un download
debit_descendant	Float	--	Valeur en Mbps
debit_montant	Float	--	Valeur en Mbps

FIGURE 2.4 – Liste des colonnes d'informations retenues pour l'application post pré-traitements et enregistrées dans la BDD serveur pour évaluer les mesures data

FIGURE 2.5 – Liste des noms de paramètres et des différentes options que l'utilisateur doit rentrer pour valider l'upload (disponible en taille complète en Annexe C)

Pour en revenir à la plateforme et à son fonctionnement, un résumé des interactions utilisateur et des traitements est détaillé dans la figure 2.6 ci-dessous.

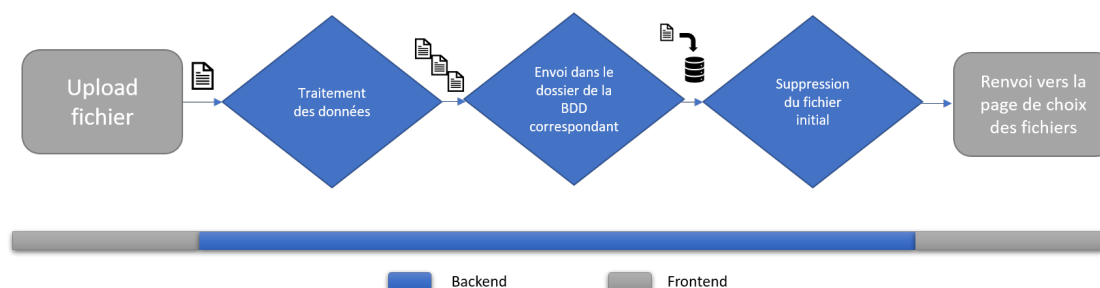


FIGURE 2.6 – Résumé des interactions entre le client et le backend au moment de l'upload d'un nouveau fichier QOS

Enfin, dans la figure 2.7 ci-dessous, vous pouvez retrouver l'organisation de la base de données sous forme de dossiers dans le serveur qui accueille l'application.

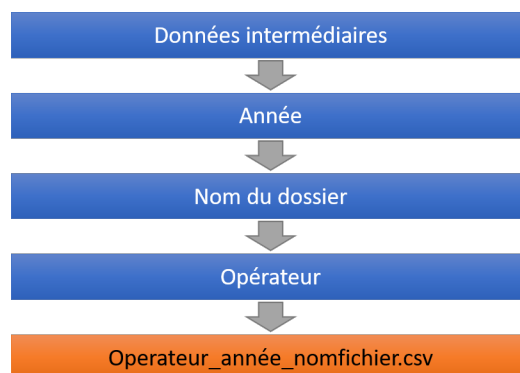


FIGURE 2.7 – Organisation des fichiers post traitement dans la base de données serveur

Toute la partie de traitement des données se fait via des scripts Python qui permettent de lire et écrire des fichiers .csv et .xlsx. Le lien avec la partie web de l'application se fait grâce au framework Flask pour Python. Les échanges entre la partie client et la partie serveur se font via de simples requêtes http. Les fichiers sont enregistrés en dur sur le serveur, les fichiers originaux sont supprimés après les traitements pour ne pas encombrer le serveur.

Côté web, une gestion minimale des données avant de les présenter est nécessaire. Cette gestion se fait via le langage Javascript sur la partie client. Enfin, l'affichage est géré en CSS et HTML. La présentation du produit fini se fait dans la partie suivante.

2.1.3 Discussion sur le rendu

Comme évoqué dans la partie précédente, la complexité de la mise en place de cette application repose principalement sur sa nécessité de pouvoir s'adapter aux différents types de données que l'on peut trouver d'une année sur l'autre et en fonction du prestataire.

Cette application était la première que je réalisais avec le framework Flask qui est utilisé occasionnellement à l'ARCEP même si la plupart des applications du même type utilisent le langage R et son extension WEB R-Shiny.

Les objectifs initiaux ont tous été complétés, et l'application est déployée sur le serveur interne de l'ARCEP. La mise en place de l'application aurait pu se faire sur une méthode se rapprochant plus de la méthode SCRUM vue lors du Master TSI à l'ENSG. En effet, il n'y a pas eu d'échanges entre le recueil du besoin et la présentation de la première version de l'application aux agents. Cela a entraîné de fortes révisions qui ont rallongé la durée du projet de 2 semaines à 3 semaines et demie pour le déploiement final.

L'application est disponible sur le logiciel Gitlab interne à l'ARCEP et pourra être entretenu par les personnes ayant supervisé la partie technique de la mise en place de ce projet.

Des screenshots de l'application et surtout de la partie de présentation des résultats est disponible en Annexe [C](#)

2.2 Les sites mobiles et le programme de dispositif de couverture ciblée

La connexion mobile en France se fait via des antennes mobiles aussi appelées sites mobiles. En France métropolitaine, on recense 98 458 sites mobiles au second trimestre 2022. Ce chiffre peut paraître important, cependant, il reste de nombreuses collectivités qui n'ont pas d'accès sûr aux réseaux mobiles.

En règle générale, les fréquences utilisées par les opérateurs mobiles sont attribuées via des enchères par l'Etat, qui en est le propriétaire. Cela dit, en 2018, le Gouvernement, sur proposition de l'Arcep, avait alors fait le choix de prioriser l'objectif politique de cohésion des territoires dans les modalités d'attribution. Ainsi, plutôt que de privilégier un critère financier, l'Etat avait décidé d'orienter l'effort des opérateurs vers l'investissement, en prévoyant des obligations de couverture inédites par leur ambition. Les opérateurs ont alors pris des engagements forts, retranscrits ensuite sous forme d'obligations juridiquement contraignantes, dont la participation au « dispositif de couverture ciblée ». Nous allons revenir dans le détail sur ce dispositif dans la partie de présentation des besoins et objectifs qui suit.

2.2.1 Présentation des besoins et objectifs

Le Dispositif de Couverture Ciblée (DCC), consiste en l'obligation pour les opérateurs de déployer 5 000 nouveaux sites chacun pour couvrir des « zones dans lesquelles un besoin d'aménagement numérique du territoire a été identifié par les collectivités et le gouvernement » [6]. Ce programme se construit donc en accord avec les collectivités locales qui font remonter des besoins de couverture sous forme de points d'intérêts (POI) ou de zones, ces besoins sont revus par les opérateurs qui jugent s'ils sont pertinents au vu des programmes de déploiements qu'ils avaient déjà mis en place. Si les opérateurs valident la demande, la zone à couvrir est inscrite dans un arrêté par le ministre et les opérateurs ont alors une obligation légale de 24 mois pour couvrir la zone avec un ou plusieurs nouveaux sites selon les besoins. Au 31 décembre 2021, 3 677 sites avaient été demandés par le gouvernement et déjà 1 294 ont été mis en service.

Un agent est responsable au sein de l'UCIM de ce programme et s'assure du bon déroulé du dispositif en faisant le lien entre les opérateurs, les collectivités et l'ARCEP.

L'ARCEP tient ici son rôle de régulateur en s'assurant du respect des obligations des opérateurs. Pour cela, elle publie régulièrement en open-data l'avancée du programme et suit en interne l'organisation des déploiements.

Lors de mes entretiens avec les membres de l'équipe mais surtout par nos échanges réguliers sur le sujet, j'ai pu recueillir un besoin d'avoir une représentation cartographique et sous forme de graphiques de l'avancée du DCC sous toutes ses coutures.

Les représentations sur le site arcep.fr [6] de l'avancée du programme ne prennent qu'une « photographie » à un temps donné du dispositif. L'équipe souhaitait pouvoir voir son évolution dans le temps sur une carte. De plus, l'équipe souhaitait pouvoir avoir sur la même représentation les POIs soulevés par les collectivités et inscrits dans les arrêtés et les sites effectivement déployés.

Un des derniers objectifs était de pouvoir relier graphiquement un site au(x) POI(s) correspondant(s) afin de calculer des statistiques et de se rendre compte des distances entre les POIs et les sites implémentés.

2.2.2 La visualisation graphique pour trouver les erreurs

Pour cette seconde partie, j'ai décidé de concentrer la présentation sur un aspect qui n'avait pas été envisagé au moment de poser les objectifs.

Mais revenons rapidement sur le processus permettant de mettre en place une cartographie des sites et des POIs du programme DCC. Les données sont séparées en deux fichiers Excel entretenus par un agent de l'UCIM, l'un recueillant les zones et POIs mis en place dans les arrêtés ministériels et l'autre, plus large, concentrant l'ensemble des sites de France métropolitaine avec une colonne indiquant ou non leur appartenance à un programme d'Etat tel que le DCC.

Toutes ces données sont géolocalisées via deux colonnes de coordonnées dans le système Lambert 93 qu'il faut transformer en WGS84 afin de pouvoir les intégrer dans l'outil de représentation de cartes dynamiques Leaflet.

Comme pour l'application des enquêtes de QOS, cette application a une partie en Python reliée à la partie WEB via le framework Flask. Ici, la mise à jour des données directement via un upload sur l'application n'a pas été mise au point et les fichiers devront donc être changés directement sur le serveur. A terme, avec la mise en place de base de données PostgreSQL, l'objectif est de simplement obtenir les données avec une connexion vers la base de données, ce qui pourra mettre à jour automatiquement l'application. Ces processus ne sont pas encore mis en place et les bases de données SQL existantes n'étant pas à jour, j'ai préféré passer directement par les fichiers au format Excel. Il s'agit cependant d'une piste d'amélioration.

La figure 2.8 ci-dessous montre la représentation qui a été choisie pour les POIs, les sites et la façon de les relier.

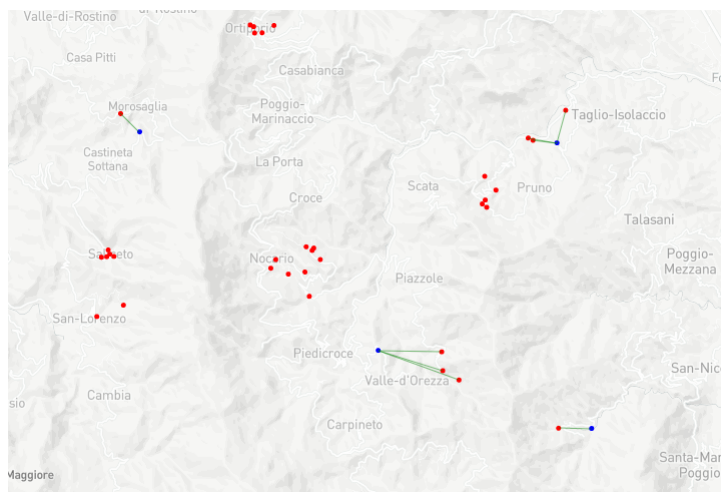


FIGURE 2.8 – Capture d'écran de l'application mise en place pour montrer les POIs, les sites et les distances associées

Comme on le voit, les sites sont représentés en bleu et les POIs en rouge, ces derniers sont reliés via un numéro de dossier commun qui permet de tracer les lignes vertes que vous voyez. Cette façon de représenter les liaisons sites-POIs a permis d'identifier les décalages dans les numéros de dossiers et les incohérences d'association.

La figure 2.9 ci-dessous montre un exemple de zone où les numéros de dossiers de POIs et de sites étaient défectueux. Ce type d'erreur est très difficile à repérer uniquement avec les tableurs Excel comptabilisant les sites et les POIs séparément.

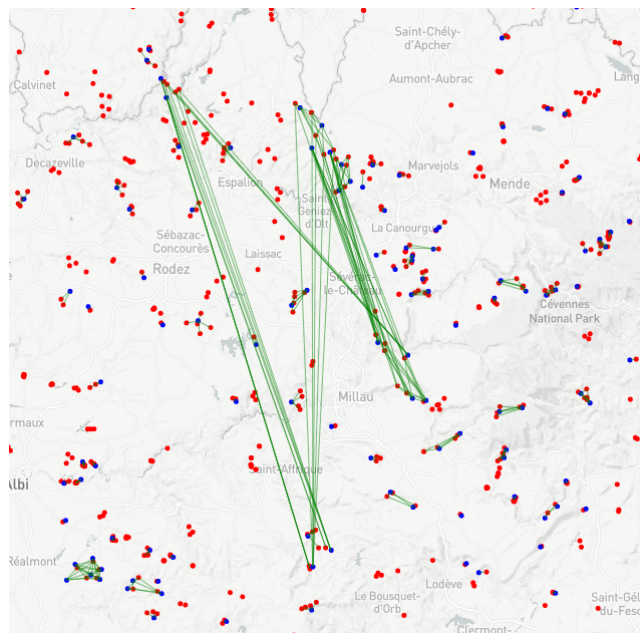


FIGURE 2.9 – Problèmes d'association entre les numéros de dossier des POIs et des sites associés

Ces erreurs ont été repérées à temps avant la publication sur l'Open data des données et signalées aux opérateurs qui ont pu les corriger.

2.2.3 Présentation du résultat

Le rendu final se compose de trois parties (voir Annexe D). Dans une première partie et conformément aux objectifs, deux sliders permettent de choisir des intervalles d'études à représenter. Un slider est utilisé pour les POIs provenant des arrêtés et un autre pour les sites déployés ce qui permet une représentation dans le temps. La deuxième partie est donc la carte des sites et POIs à afficher. La troisième partie est un graphique récapitulatif de l'évolution des déploiements de sites et des inscriptions de POIs dans le dispositif de couverture ciblée.

Cette application souffre encore de temps de chargement un peu long et d'une réactivité qui n'est pas parfaite dans les déplacements sur la carte mais cela vient de la limitation des outils utilisés face à des jeux de données couvrant toute la France avec au deuxième trimestre 2022 8 601 POIs signalés et 1588 sites déployés.

La plateforme a permis de trouver une distance moyenne entre les POIs et les sites de 3km et une distance médiane de 1.2km, ce qui démontre que les opérateurs semblent bien faire l'effort de déployer les sites au plus près des zones identifiées dans les arrêtés.

Si elle est reprise et entretenue dans le futur, cette application peut servir de tableau de bord interne à l'ARCEP permettant de suivre l'évolution du programme DCC qui doit s'étendre encore sur quelques années.

La deuxième partie de présentation des projets que j'ai pu effectuer au cours de ce stage se concentre sur des études ponctuelles avec un rendu sous la forme de graphiques ou de rapports avec mon maître de stage permettant de mettre en lumière des indicateurs sur les déploiements mobiles. Les deux analyses que j'ai choisi de présenter sont transverses à la Direction Mobile et Innovation et se basent sur un recoupement de données d'origines diverses.

Ces projets d'analyse se concluent sur une restitution ponctuelle en réunion avec mon maître de stage mais j'ai pris un grand soin de documenter les codes que j'ai pu utiliser et de les rendre attractifs même pour une personne qui ne serait pas experte avec le langage Python grâce à l'utilisation des notebooks Jupyter [8]. Cet outil permet de partager des travaux de code de manière simple, claire et compréhensive.

Dans un premier temps, je vais revenir sur l'analyse de l'implantation de sites du DCC présenté dans le [chapitre 2](#). Ce sujet est transverse à mon unité : l'UCIM ainsi qu'à l'UMM et l'UGF, les Unités de Régulation des Marchés Mobiles et de Gestion des Fréquences.

Dans un second temps, je présenterai le projet que je suis en train de finaliser au moment de l'écriture de ce rapport. Ce dernier se concentre sur la relation entre largeur de bande disponible et densité de population ainsi que sur les technologies disponibles selon la densité de population.

Contrairement au [chapitre 2](#), je vais ici expliquer l'entièreté des démarches d'analyses qui ont permis de produire des résultats utiles aux différentes unités mentionnées ci-dessus. Cela permettra notamment, dans un souci de reprise de ces analyses, de pouvoir suivre l'entièreté du processus accompagnant le code.

3.1 Etude exploratoire sur l'implantation des sites

Ce projet est un projet exploratoire et ambitieux. Le but annoncé est de mettre en place un algorithme permettant de trouver des points d'implantation optimaux afin de satisfaire l'ensemble des besoins de couverture mobile des collectivités territoriales. Concrètement, il s'agit de se servir du programme DCC et d'inverser son fonctionnement. Au lieu que les collectivités ne fassent remonter des besoins en couverture mobile, l'étude doit pouvoir identifier les collectivités où l'implantation d'un nouveau site serait le plus bénéfique même si elles n'en ont pas fait la demande.

3.1.1 Objectifs de l'étude

Pour rappel, la France métropolitaine est l'un des territoires européens et même mondiaux les mieux couverts par ses opérateurs. D'après les données du site monreseau.mobilite.arcep.fr [4] en 2022, 99% de la population française de métropole et 94% du territoire sont couverts par un opérateur téléphonique. Les données complètes de couverture datant de septembre 2020 sont disponibles en Annexe E à la fin de ce rapport.

Cette couverture mobile affichée laisse cependant encore des territoires avec une couverture limitée. Dans le monde actuel, une couverture mobile limitée signifie moins d'opportunités et un handicap important pour les populations affectées. Il est du rôle de l'ARCEP d'aider ces populations à se doter des mêmes moyens que le reste de leurs concitoyens. Cela entraîne la mise en place de programmes tels que celui du DCC et aussi la création d'indicateurs permettant d'identifier des moyens efficaces de parvenir à résoudre ces disparités. L'étude présentée ici s'inscrit complètement dans cette démarche.

Les objectifs de l'étude sont doubles. D'une part, le projet est concordant avec le projet DCC. Un premier rapport doit pouvoir identifier les caractéristiques qui font qu'un site est un site DCC parmi la liste de tous les sites présents sur le territoire métropolitain. D'autre part, la phase exploratoire doit se servir des résultats de la première partie afin d'identifier les zones les plus propices à une implantation de nouveaux sites. Le programme DCC étant déjà en place, les résultats de ces études ne serviraient pas directement à la mise en place de nouveaux sites mais permettraient d'avoir de nouveaux indicateurs pour l'ARCEP.

En ce qui concerne les données utilisées, l'objectif est d'accumuler le plus d'informations possibles sur les sites mobiles. On part donc de la base de données des sites mobiles qui recueille des informations sur les plus de 96 000 sites déployés en France tous opérateurs confondus. Sur ces sites, 1 588 sont identifiés comme sites DCC et 10 576 sites font partie du programme Zone Blanche Centre-Bourg (ZBCB), l'ancêtre du programme DCC. Ces sites sont donc des sites mis en place sous incitative de l'Etat et vont nous servir de points de références pour identifier les caractéristiques d'un site non choisi initialement par les opérateurs d'un point de vue économique mais qui sert une fonction pour le grand public.

Ensuite, j'ai essayé d'agréger un maximum de données sur les territoires desservis par ces sites mobiles. J'ai donc utilisé des données de l'Institut National de la Statistique et des Etudes Economiques (INSEE). Un des facteurs qui nous paraissait a priori des plus important est la densité de population. Une nouvelle grille INSEE à 7 niveaux de densité permet de classer les communes en fonction de ce paramètre et j'ai donc utilisé ces données.

Les autres données INSEE disponibles l'étaient au niveau de la commune et du département et sont résumées dans le tableau 3.1 ci-dessous :

Nom de la donnée	Echelle	Description
Insee_dep	Département	Département de la commune d'implantation du site
ADM_INSEE_REG	Région	Région de la commune d'implantation du site
Num_insee_zone_density	Commune	Densité de population de la commune d'implantation du site
Pop_2019	Commune	Population de la commune d'implantation du site
Part_moins_25a	Commune	Part des habitants de moins de 25 ans de la commune d'implantation du site
Part_25a_64a	Commune	Part des habitants de 25 à 64 ans de la commune d'implantation du site
Part_plus_64a	Commune	Part des habitants de plus de 64 ans de la commune d'implantation du site
Partagriculteurs	Département	Part d'agriculteur dans le département de la commune d'implantation du site
Part_cadres	Département	Part de cadres dans le département de la commune d'implantation du site
Programme	Site	Valeur cible, permet de savoir si le site étudié fait partie ou non d'un programme d'état

FIGURE 3.1 – Tableau permettant de résumer les données utilisées dans l'analyse des composantes des sites mobiles

Afin de pouvoir utiliser l'ensemble des compétences apprises lors de ma scolarité à l'ENSG, j'ai proposé de monter un protocole de machine learning afin de trouver dans un premier temps la « fiche type » du territoire où un site faisant partie d'un programme d'Etat est implémenté. Dans un second temps, j'ai réutilisé la méthodologie présentée dans l'article "Maximum entropy modeling of species geographic distributions" de S. J. Phillips, R. P. Anderson et R. E. Schapire [11] et l'ai adapté afin d'essayer d'en tirer des implantations optimales de sites sur le territoire métropolitain. Les étapes de ces deux analyses sont détaillées ci-dessous.

3.1.2 Explication de la démarche par machine learning

Pour la première étude, nous nous concentrons sur la liste des sites en y agrégeant les données INSEE en fonction de la commune d'implantation du site. J'ai choisi sur la base de mes connaissances et des projets que j'ai pu réaliser à l'ENSG d'utiliser la librairie scikit-learn [7] afin de faire cette

analyse. Après plusieurs tests avec des modèles de classification différents, j'ai décidé de choisir un algorithme de régression logistique de la librairie scikit-learn [7].

Concrètement, on divise le jeu de données en deux parties, un training et un testing set. Notre variable Y, la cible de l'algorithme, c'est à dire la variable que l'on cherche à prédire, est la colonne du dataframe contenant l'information « site faisant partie d'un programme » identifiée par un 1 si le site appartient à un programme mis en place par l'Etat et un 0 dans le cas contraire. L'algorithme de régression logistique s'entraîne d'abord sur le training set et essaye de prédire sur le test set l'appartenance d'un site ou non à un programme. On compare ensuite les résultats effectifs entre la vérité terrain du test set et les prédictions de l'algorithme. Une fois que l'on a trouvé un modèle et un paramétrage qui correspondent à notre jeu de données, on regarde quels paramètres l'algorithme a utilisés afin de tirer une sorte de « fiche du site type appartenant à un programme ».

L'ensemble des étapes de l'algorithme ainsi que les technologies employées sont résumés dans la figure 3.2 ci-dessous.

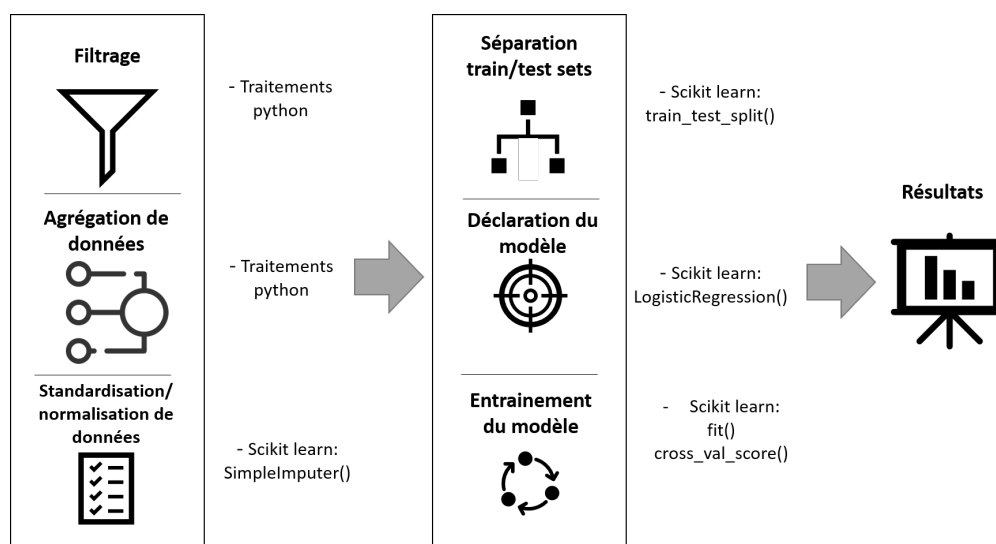


FIGURE 3.2 – Schéma décrivant les étapes de l'algorithme mis en place pour la partie 3.1.2 par site

Les résultats de cette étude seront présentés dans la partie suivante.

Pour ce qui est de la seconde étude, j'ai dû changer de méthodologie. En effet, on ne peut plus s'appuyer uniquement sur la base des emplacements des sites existants pour effectuer l'étude. Pour rappel, l'objectif est de trouver les zones où l'implantation d'un site serait la plus en adéquation avec un programme mais qui n'en aurait pas encore fait la demande.

Pour cela, et pour gagner du temps sur les calculs, j'ai commencé par me concentrer sur la région Bretagne. La méthodologie consiste à diviser la région en dalles de tailles égales et d'appliquer à chaque dalle, par une jointure spatiale et les librairies Python geopandas, pyproj et shapely, les attributs des données détaillées précédemment dans la figure 3.1. En ce qui concerne les sites, on prend chaque dalle et on additionne le nombre de sites qu'elle contient avec toujours la distinction entre simple site et site faisant partie d'un programme.

La figure 3.3 montre ce parcellement de la région.

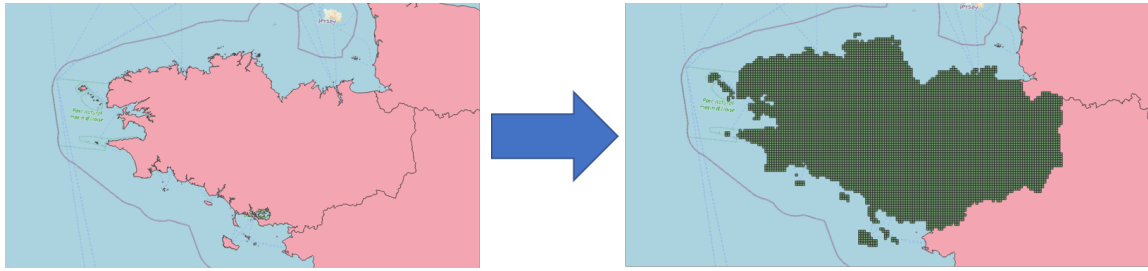


FIGURE 3.3 – Découpage du territoire en zones d'aires égales

J'ai ensuite essayé d'adapter l'algorithme utilisé dans l'article scientifique écrit par S. J. Phillips, R. P. Anderson et R. E. Schapire [11], qui utilise l'outil de modélisation OneClassSVM afin de déterminer une estimation de densité des sites appartenant à des programmes dans les dalles que nous avons construites. Comme pour la première étude, la figure 3.4 ci-dessous résume le processus employé.

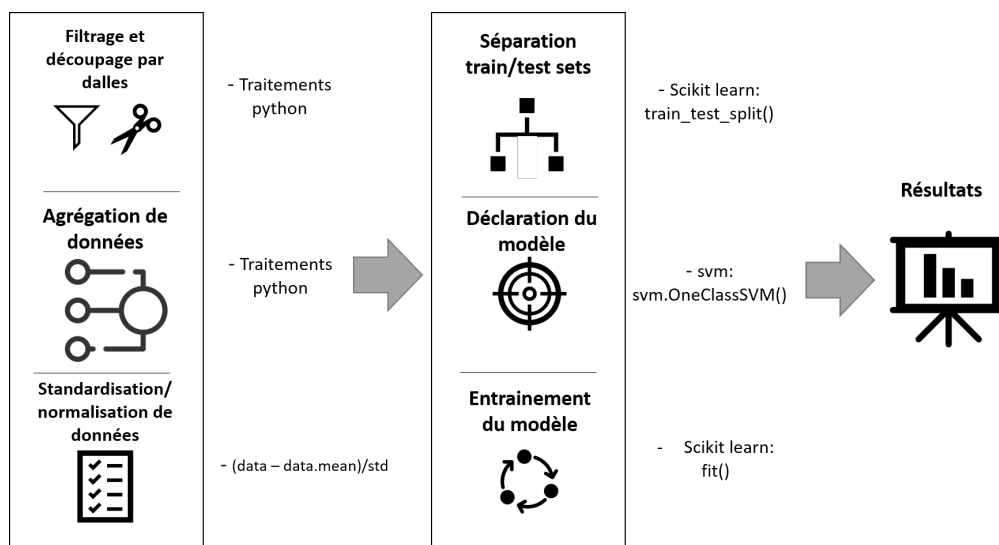


FIGURE 3.4 – Schéma décrivant les étapes de l'algorithme mis en place pour la partie 3.2 pour des zones données

Cependant, face à des difficultés à mettre en place l'adaptation de l'algorithme, je n'ai pas pu avoir de résultats concluants et il a été décidé avec mon maître de stage de passer à un autre sujet quitte à réessayer la procédure en fin de stage. Je reviendrai en détail sur les points qui ont bloqués dans la partie de discussion des résultats qui suit.

3.1.3 Présentation et discussion des résultats

Pour la partie d'identification des sites, les résultats suivants ont été déterminés. Dans un premier temps, on retrouve la matrice de confusion dans la figure 3.5 ci-dessous.

Matrice de confusion	Positive	Negative
Positive	30670	1416
Negative	3130	2917

FIGURE 3.5 – Matrice de confusion de l'algorithme d'identification de sites.
SPECS : [LogisticRegression(penalty = 'l2', C = 100, solver = 'lbfgs', max iter = 15000)]

D'après la matrice de confusion, on se rend compte que l'algorithme est efficace pour retrouver les sites ne faisant pas partie d'un programme mais qu'il y a beaucoup trop de faux négatifs. C'est-à-dire que l'algorithme identifie trop souvent un site ne faisant pas partie d'un programme comme l'étant.

Cela est validé par le « classification report » fourni par la librairie scikit-learn [7] visible dans la figure 3.6 ci-dessous. De ces données, on identifie que la proportion de sites identifiés par l'algorithme comme faisant partie d'un programme (Ligne « 1 » du tableau) est bien trop faible (67%) pour tirer la moindre conclusion. Comme précisé précédemment, ces scores bas sont malgré tout les meilleurs que j'ai pu retirer en utilisant plusieurs types d'algorithmes de classification. Une piste d'amélioration serait d'agréger beaucoup plus de données que celles présentées dans le tableau de la figure 3.1. Il faut aussi trouver des données qui sont plus susceptibles d'influer sur la présence ou non d'un site puisqu'en sortie de cette étude, seule la densité de population s'est révélée être un facteur influant.

	precision	recall	f1-score	support
0	0.91	0.96	0.93	32086
1	0.67	0.48	0.56	6047
accuracy			0.88	38133
macro avg	0.79	0.72	0.75	38133
weighted avg	0.87	0.88	0.87	38133

FIGURE 3.6 – Classification report de l'algorithme d'identification de sites.
SPECS : [LogisticRegression(penalty = 'l2', C = 100, solver = 'lbfgs', max iter = 15000)]

Pour la seconde partie d'étude des implantations potentielles, je suis tombé sur les mêmes problèmes. Les paramètres sur chaque dalle influaient trop peu sur le résultat final et je pense ne pas avoir assez bien adapté l'algorithme de S. J. Phillips, R. P. Anderson et R. E. Schapire présent dans les exemples de la librairie scikit learn [7]. Après plusieurs tentatives infructueuses et les temps de calculs élevés dus aux grands nombres de géométries sur lesquels faire des jointures spatiales et des calculs, mon maître de stage et moi avons pris la décision d'arrêter ce projet prématurément.

Malgré l'absence de résultats pour ces études, les processus de préparation de la donnée sont en place et cela peut aider de futures tentatives de passer par cette méthode afin de réaliser une étude

« from the top downwards » dans le futur. Les codes ne sont pas à jeter et seront peut-être réutilisés dans le futur.

Au cours de la suite de mon stage, j'ai rencontré de nouveaux types de données pouvant s'inclure dans un projet de ce type telles les données « Pitney-Bowes » [9] de populations fournies par un prestataire à l'ARCEP. J'aurai l'occasion de travailler avec ces données très complètes dans le projet que je vais présenter dans la suite de ce rapport. L'idée toujours sur la table au moment de l'écriture de ce rapport est de reprendre la démarche de machine learning en fin de stage avec une nouvelle approche du découpage du territoire sous la forme d'hexagones contenant des données très précises et complémentaires à celles présentées dans le tableau de la figure 3.1 pour voir si la précision s'améliore et les résultats sont plus concluants.

Un autre point à soulever est que les données de « sites faisant partie d'un programme d'Etat » se reposent majoritairement sur les sites du programme DCC qui sont toujours en cours de déploiement avec près de 1 500 sites déployés. A terme, ce seront 5 000 sites déployés qui pourront venir enrichir le « training set » de l'algorithme et donc assurément augmenter sa fiabilité. De plus, les critères qui font qu'une communauté territoriale fait inscrire une zone dans un programme d'Etat ne sont pas uniformes.

Pour conclure, cette étude exploratoire s'est soldée par un échec de résultats mais a posé les bases pour de futures études et a surtout été une tentative qui, si elle avait fonctionné, aurait pu apporter beaucoup aux différents plans de déploiements mis en place par l'ARCEP.

3.2 Capacité/Largeur de bande et qualité de service

3.2.1 Rappels des technologies mobiles

En France, la propriété et la gestion des fréquences hertziennes appartient à l'Etat. L'utilisation des bandes de fréquences est très large et va des ondes radios aux bas infrarouges permettant les observations satellites par exemple. Leurs attributions et leur gestion, en ce qui concerne la téléphonie mobile, est géré conjointement par l'ARCEP et par l'agence nationale des fréquences (ANFR).

L'attribution de ces fréquences aux opérateurs, dont nous avons brièvement parlé dans la partie 2.2 de ce rapport, se fait via des AUF (Autorisations d'Utilisation de Fréquences) généralement pour 10, 15 ou 20 ans selon les cas.

Au sein de la DMI, l'unité gestion des fréquences (UGF) est en charge de la partie technique de leur attribution et de leur surveillance et toutes les autres unités sont liées de près ou de loin à leur gestion.

La répartition et l'utilisation de ces fréquences est résumé dans le tableau de la figure 3.7 ci-dessous. L'objectif soulevé par mon maître de stage était de croiser les données de densité de population avec les largeurs de bandes de fréquences disponibles. En effet, en fonction des technologies, une plus grande largeur de bande entraîne une meilleure qualité de service car plus de bande passante est disponible pour l'utilisateur. Cela dit, si la zone étudiée est très densément peuplée, et que les utilisateurs font usage de leur mobile au même moment, la bande passante disponible est une valeur finie et la qualité de service est donc réduite pour chaque utilisateur.

**LES FRÉQUENCES : LES BANDES PIONNIÈRES DE LA 5G
ET LES AUTRES BANDES ATTRIBUÉES AUX OPÉRATEURS**

Fréquences	Technologies utilisées actuellement en France métropolitaine	Date	Pénétration à l'intérieur des bâtiments	Portée	Débit maximum théorique
700 MHz	4G et 5G	Attribuée en 2015	★★★★	★★★★	★
800 MHz	4G	Attribuée en 2012	★★★★	★★★★	★
900 MHz	2G et 3G	Attribuée en 1986	★★★★	★★★★	★
1,8 GHz	2G, 4G	Attribuée en 1994	★★★	★★★	★★
2,1 GHz	3G, 4G et 5G	Attribuée en 2001	★★★	★★★	★★
2,6 GHz	4G	Attribuée en 2012	★★	★★	★★
3,5 GHz	5G	Attribuée en 2020	★★	★★★	★★★

Source : Arcep

FIGURE 3.7 – Répartition des bandes de fréquences mobiles et technologies en France métropolitaine en 2022

Le but de cette étude est de produire des cartes et des graphiques permettant de se rendre compte de la répartition territoriale de l'utilisation des différentes bandes de fréquences par les opérateurs. Un des phénomènes à objectiver est de déterminer si les territoires ruraux déjà couverts, qui disposent d'une largeur de bande disponible plus faible (soit du fait de l'éloignement plus important entre les sites mobiles déployés, soit parce que les opérateurs ne déploient pas toutes leurs bandes de fréquences sur les sites en zones peu denses), bénéficient d'une qualité de service comparable aux

zones denses.

Si cela était confirmé, comme semblent le prouver plusieurs analyses internes de l'UCIM, cela signifie que l'enjeu pour l'accès au service mobile en zone rurale relève essentiellement de la couverture (il faut déployer plus de sites) et non de la capacité des réseaux mobiles (pas besoin d'augmenter la bande passante des sites déjà existants en zones rurales).

Les données utilisées seront encore une fois les données INSEE de densité de population. Une commune classée comme ayant une densité très forte est classée 1 tandis qu'une commune avec une densité faible est classée 7.

L'autre jeu de données utilisé est la base de données des sites mobiles en France métropolitaine. Comme évoqué sur d'autres parties de ce rapport, ce jeu de données recense sous la forme d'un tableur Excel contenant une ligne pour chaque site référencé avec ses spécificités. Il rassemble aussi sous un même fichier les données des quatre opérateurs présents en France métropolitaine : Bouygues Télécom, Free mobile, Orange France et SFR. Pour cette étude précise, nous retiendrons les coordonnées d'implantation des sites ainsi que les fréquences installées sur chaque site. Un site peut émettre plusieurs fréquences à la fois et selon les opérateurs, avoir plusieurs technologies (2G, 3G, 4G ou 5G) sur une même fréquence.

Une dimension temporelle est aussi incluse dans cette étude. Mon maître de stage souhaite voir l'évolution des statistiques sur les dernières années. Les données de sites sûres les plus anciennes remontent à 2018 et cela coïncide aussi avec la mise en place du « New deal », on souhaitera donc voir l'évolution entre 2018 et 2022.

Les rendus de cette étude sont de deux types différents. Dans un premier temps, des cartes imprimées recensant les technologies déployées par opérateur et par bande de fréquence. Dans un second temps, des diagrammes avec en abscisse les 7 types de communes par densité et en ordonnées des décomptes de sites ou de fréquences associées. Je vais présenter dans le détail la démarche que j'ai suivie dans la partie suivante.

3.2.2 Présentation des données et de la démarche

A l'image de toutes les autres études présentées dans ce rapport, la première phase de ce projet consiste en une phase de data engineering afin de nettoyer les données brutes et de les optimiser pour le reste des analyses. Tous les traitements sont ici effectués via Python et les notebooks Jupyter [8] dans un souci de réutilisation du code. Les pré-traitements consistent ici à retravailler les données de site en les séparant par opérateur et en agrégeant les données de densité de population à travers une jointure sur la colonne de code commune INSEE des sites et des communes référencées dans le fichier de densité. Cette phase permet aussi de ne garder que les colonnes d'intérêt pour l'étude dans la cinquantaine qui compose les données de sites.

A la fin de ces pré-traitements, on passe d'un fichier site de 126 Mégas à 4 fichiers (un par opérateur) de 3 Mégas chacun en moyenne. Cette étape est aussi importante car comme précisé précédemment, l'étude est à répéter sur différentes bases de données dans le temps. J'ai donc mis en place un processus automatisé permettant de réduire la taille des fichiers à utiliser et utilisable sur n'importe quelle base de données sites depuis 2018.

Trois types de rendus ont été mis en place pour cette étude.

Dans un premier temps, j'ai utilisé la librairie Basemap de matplotlib afin de réaliser des cartes d'implantation de sites en fonction de la technologie employée par opérateur et par fréquence. Un

exemple d'une de ces cartes est disponible dans la figure 3.8 ci-dessous et une plus large sélection avec notamment les évolutions temporelles en Annexe F. Pour réaliser cette étude, j'ai effectué des filtrages par opérateurs et par fréquences sur les bases de données afin de recueillir les technologies déployées dans chaque cas. Il a aussi fallu mettre en place des fonctions permettant de gérer la lecture de la colonne des technologies (2G, 3G, 4G, ...) qui avait sûrement été remplie à la main dans la base de données et n'étaient donc pas toutes au même format.

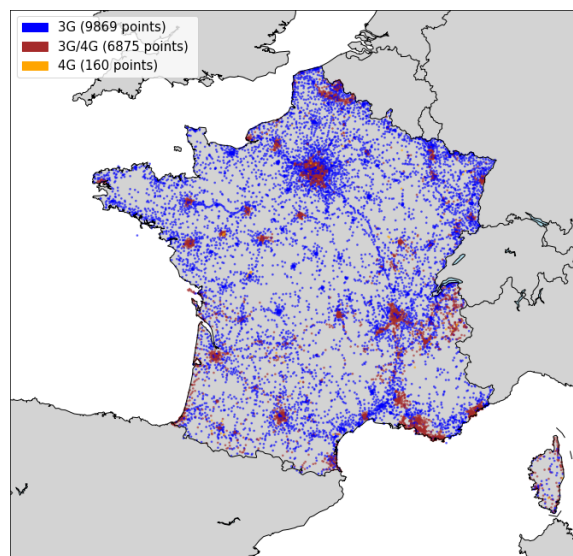


FIGURE 3.8 – Carte de répartition des technologies pour SFR en bande 2100 au troisième trimestre 2020

Le deuxième type de rendu est un diagramme qui a pour ordonnée le nombre de sites mobiles sur l'ensemble du territoire métropolitain en fonction de la densité de population. Ce diagramme est répliqué pour chaque bande de fréquence et pour chacun des opérateurs. Cela permet aux régulateurs d'analyser la répartition des allocations des bandes de fréquence en fonction de la densité de population (voir figure 3.9 ci-dessous). Les technologies déployées en fonction des bandes de fréquences sont aussi indiquées.

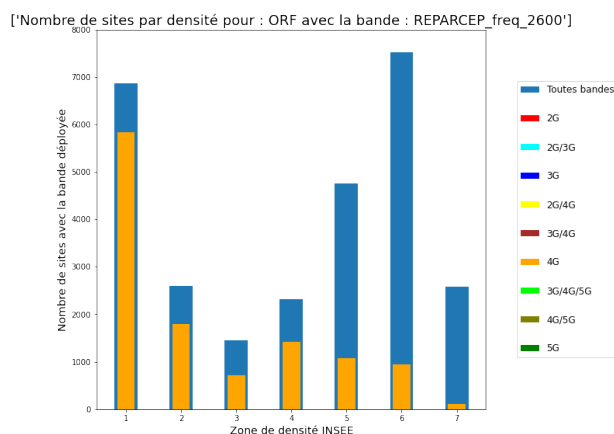


FIGURE 3.9 – Diagramme du compte de sites en fonction de la densité de population et des technologies au second trimestre 2022 pour l'opérateur Orange et la bande de fréquence 2,6 GHz

Ces diagrammes sont complémentaires aux cartes en reprenant notamment le code couleur pour les technologies employées. Ils apportent cependant un autre axe d'analyse qui sera détaillé dans la partie suivante de discussion des résultats. Enfin, comme avec les autres analyses, la dimension d'évolution dans le temps est intéressante et un exemple d'une telle évolution est disponible en Annexe F à la fin de ce rapport.

Le troisième type de rendu est une carte de chaleur (heatmap) par opérateur et par période comme montré dans la figure 3.10 ci-dessous. Comme pour les cartes d'implantation des sites, un exemple complet notamment d'évolution dans le temps est inclus dans l'Annexe F.

Sur l'axe y des heatmaps, on trouve les bandes de fréquences et sur l'axe x, toujours les densités de population. La valeur des cases représente la somme des sites qui tombent dans le recroisement de ces deux paramètres. Ce type de représentation a un avantage particulier pour rapidement identifier les « outliers » c'est-à-dire les données sous représentées qui n'apparaissent pas sur les diagrammes présentés ci-dessus. Une autre utilité de ce type de représentation est de pouvoir facilement visualiser l'évolution dans le temps grâce au code couleur et à la taille similaire des figures en fonction des années.

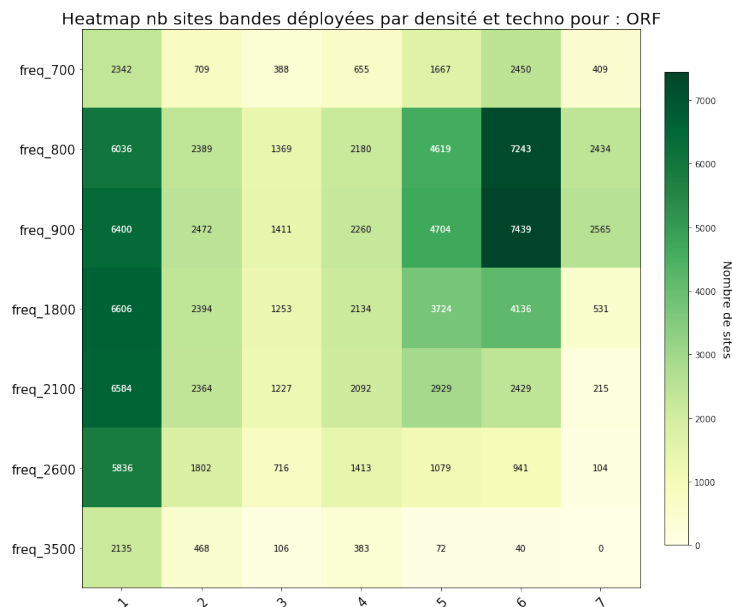


FIGURE 3.10 – Heatmap de la répartition des sites pour l'opérateur Orange France au second trimestre 2022

Toutes les analyses que l'on peut retirer de ces représentations sont présentées dans la partie de discussion des résultats concluant la présentation de ce projet.

3.2.3 Présentation et discussion des résultats

Les trois types de visualisation de la donnée présentés précédemment sont complémentaires et permettent aux agents de l'ARCEP de se reposer sur des axes de lecture différents selon l'information qu'ils recherchent. Dans le cas de ce projet, j'ai pu proposer plusieurs types de visualisation différents et ceux présentés ci-dessus ont été validés dans la démarche par mon maître de stage. Les cartes et les diagrammes en barres faisaient partie des demandes initiales mais j'ai aussi pris l'initiative de la représentation sous forme de heatmaps.

Pour les cartes, elles permettent une lecture large et visuelle de l'implantation des sites en fonction des fréquences et par opérateur sur le territoire. On peut suivre l'évolution dans le temps des fermetures d'anciennes technologies et de l'implantation de nouvelles comme la 5G sur la bande 3.5 GHz. Ces cartes ont vocation à être plus des outils de communication que d'analyse mais permettent d'avoir une vue d'ensemble.

Pour ce qui est des diagrammes, le croisement avec les données INSEE de population permet de juger efficacement de la répartition des technologies et des bandes de fréquences en fonction des territoires. Un des questionnements de mon maître de stage concernait l'utilisation de la bande 2.6 GHz en zone rurale, qu'il jugeait sous utilisée par les opérateurs, ce qui a été confirmé par l'étude (voir figure 3.9). Une des autres assumptions logiques qui a pu être confirmée est l'utilisation beaucoup plus forte de la bande 3.5 GHz, récemment inclus dans les AUF et spécifique à la 5G, dans les zones de fortes densités quand elle est quasi absente des autres types de territoires. Enfin, les heatmaps sont une concentration d'informations. Elles permettent un regard d'ensemble sur la situation avec un seul graphique pour toutes les longueurs d'ondes disponibles. Elles se révèlent donc très utiles pour analyser les évolutions temporelles. Ainsi, l'Annexe F montre bien pour Free mobile, une fermeture des dernières antennes 800 Mhz qu'ils avaient en 2020 et un renforcement de leur présence en zones moyennement denses entre 2020 et 2022.

Pour la suite de l'étude et afin de changer de référentiel de densité de population, je compte utiliser les données « Pitney-Bowes »^[9] évoquées précédemment dans ce rapport afin de faire une analyse par jointures spatiales entre les données de population et d'implantations de sites. Enfin, un recroisement avec les données de QOS est envisagé si le temps est suffisant.

Ces analyses ne sont qu'un échantillon de ce qui a pu être retiré de ces représentations mais elles permettent d'en illustrer les utilités.

4.1 Difficultés rencontrées

Au moment de la rédaction de ce rapport, j'ai effectué deux tiers de mon stage au sein de l'Autorité. Il est donc difficile de prendre du recul sur toutes les difficultés que j'ai pu rencontrer ou que je risque de rencontrer par la suite car, comme le montre le diagramme de GANTT établi à la fin du mois de juin et disponible en Annexe [A](#) je suis toujours en phase d'implémentation des projets.

Cela dit, après ces quatre mois écoulés, je peux revenir sur certaines difficultés que j'ai pu rencontrer et sur les méthodes que j'ai utilisées pour les surmonter.

La première difficulté assez évidente en arrivant dans l'unité Couverture et Investissements Mobiles, est que je n'avais quasiment aucune connaissance dans le monde des télécommunications. Malgré l'intitulé de mon stage qui consistait en une analyse géographique, j'ai dû me familiariser avec les terminologies et les technologies employées actuellement dans le secteur des télécoms français. Pour cela, j'ai été grandement aidé par le fait de faire partie et d'être au contact de l'UCIM au quotidien. Lors de ma phase de bibliographie, d'entretien et de familiarisation avec l'ARCEP, mes collègues ont toujours su apporter les réponses à mes questions et j'ai été très bien accompagné. Après quatre mois de stage, je suis aujourd'hui familier avec les terminologies et le fonctionnement général de la téléphonie mobile en France même si je ne comprends bien sûr pas encore tous ses aspects technologiques. Durant le début de mon stage, j'ai usé d'une approche pratique pour me familiariser avec le milieu en manipulant dès la première semaine des données de sites, de qualité de service et de couverture.

La deuxième difficulté de ce stage est la gestion du temps. Avec des projets multiples et très différents les uns des autres (tant au niveau des données que des interlocuteurs) qui s'enchaînent selon un planning défini, il m'a été très difficile de bien rentrer à 100% dans chaque projet et de réussir à séparer efficacement les tâches. En effet, lesancements des projets ont respecté le planning, contrairement à leurs clôtures. J'ai dû, à de nombreuses reprises, revenir sur un projet que je n'avais pas touché depuis plusieurs semaines afin d'y apporter des modifications au cours de l'utilisation ou de la relecture du projet par un de mes collègues. Cela a entraîné un retard important comparé au planning initial. Cela dit, sur ce point les échanges réguliers avec mon maître de stage ont pu me rassurer car il était compréhensif et avait prévu l'éventualité de ces retards. En début de stage, nous avions discuté de cette possibilité mais avons préféré garder un planning ambitieux. La plupart des travaux étant exploratoires et sachant que chacun d'entre eux aurait pu constituer un stage à lui tout seul, il fallait mettre des limites comme pour l'étude d'implantations potentielles des sites. La priorisation du travail était aussi essentielle pour gérer ce type de difficultés.

Toujours sur la gestion du temps, un aspect a été sous-estimé. Les projets que j'ai réalisés étaient tous analytiques avec un rendu relevant d'une mise en place de logiciel ou d'un rapport détaillé. Cela dit, j'ai sous-estimé la partie de data-engineering précédant la moindre possibilité d'analyse. Les jeux de données à l'ARCEP sont en cours de standardisation mais la plupart d'entre eux sont encore sous des formats .csv ou .xlsx avec de nombreuses disparités d'une année sur l'autre, des erreurs mineures

ou encore des notations différentes pour des produits similaires. La phase de gestion des exceptions, afin de standardiser et nettoyer les fichiers obligatoires avant chaque analyse, a donc pris énormément de temps et de ressources en début de stage. Cela dit, cette phase était très importante dans le sens où elle facilitera les futures intégrations de données et les codes pourront être réutilisés.

Enfin, un dernier point mineur est l'adaptation aux outils utilisés à l'ARCEP. Même si je connaissais la plupart des outils utilisés lors de ce stage grâce à ma formation à l'ENSG, j'ai dû me former aux outils utilisés en interne dont je n'avais pas l'habitude. Pour cela, j'ai pu compter sur le soutien et l'aide précieuse de l'unité de régulation par la donnée (URD) qui se compose de profils techniques et de géomaticiens avec qui j'ai donc pu échanger facilement.

4.2 Améliorations des outils

Du point de vue des pistes d'amélioration des outils, comme évoqué dans la partie 2, une intégration de toutes les données sous une base de données géographique comme PostgreSQL permettrait de rendre les outils plus résilients, avec une mise à jour automatique et permettrait de faciliter les analyses. L'URD met bien en place ces outils mais cela n'est pas encore automatisé pour tous les types de données et est fait plus dans un but d'archiver les données pour certains datasets. Donc en général, les données existent sur la base de données URD mais si l'on veut travailler avec les dernières données disponibles, il faut forcément passer par un format de fichier Excel ou .csv.

Pour ce qui est de l'outil de suivi en temps réel des données de campagne QOS, une piste d'amélioration serait d'améliorer la reconnaissance du type de données par le logiciel. En l'absence de données standardisées, l'utilisateur de la plateforme doit lui-même rentrer dans l'application le nom des colonnes dont il a besoin pour faire fonctionner l'application (cf. figure 2.5).

Une piste d'amélioration pour les outils web serait de renforcer la fluidité des outils. D'un point de vue du chargement tout d'abord, même si les données sont filtrées et sectionnées au maximum, le temps de chargement des pages web n'est pas instantané au moment de l'affichage et dure en moyenne entre 5 et 10 secondes quand on choisit d'afficher l'ensemble des données pour les deux outils. Une solution que j'ai explorée notamment pour l'upload de fichiers serait de créer plusieurs threads qui permettraient au backend de continuer de travailler en affichant la page web au fur et à mesure. De plus, toujours sur le sujet de la fluidité, le grand nombre de données entraîne une navigation saccadée sur les cartes interactive WEB faites à l'aide de la librairie Javascript Leaflet.

4.3 Fin du stage, suite et utilisations futures des projets

Durant les deux mois de stage suivant la rédaction de ce rapport, je devrai tout d'abord terminer l'implémentation des projets signifiés dans le diagramme de GANTT (Annexe A) en en faisant le maximum avant la date limite du 10 novembre. Je devrai aussi m'assurer de la bonne intégration de mes outils et de mes études dans l'arsenal d'indicateurs de l'ARCEP afin qu'ils puissent être utilisés à leur maximum. Des plateformes de partage de données et d'indicateurs telle que Redash sont en place au sein de l'ARCEP et je devrai très certainement y intégrer certains des résultats de mes études. En ce qui concerne l'entretien des outils WEB ou la réutilisation potentielle de mes codes pour les analyses, je devrai fournir une documentation programmeur à l'Unité de Régulation par la Donnée (URD) qui jugera de l'utilité d'entretenir et de garder tout ou partie de mon travail. Les codes que j'ai pu écrire sont déjà commentés dans cette optique et disponibles sur le logiciel Gitlab géré par l'URD.

Pour chaque outil que je mettrai en place dans le futur, comme je l'ai fait par exemple pour l'outil de suivi de QOS et de suivi du dispositif DCC, il faudra former les agents susceptibles de les utiliser afin que leur utilisation reste et que les fonctionnalités puissent être enrichies.

Enfin, ce rapport de stage sera mis à disposition de l'ARCEP et plus spécifiquement de l'UCIM au cas où un autre stagiaire ou un agent serait amené à travailler sur les mêmes sujets que moi.

Conclusion

Afin de conclure ce rapport, nous pouvons rappeler les objectifs initiaux du stage. Les objectifs étaient dans un premier temps d'apporter un regard extérieur sur les jeux de données disponibles au sein de la Direction Mobile et Innovation (DMI). Ensuite, via ce regard extérieur, l'objectif était de produire des analyses et des indicateurs s'intégrant dans l'arsenal d'outils présents à l'ARCEP afin d'avoir une évaluation la plus juste possible des réalités du marché de la téléphonie mobile en France.

A travers les différents outils et les analyses que j'ai pu mettre en place, l'unité couverture et investissements mobiles a maintenant en sa possession une série de nouveaux éléments qu'elle peut utiliser au quotidien pour alléger sa charge de travail, parfaire ses prises de décisions ou encore s'informer sur les travaux des autres agents en lien avec ses activités.

Pour donner une conclusion chiffrée à la rédaction de ce rapport, j'ai mis en place 3 applications web de gestion ou de visualisation, j'ai aussi réalisé 2 enquêtes avec en conclusion des résultats illustrés et chiffrés et 4 projets annexes de types divers de quelques jours à la demande de l'équipe ou de mon maître de stage.

Tous les projets envisagés en début de stage n'ont pas été complétés mais je retire de cela une leçon sur le travail avec des dates limites resserrées et l'enchaînement des projets en collaboration avec des équipes qui ne peuvent pas prendre de retard en attendant vos travaux.

D'un point de vue personnel, ce stage est une première expérience réussie au sein d'une autorité de régulation après un passage dans celui de la recherche. J'ai pu développer mon réseau professionnel et découvrir le secteur très intéressant des télécoms et des autorités administratives indépendantes. Ce stage était aussi l'occasion d'appliquer les connaissances acquises lors de mon master TSI de fin de cycle ingénieur tout en continuant de développer ces mêmes compétences et d'en acquérir de nouvelles.

Le fait d'avoir fait partie d'une unité métier telle que l'UCIM, en étant le seul géomaticien m'a permis de communiquer sur le savoir-faire acquis au sein de ma formation à l'ENSG, de proposer de nouvelles façons de faire et de voir les données géographiques et leur utilisation. Cela m'a aussi permis de travailler sur la transmission des résultats de mes travaux à destination de personnes non expertes en géomatique.

Je retire donc beaucoup de positif de cette expérience. Je suis particulièrement fier d'avoir pu clôturer ma formation d'ingénieur en géomatique avec cette expérience, où j'ai pu mettre en pratique mes compétences, découvrir un nouveau domaine, et construire des outils qui, je l'espère, aideront l'Arcep à mener son action pour assurer une couverture mobile satisfaisante pour l'ensemble des citoyens.

Webographie / Bibliographie

- [1] ARCEP. *Le manifeste de l'ARCEP*. <https://www.arcep.fr/larcep/le-manifeste-de-larcep.html>. 2022.
- [2] ARCEP. *Livret d'accueil*. Disponible sur l'intranet de l'ARCEP. 2021.
- [3] ARCEP. *Plaquette institutionnelle de l'ARCEP*. Disponible sur l'intranet de l'ARCEP. 2020.
- [4] ARCEP. *Site de partage d'informations en open-source mon réseau mobile*. <https://monreseau-mobile.arcep.fr/>. Consulté en 2022.
- [5] ARCEP. *Site j'alerte l'ARCEP*. <https://jalerte.arcep.fr/>.
- [6] ARCEP. *Tableau de bord du New deal*. <https://www.arcep.fr/cartes-et-donnees/tableau-de-bord-du-new-deal-mobile.html>. Consulté en 2022.
- [7] Multiple AUTHORS. *Site de la librairie python scikit learn*. <https://scikit-learn.org/stable/>. Consulté en 2022.
- [8] Multiple AUTHORS. *Site du système notebook jupyter*. <https://jupyter.org/>. Consulté en 2022.
- [9] Pitney BOWES. *Référence du site Pitney Bowes, fournisseur de données de population*. <https://www.pitneybowes.com/fr>. Consulté en 2022.
- [10] Conseil d'état FRANÇAIS. *La régulation*. <https://www.conseil-etat.fr/publications-colloques/discours-et-interventions/la-regulation>. Consulté en 2022.
- [11] R. E. Schapire S. J. PHILLIPS R. P. Anderson. *Maximum entropy modeling of species geographic distributions*. <http://rob.schapire.net/papers/ecolmod.pdf>. 2006.

Table des figures

1.1	Logo de l'ARCEP	6
2.1	Image de promotion d'un prestataire en enquêtes de qualité de services mobiles, [https://www.connect-testlab.com/the-netherlands-2020-methodology]	11
2.2	Liste des colonnes d'informations retenues pour l'application post pré-traitements et enregistrées dans la BDD serveur pour qualifier les mesures voix/sms et data	13
2.3	Liste des colonnes d'informations retenues pour l'application post pré-traitements et enregistrées dans la BDD serveur pour évaluer les mesures voix/sms	13
2.4	Liste des colonnes d'informations retenues pour l'application post pré-traitements et enregistrées dans la BDD serveur pour évaluer les mesures data	13
2.5	Liste des noms de paramètres et des différentes options que l'utilisateur doit rentrer pour valider l'upload (disponible en taille complète en Annexe C)	14
2.6	Résumé des interactions entre le client et le backend au moment de l'upload d'un nouveau fichier QOS	14
2.7	Organisation des fichiers post traitement dans la base de données serveur	14
2.8	Capture d'écran de l'application mise en place pour montrer les POIs, les sites et les distances associées	17
2.9	Problèmes d'association entre les numéros de dossier des POIs et des sites associés	18
3.1	Tableau permettant de résumer les données utilisées dans l'analyse des composantes des sites mobiles	21
3.2	Schéma décrivant les étapes de l'algorithme mis en place pour la partie 3.1.2 par site	22
3.3	Découpage du territoire en zones d'aires égales	23
3.4	Schéma décrivant les étapes de l'algorithme mis en place pour la partie 3.2 pour des zones données	23
3.5	Matrice de confusion de l'algorithme d'identification de sites. [LogisticRegression(penalty = 'l2', C = 100, solver = 'lbfgs', max iter = 15000)]	24
3.6	Classification report de l'algorithme d'identification de sites. [LogisticRegression(penalty = 'l2', C = 100, solver = 'lbfgs', max iter = 15000)]	24
3.7	Répartition des bandes de fréquences mobiles et technologies en France métropolitaine en 2022	26
3.8	Carte de répartition des technologies pour SFR en bande 2100 au troisième trimestre 2020	28
3.9	Diagramme du compte de sites en fonction de la densité de population et des technologies au second trimestre 2022 pour l'opérateur Orange et la bande de fréquence 2,6 GHz	28
3.10	Heatmap de la répartition des sites pour l'opérateur Orange France au second trimestre 2022	29
A.1	Organigramme de l'ARCEP du 01 octobre 2021	41
A.2	Diagramme de GANTT préliminaire du 26 juin 2022	42
B.1	Architecture type d'une application via le framework Python Flask	43

C.1	Screenshot descriptif de l'application de suivi des campagnes QOS, page d'accueil de l'application	45
C.2	Screenshot descriptif de l'application de suivi des campagnes QOS, page d'upload de nouveau fichier et choix des paramètres par l'utilisateur	45
C.3	Screenshot descriptif de l'application de suivi des campagnes QOS, page de choix des données présentes dans l'application	46
C.4	Screenshot descriptif de l'application de suivi des campagnes QOS, page de présentation des résultats, tableaux des statistiques en fonction des zones de densité	46
C.5	Screenshot descriptif de l'application de suivi des campagnes QOS, page de présentation des résultats, diagrammes des statistiques en fonction des zones de densité . .	47
C.6	Screenshot descriptif de l'application de suivi des campagnes QOS, page de présentation des résultats, carte des zones déjà couvertes au moment de l'upload du fichier	47
D.1	Screenshot descriptif de l'application de suivi des déploiements DCC (1/3)	48
D.2	Screenshot descriptif de l'application de suivi des déploiements DCC (2/3)	48
D.3	Screenshot descriptif de l'application de suivi des déploiements DCC (3/3)	49
E.1	Etat des lieux des la couverture en métropole et Outre-mer en septembre 2020	51
F.1	Exemple d'évolution des déploiements de sites sous forme de cartes pour Bouygues Télécoms entre 2020 et 2022 pour les bandes 800 et 900	53
F.2	Exemple d'évolution des déploiements de sites sous forme de diagramme pour SFR entre 2020 et 2022 pour la bande 2600 MGHZ	54
F.3	Exemple d'évolution des déploiements de sites sous forme de heatmap pour Free mobile entre 2020 et 2022 toutes bandes confondues	55

Annexes

A	Contexte	41
B	Architecture des applications	43
C	Application QOS	45
D	Application DCC	48
E	Etude de déploiements	51
F	Data visualisation sur la largeur de bande	53

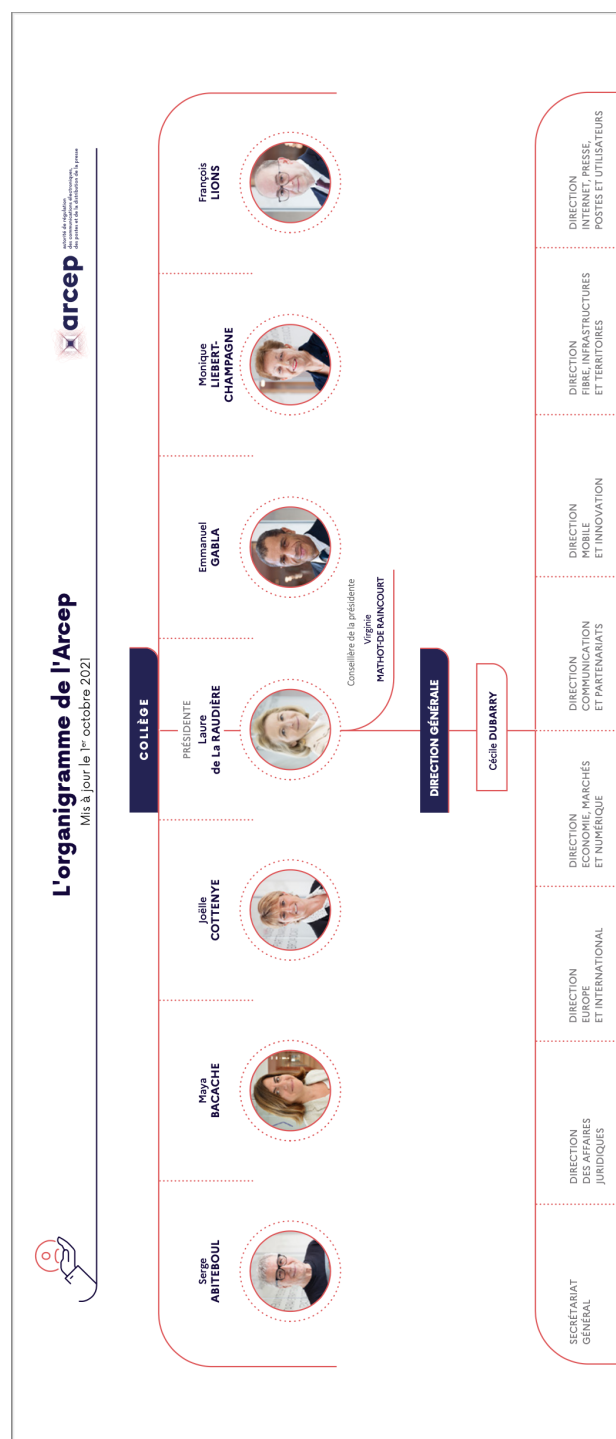


FIGURE A.1 – Organigramme de l'ARCEP du 01 octobre 2021

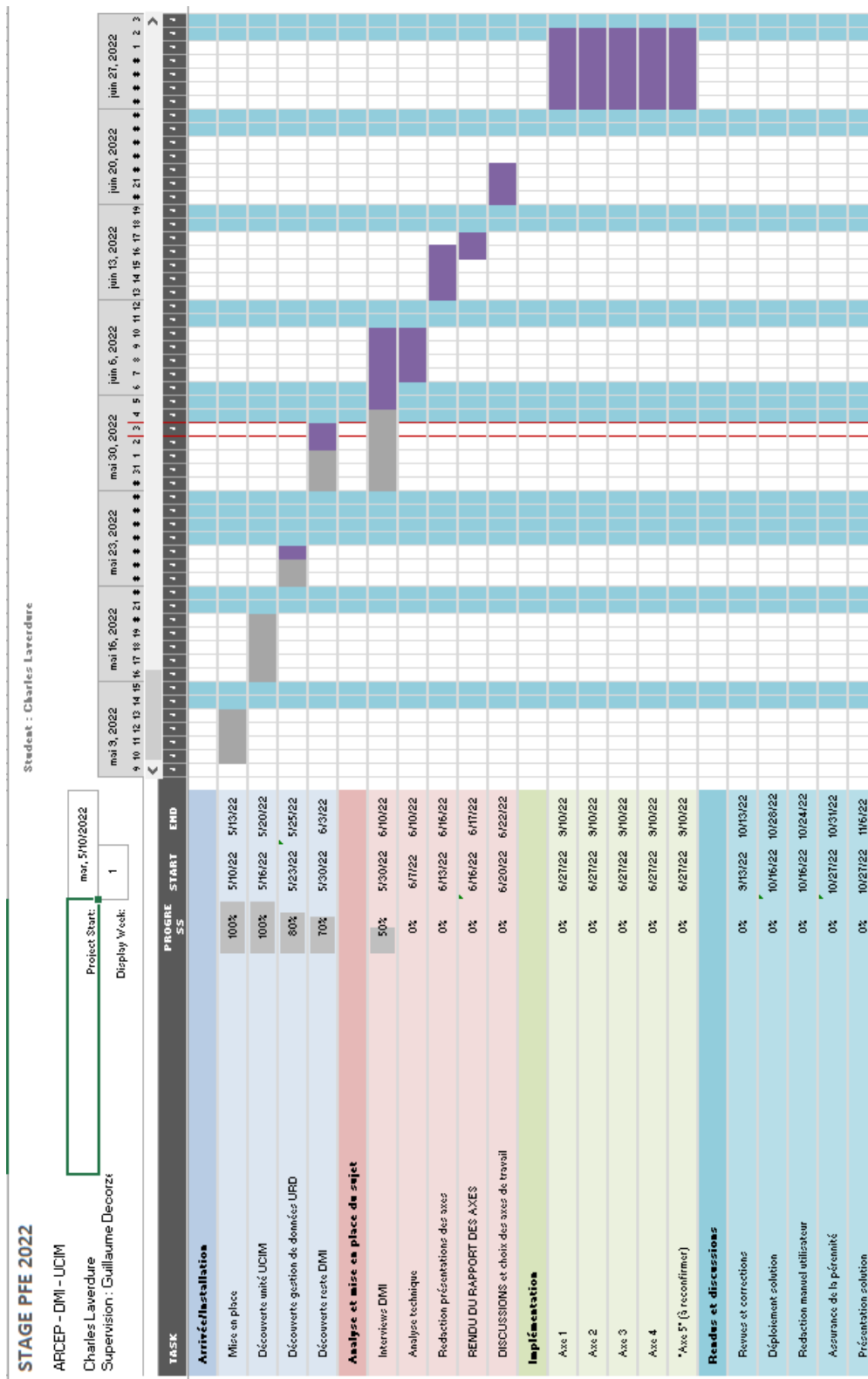


FIGURE A.2 – Diagramme de GANTT préliminaire du 26 juin 2022

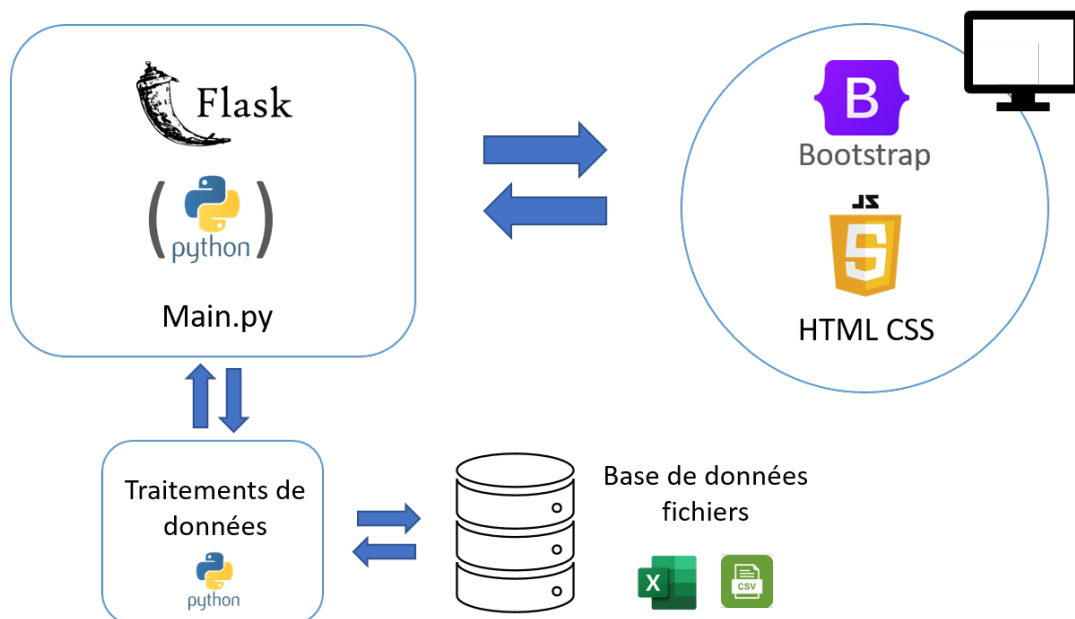


FIGURE B.1 – Architecture type d'une application via le framework Python Flask

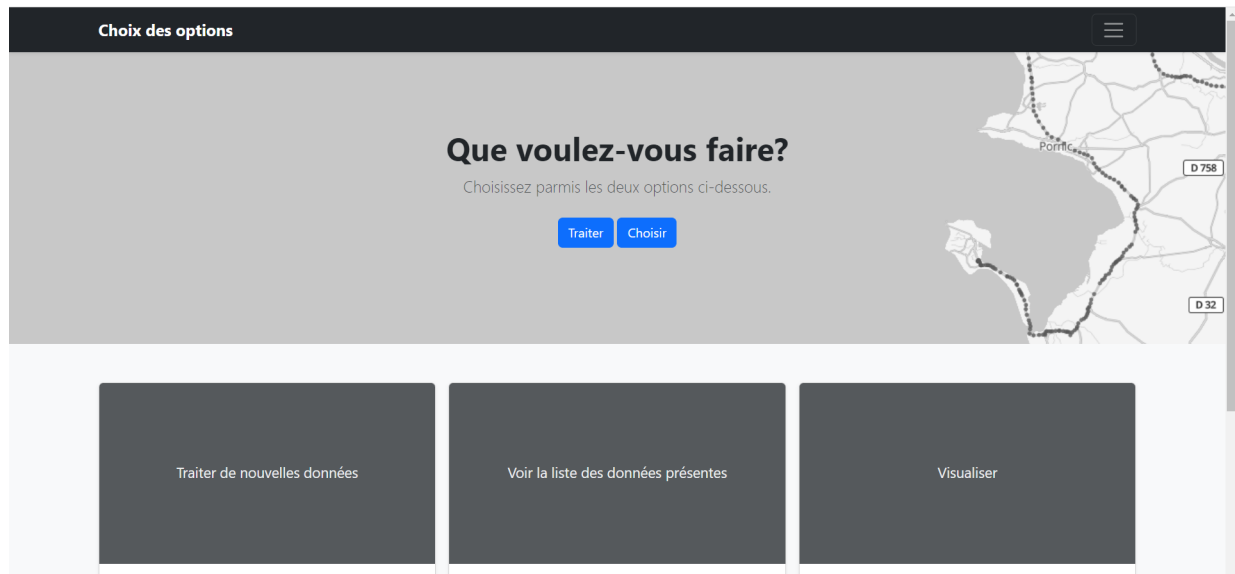


FIGURE C.1 – Screenshot descriptif de l'application de suivi des campagnes QOS, page d'accueil de l'application

The screenshot shows the upload page of the QOS application. At the top, there are two radio buttons: "Data" (unselected) and "Voix/SMS" (selected). Below this, there is a dropdown menu labeled "Renseignez l'année de vos données". Underneath, a text box contains the instruction "Si votre fichier est un .xlsx, renseignez le nom de la spreadsheet". Below this, there is a text input field labeled "Nom spreadsheet" with the value "BDD". Further down, the text "Renseignez les champs suivants pour la voix et sms" is displayed. Below this, there are several text input fields, each with a label and a value: "Numero" (No), "Conditions" (Usage), "Couple" (Couple), "Type test" (Service), "Result" (Result), "Latitude" (Lat_debut), "Longitude" (Lon_debut), and "MOS Moyen" (MOS Moyen). At the bottom, there is a large blue button labeled "Valider l'upload".

FIGURE C.2 – Screenshot descriptif de l'application de suivi des campagnes QOS, page d'upload de nouveau fichier et choix des paramètres par l'utilisateur

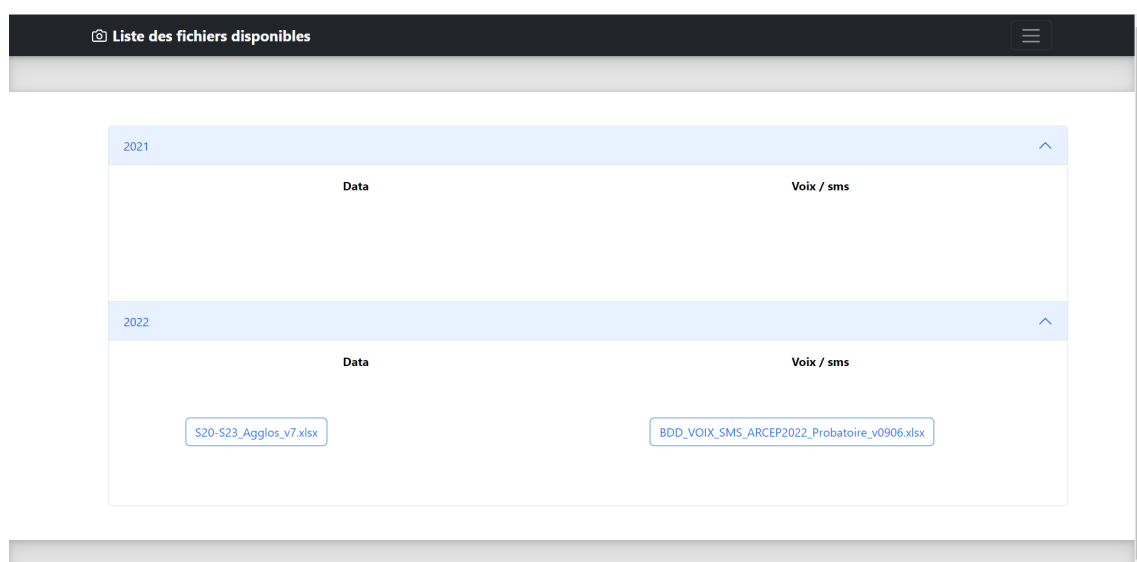


FIGURE C.3 – Screenshot descriptif de l'application de suivi des campagnes QOS, page de choix des données présentes dans l'application

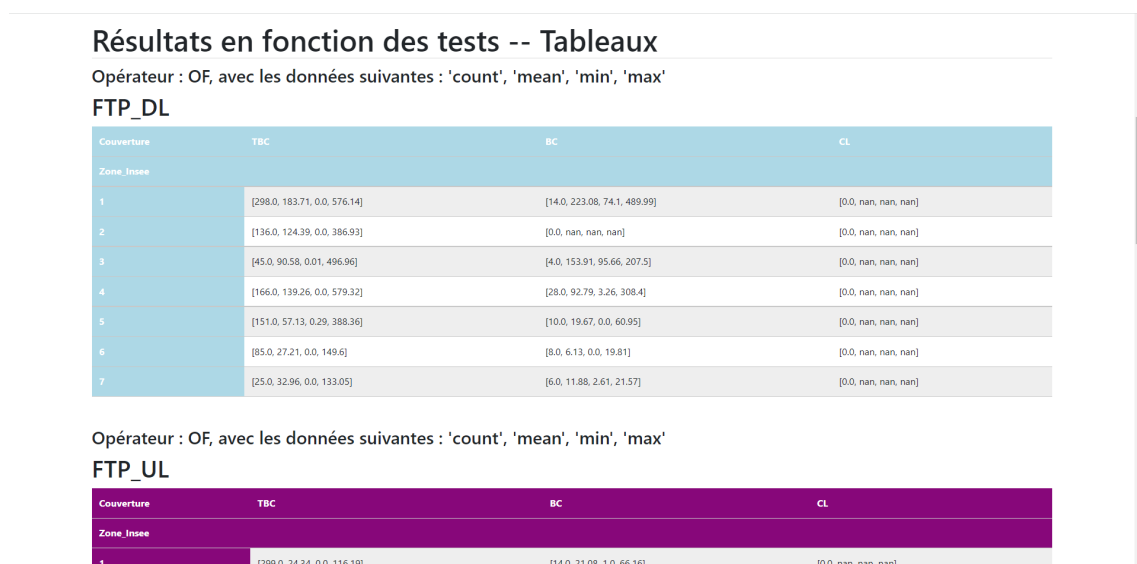
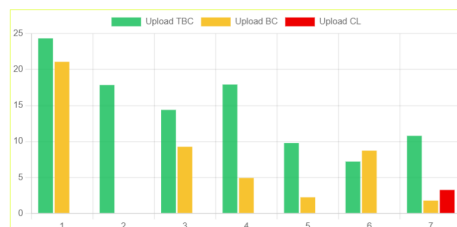


FIGURE C.4 – Screenshot descriptif de l'application de suivi des campagnes QOS, page de présentation des résultats, tableaux des statistiques en fonction des zones de densité

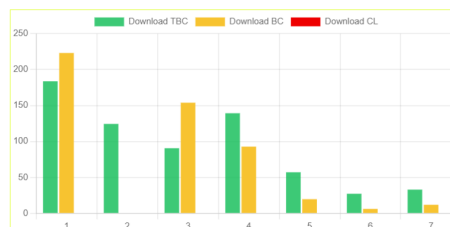
Plots des résultats

PLOTS

On plot les valeurs de mean pour les tests d'uploads



on plot les valeurs de mean pour les tests de download



Carte des relevés

FIGURE C.5 – Screenshot descriptif de l'application de suivi des campagnes QOS, page de présentation des résultats, diagrammes des statistiques en fonction des zones de densité

Carte des relevés

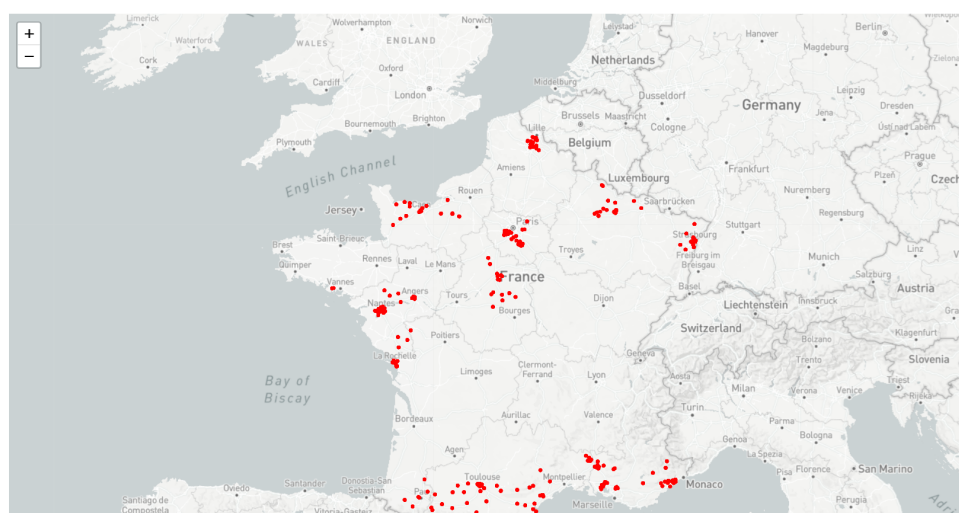


FIGURE C.6 – Screenshot descriptif de l'application de suivi des campagnes QOS, page de présentation des résultats, carte des zones déjà couvertes au moment de l'upload du fichier

APPLICATION DCC

ANNEXE
D

Visualiser l'évolution du projet DCC

Choisissez les dates de vos données de POI

Trimestre 2021-07-02

Choisissez les dates de vos données de sites

Date 2021-07-30

☐ Voir les POIs proposés avant la mise en place des arrêtés?

Afficher

FIGURE D.1 – Screenshot descriptif de l'application de suivi des déploiements DCC (1/3)

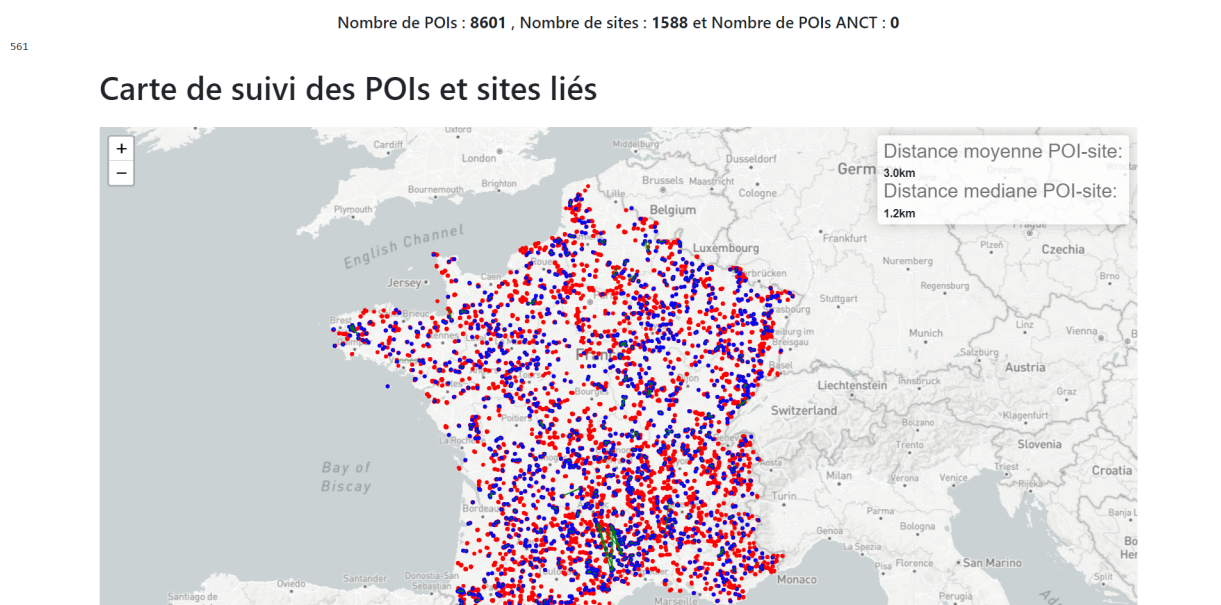


FIGURE D.2 – Screenshot descriptif de l'application de suivi des déploiements DCC (2/3)

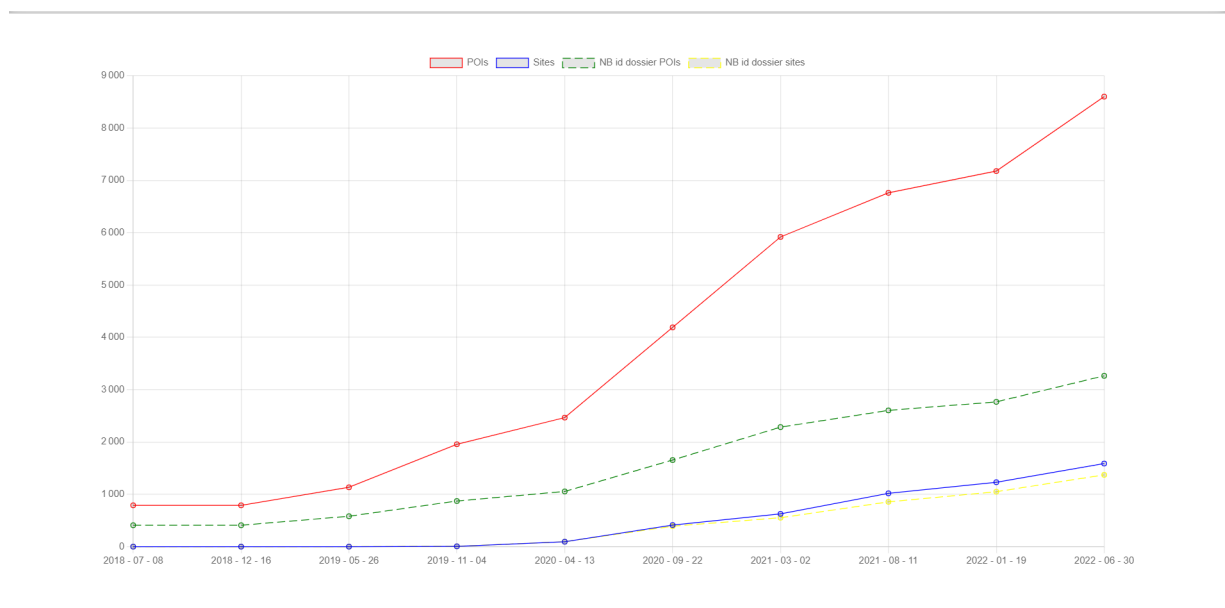
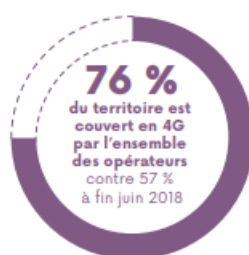
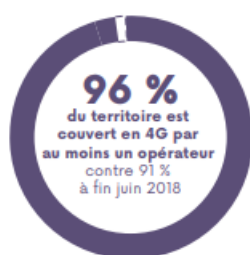


FIGURE D.3 – Screenshot descriptif de l'application de suivi des déploiements DCC (3/3)

Couverture en France métropolitaine à fin septembre 2020



**Entre 600
et 800 sites**
sont identifiés chaque année
par les collectivités. Ils doivent
être construits dans les 2 ans
qui suivent leur identification³.

63 % des sites zones blanches
– centres-bourgs sont équipés
en 4G à fin septembre 2020,
contre aucun à fin juin 2018

Ces taux de couverture sont issus
des cartes de couverture 4G théo-
riques des opérateurs mobiles, qui
représentent les zones où un usager
devrait pouvoir échanger des données
en 4G à l'extérieur des bâtiments
dans la plupart des cas. Ces cartes
sont réalisées à partir de simulations
numériques. Aussi précises soient-elles,
elles donnent une information sur l'en-
semble du territoire et représentent,
par nature, des visions simplifiées de la
réalité. Néanmoins, ces cartes doivent
respecter un niveau de fiabilité minimal
établi par l'Arcep à 98 %.

Couverture en Outre-mer à fin septembre 2020

Guadeloupe



Martinique



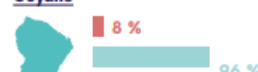
La Réunion



Mayotte



Guyane



Saint-Barthélemy



Saint-Martin



■ TAUX DE SURFACE COUVERTE
EN 4G PAR AU MOINS 1 OPÉRATEUR
■ TAUX DE POPULATION COUVERTE
EN 4G PAR AU MOINS 1 OPÉRATEUR

1. <https://www.arcep.fr/cartes-et-donnees/tableau-de-bord-du-new-deal-mobile.html>

2. monreseaumobile.fr

3. Ou dans un délai réduit si l'emplacement est mis à disposition par la collectivité.

FIGURE E.1 – Etat des lieux de la couverture en métropole et Outre-mer en septembre 2020

DATA VISUALISATION SUR LA LARGEUR DE BANDE

ANNEXE
F

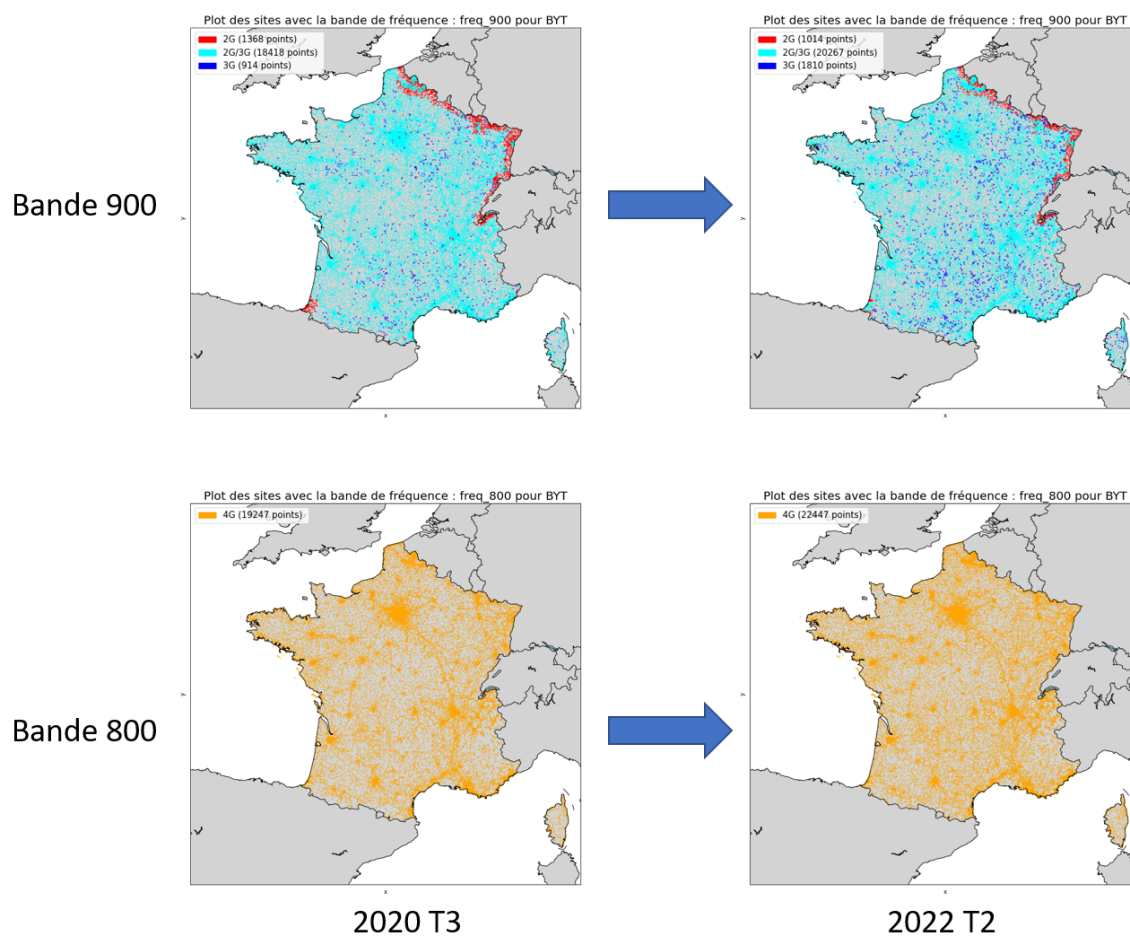


FIGURE F.1 – Exemple d'évolution des déploiements de sites sous forme de cartes pour Bouygues Télécoms entre 2020 et 2022 pour les bandes 800 et 900

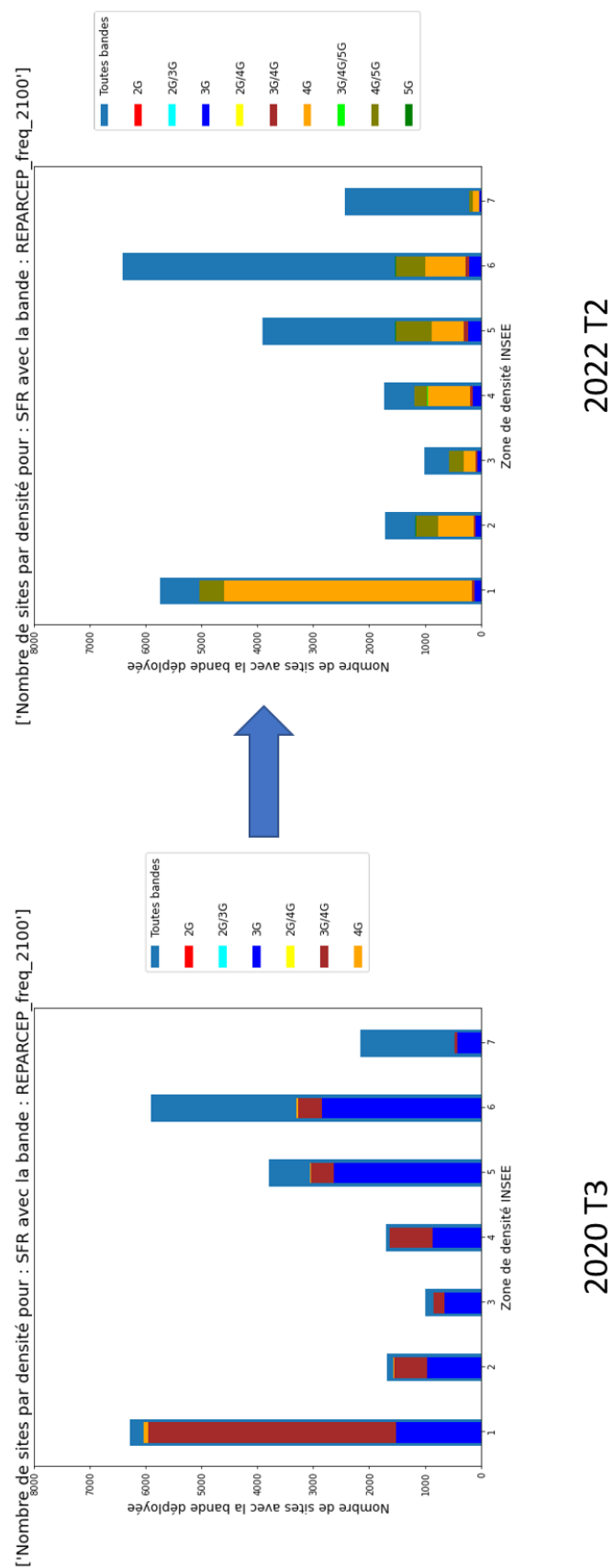


FIGURE F.2 – Exemple d'évolution des déploiements de sites sous forme de diagramme pour SFR entre 2020 et 2022 pour la bande 2600 MGHz

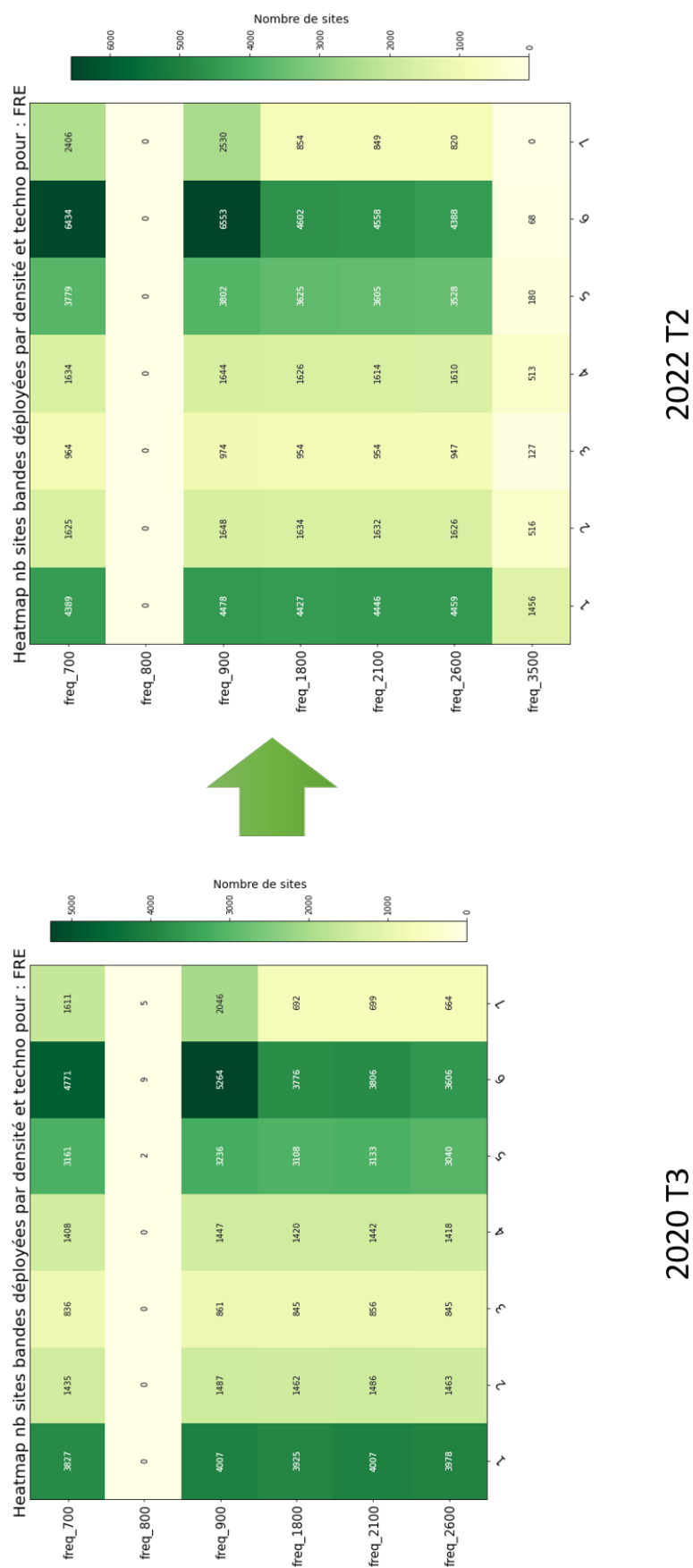


FIGURE F.3 – Exemple d'évolution des déploiements de sites sous forme de heatmap pour Free mobile entre 2020 et 2022 toutes bandes confondues