



HAL
open science

Valorisation des données dans le cadre des “ données ouvertes ” (open data) des entreprises : le cas du portail data.sncf de la SNCF

Stella Ngandjou Diki

► **To cite this version:**

Stella Ngandjou Diki. Valorisation des données dans le cadre des “ données ouvertes ” (open data) des entreprises : le cas du portail data.sncf de la SNCF. Sciences de l’information et de la communication. 2017. dumas-01761339

HAL Id: dumas-01761339

<https://dumas.ccsd.cnrs.fr/dumas-01761339>

Submitted on 9 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Valorisation des données dans le cadre des
« données ouvertes » (open data) des entreprises:
Le cas du portail data.sncf de la SNCF.**

Par

NGANDJOU DIKI Stella

*Mémoire présenté au Département SID
en vue de l'obtention du Master 2 en Gestion de l'Information et du Document en Entreprise*

UNIVERSITE DE LILLE 3

Lille, France. 2016-2017

Résumé

A l'heure où le numérique et ses usages transforment et redéfinissent à la fois notre économie et nos espaces, les collectivités, administrations publiques et même les entreprises privées aujourd'hui ont tendance à s'approprier des technologies spécifiques en occurrence l'open data pour valoriser les données issues de leurs systèmes d'information. Ils s'engagent de ce fait sur une politique de mise à disposition et de réutilisation libre et facile des données. Cette propension à mettre en avant les données fait radicalement évoluer les relations entre les producteurs et les usagers même si les enjeux liés à cette ouverture et aux usages qui peuvent en découler ne sont pas encore totalement perçus. Faire pourtant usage de cette matière brute et exploitable pour former un écosystème innovant est la base dans la création de valeur de l'Open Data. Nous espérons que ce travail aura contribué à la mise en cohérence de la politique open data au sein de l'entreprise.

Mots-clés: Open data, données ouvertes, réutilisations, valorisation des données, ouverture des données des entreprises, patrimoine immatériel

Abstract :

At a time when digital technology and its uses are transforming and redefining both our economy and our spaces, communities, public administrations and even private companies today tend to appropriate specific technologies like Open Data to value data from their information systems. They therefore commit themselves to a policy of free and easy data re-use. This propensity to put forward the data radically changes the relations between producers and users. During our professional career at the SNCF, it was quite clear that some in-house services did not fully perceive the issues related to the opening of the data and the uses that can result from it. However making use of this raw and exploitable material to form an innovative ecosystem is the basis in the value creation of Open Data. We hope that this work will have contributed to the coherence of the open data policy within the company.

Keywords: Open data, reuses, valuation of data, company open data, intangible heritage

Remerciements

Je ne saurais amorcer la rédaction de ce mémoire sans toutefois exprimer ma profonde gratitude à tous ceux qui ont œuvré de quelque manière que ce soit à l'aboutissement de ce présent travail.

Je tiens à cet effet à exprimer tout particulièrement ma reconnaissance aux responsables de la formation, du master GIDE et à toute l'équipe enseignante pour avoir, des mains de maître su nous conduire à la réussite et l'intégration des enseignements de cette deuxième année de master en gestion de l'information et du document.

Mes remerciements vont également à l'endroit de Madame Laurence FAVIER, et Monsieur Guillaume LEBORGNE, respectivement tutrice académique et tuteur professionnel, ainsi qu'à tout le personnel de la Direction de la Communication et en particulier celui du service Transparence et Open Data de la SNCF pour leur accueil chaleureux, leur disponibilité permanente tout au long de ce parcours professionnalisant.

Je remercie mes camarades de promotion pour leur esprit de solidarité manifeste tout au long de cette année académique.

Enfin, à mes parents bien aimés Monsieur et Madame NGANDJOU pour tous les efforts inestimables qu'ils n'ont jamais cessé de consentir pour mon éducation et mon bien-être.

Table des matières

Résumé	ii
Remerciements	iii
Table des matières	iv
Liste des tableaux	vi
Liste des figures	vii
INTRODUCTION GÉNÉRALE.....	1

PREMIERE PARTIE: RAPPORT DE STAGE

Introduction	4
Chapitre 1 : Etude institutionnelle et présentation de la mission	6
I- Présentation de la SNCF	6
1.1. Fiche synoptique de présentation	6
1.2. Historique et évolution	6
1.3. Activités et fonctionnement	7
II- Le service Transparence & Open data	8
2.1. La démarche d'ouverture des données de transports	9
2.2. La mise en oeuvre de l'open data dans le contexte SNCF	10
2.3. Organisation et Missions	11
III- Mission et objectifs	11
Chapitre 2 : Mise à jour du catalogue data.sncf, optimisation des graphiques d'analyse et reporting de l'activité des domaines.....	13
I- Bien-fondé de la réalisation.....	14
II- Méthodologie	15
2.1. Analyse de l'existant et définition des besoins	15
2.2. Benchmarking	17
III- Réalisations et préconisations	18
3.1. Mise à jour du catalogue, méthodes et alertes.....	18

3.2. Moissonnage des données externes et interopérabilité	23
3.3. Génération des graphiques d'analyse statistique, et décryptage des données.....	30
3.4. Activité de reporting: DÉFINIR LES MÉTRIQUES DE SUIVI DES USAGES et création des dashboard	32
Chapitre 3 : Discussion du travail réalisé.....	35
Conclusion.....	37

DEUXIEME PARTIE: RAPPORT DE RECHERCHE - L'open data des entreprises pour quels usages?

Introduction	40
I- Définition de la problématique et présentation de son positionnement dans la littérature	42
1.1. Panorama de la littérature scientifique	42
1.2. Panorama de la littérature professionnelle	46
1.3. Observations InSitu	48
II- Pistes à étudier pour une recherche approfondie du sujet + méthodologie d'étude	48
Conclusion.....	50
Bibliographie.....	51
Annexes.....	56
Annexe1 Organigramme	57
Annexe2 Cartographie OpenStreetMap des gares SNCF Transilien en Ile-de-France.....	59
Annexe3 Charte Open data SNCF	60
Annexe4 Tableau de correspondances-catégories Gallica	62
Annexe5 Extrait du code de conception des tableaux de bord et graphiques	64

Liste des tableaux

Tableau 1 – Synthèse d’audit du catalogue de données SNCF	17
Tableau 2 – Structures ayant servi au benchmarking.....	18
Tableau 3 – Requêtes utilisées pour l’exploration des jeux de données	20
Tableau 4 – Plan d’action des mises à jour du catalogue.....	21
Tableau 5 – Extrait du référentiel de mise à jour	22
Tableau 6 – Analyse des risques liés à la fédération des jeux de données KEOLIS	26

Liste des figures

Figure 1 – Segmentation des EPIC SNCF	7
Figure 2 – Procédure d’export du catalogue de données	19
Figure 3 – Cartographie des jeux de données KEOLIS	25
Figure 4 – Résultats du test de fédération des données KEOLIS.....	27
Figure 5 – Typologie Gallica	28
Figure 6 – Représentation graphique du jeu de données « Baromètre d’opinions SNCF »....	31
Figure 7 – Exemple de notice élaborée à partir des archives mis en open data	31
Figure 8 – Analyse des données de régularité TER	32
Figure 9 – Extrait du tableau de bord des données Réseau.....	33
Figure 10 – Tableau de bord du catalogue data	34

INTRODUCTION GÉNÉRALE

La démarche d'ouverture des données commence réellement à prendre forme en France après que celle-ci ait intégré le partenariat pour le gouvernement ouvert. Elle désigne l'effort que font les institutions, notamment gouvernementales, qui partagent les données dont elles disposent. Ce partage doit être gratuit, dans des formats ouverts, et permettre la réutilisation des données. La politique du Gouvernement s'appuie notamment sur le droit d'accès aux documents administratifs (loi CADA 1978), qui considère que les données produites ou détenues par les administrations, dans le cadre de leurs missions de service public, doivent être mises à disposition du public. Cela ne concerne ni les informations personnelles, ni celles touchant à la sécurité nationale, et celles couvertes par les différents secrets légaux non plus.

L'open data est donc un outil numérique ayant comme objectifs :

- D'améliorer le fonctionnement démocratique, non seulement par la transparence mais par la concertation et l'ouverture à de nouveaux points de vue ;
- D'améliorer l'efficacité de l'action publique ;
- De proposer de nouvelles ressources pour l'innovation économique et sociale : les données partagées trouvent des réutilisateurs qui les intègrent dans de nouveaux services à forte valeur ajoutée économique ou sociale.

L'Open Data est souvent abordée sous l'angle secteur public/collectivités. Mais un axe futur de développement majeur passe également par le secteur privé et le monde des entreprises. Cette nouvelle communication numérique permet non seulement la valorisation des données et le développement en outre de nouveaux usages mais aussi de nouveaux outils de mesure d'efficacité. Il s'agit en quelque sorte d'une communication ouverte qui vient redéfinir les termes des stratégies engagées au sein des entreprises.

Dans le cas des entreprises, c'est un concept tout à fait exhaustif. Cela implique de mettre à disposition autant d'informations que possible à l'externe tout en protégeant les informations à caractère privé et tout ce qui a trait au secret d'affaires, en créant des solutions pour améliorer le service et en communiquant au sein de l'organisation l'engagement d'ouvrir les données. Ce faisant, une culture de données peut être créée qui devient plus une pratique.

L'open data connaît aussi des évolutions au niveau juridique mais au-delà, elle tend à devenir pour les entreprises un axe stratégique de développement enveloppant toutes les directions et services. Les buts sont pluriels : tout d'abord il s'agit d'un devoir de redevabilité de la part d'une entreprise dont la mission est publique à l'égard des citoyens, c'est aussi une façon de présenter la réalité auprès des usagers et des médias, et enfin cette démarche constitue un accélérateur d'innovations.

Comme dans toute organisation, convaincre les équipes d'ouvrir leurs données à la fois pour les autres membres de l'entreprise et pour le grand public peut être plutôt difficile. L'initiative peut motiver ou enthousiasmer certains et d'autres pas. Toutefois, elle se veut être une politique de valorisation des données. Une valorisation qui passe par plusieurs étapes pour rendre les données publiques utilisables et réutilisables. D'une part l'institution productrice doit définir précisément quelles données elle doit diffuser en ligne et de quelle manière, par un travail de collecte et d'analyse. D'autre part, pour faciliter et encourager les réutilisations, les institutions productrices doivent s'assurer de fournir une donnée qualifiée, à jour et documentée. Enfin, les données valorisées sont désormais au cœur de nouveaux usages, une fois mises à disposition, elles peuvent être réutilisées en interne comme outil de circulation et d'échange des flux, et en externe par la communauté des développeurs et autres entrepreneurs, qui peuvent inventer de nouveaux usages.

La SNCF se trouve ainsi engagée depuis quelques années déjà dans un processus d'ouverture des données visant essentiellement l'innovation, elle présente un intérêt pour le développement de l'entreprise en tant que transparente et moderne, et facilite les prises de décisions.

Ce mémoire a pour objet d'étude la valorisation des données dans le cadre des données ouvertes des entreprises. Dans sa première partie, il fera état des missions effectuées durant mon parcours de professionnalisation au sein de la SNCF. La deuxième partie vise à étudier les contours théoriques de la question des usages des données mises à disposition par les entreprises.

PREMIÈRE PARTIE: RAPPORT DE STAGE

Introduction

Dans le contexte de révolution numérique auquel fait face la société moderne, le mouvement open data, consistant à communiquer les données numériques des administrations connaît une avancée remarquable. L'open data c'est repenser l'action publique en intégrant les enjeux de transparence, d'innovation économique, et d'échanges d'informations. Les besoins de productivité et d'innovation dans les entreprises amènent les décideurs à s'intéresser davantage aux opportunités que cache ce nouveau marché de l'information. Fournir des services de qualité reste l'une de leurs plus grandes ambitions, disposant de nombreuses données éparses et impliquant la plupart du temps les utilisateurs, la nécessité de les partager se fait vite ressentir. De ce fait, l'écosystème se met en place et intègre dans une relation tripartite la donnée, le producteur et l'utilisateur. Quelle politique pour le partage des données? À quelles fins vont-elles servir ? Comment évaluer la plus-value?

L'émergence du concept de l'open data ou de l'ouverture des données témoigne du besoin des secteurs publics et aussi privés de disposer de manière optimale d'informations fiables, excepté quelques-unes et qui peuvent par ailleurs servir à des réutilisations qui profitent à tous.

L'ouverture des données, encore principalement l'apanage des collectivités, des organismes gouvernementaux (villes, IGN, INSEE,...) commence à gagner du terrain dans certaines entreprises publiques ou privées. Les exemples sont nombreux déjà : la SNCF qui produit des données sur la fréquence des trains, horaires et incidents (www.data.sncf.com). Le groupe Suez Environnement qui produit des données sur la gestion de l'eau, Sytadin qui génère les données du trafic routier en IDF, etc.

La SNCF, l'une des entreprises les plus importantes en France occupe un rôle essentiel. Amenée à gérer le transport de voyageurs et de marchandises, elle assure des missions de service public et est de ce fait engagée depuis peu dans une démarche de transparence et d'ouverture de ses données disponibles. Le transfert de ces données nécessite de les organiser, les contextualiser et les enrichir pour faciliter leur interprétation et réutilisations, le patrimoine informationnel que constitue ces données de transport font fi de l'activité du groupe ferroviaire et rendent compte de certaines opérations et enquêtes.


L'Open Data SNCF s'est développé à très grande vitesse. Depuis 2014, les jeux de données ouverts sont passés de 18 à 202. L'entreprise se dote par ailleurs d'une politique ambitieuse en matière d'interface de programmation applicative (API) pour diffuser de manière plus efficace ses flux de données et accompagner l'innovation ouverte. L'idée est de poursuivre et d'asseoir cette dynamique en animant les communautés open data, en interne et en externe.

S'agissant pour la plupart des données générées par les SI, le volume important et la complexité des formats d'échange ont engendré le souhait d'avoir un catalogue à jour, suivi, et d'apporter un caractère intelligible à la donnée. Communiquées très souvent de manière brute, elles ont plus de sens que ce qui paraît. Cette première partie rend fidèlement compte des activités menées durant notre stage professionnalisant au sein de l'organisme sus visé. Elle comporte trois grandes articulations réparties comme suit : Etude institutionnelles, missions réalisées, Bilan et discussions.

Chapitre 1 : Étude institutionnelle et présentation de la mission

I- Présentation de la SNCF

1.1. Fiche synoptique de présentation

Logo	
Raison sociale	SNCF
Direction	<u>Guillaume Pepy, Patrick Jeantet</u>
Siège social	<u>La Plaine Saint-Denis, Saint-Denis (France)</u>
INFORMATIONS JURIDIQUES	
Date de création	1937
Forme juridique	Etablissement Public à caractère Industriel et Commercial (EPIC)
INFORMATIONS ÉCONOMIQUES, COMMERCIALES ET FINANCIÈRES	
Secteur d'activités	Secteur tertiaire: transport ferroviaire
Chiffre d'affaires	31,4 Md€
Filiales	SNCF Réseau, SNCF Mobilités, KEOLIS

1.2. Historique et évolution

La Société nationale des chemins de fer français (SNCF) est l'entreprise ferroviaire publique française, officiellement créée le 1^{er} janvier 1938 en application du décret-loi du 31 août 1937. Elle est notamment présente dans les domaines du transport de voyageurs, du transport de marchandises et réalise la gestion, l'exploitation et la maintenance du réseau ferré national dont

elle est propriétaire¹. Depuis la réforme ferroviaire en 2015, SNCF est constituée de trois EPIC complémentaires: l'EPIC de tête, SNCF Mobilités et SNCF Réseau.... Il est détenu par l'État actionnaire unique. Depuis 2015, c'est un nouvel EPIC SNCF qui assure le contrôle et le pilotage stratégique, la cohérence économique, l'intégration industrielle et l'unité sociale de l'ensemble. Deux autres EPIC, SNCF Réseau et SNCF Mobilités sont liés dans ce même ensemble pour que l'entretien de l'infrastructure et la gestion des circulations prennent en compte les contraintes de l'exploitation des services de transport ferroviaire.

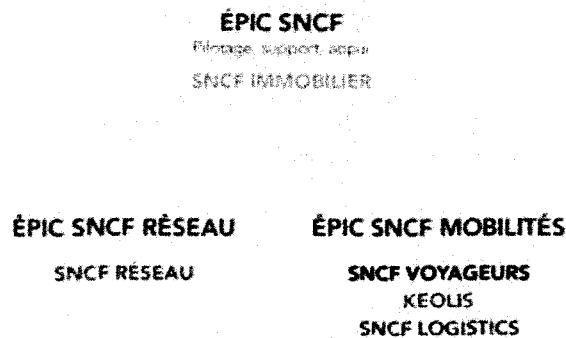


Figure 1: Segmentation des EPIC SNCF / **Source:** site sncf.com

SNCF Réseau regroupe les fonctions de gestionnaire d'infrastructures du réseau ferré national, jusqu'alors réparties entre Réseau Ferré de France, SNCF Infra et la Direction de la voie ferroviaire (DCF). Quant à SNCF Mobilités, cet EPIC assure l'ensemble des activités d'exploitation des services de transport ferroviaire qui étaient celles de « l'opérateur historique » SNCF.

1.3. Activités et fonctionnement

Structurées aujourd'hui en trois EPIC (établissements publics à caractère industriel et commercial), les compétences sont réunies autour de cinq grands métiers : SNCF Réseau, SNCF Voyageurs, Keolis, SNCF Logistics et SNCF Immobilier.

¹ https://fr.wikipedia.org/wiki/Soci%C3%A9t%C3%A9_nationale_des_chemin_de_fer_fran%C3%A7ais

SNCF Réseau gère, maintient et développe le réseau ferré national. D'ici 2020, la priorité est double : rénover et maintenir le réseau existant, et se concentrer sur l'Île-de-France. SNCF Réseau a pour ambition d'avoir le réseau le plus sûr, le plus performant et le plus innovant.

SNCF Voyageurs propose à ses clients, autorités organisatrices et voyageurs, une palette de solutions de mobilité partagée de porte à porte pour répondre à leurs besoins en termes d'offre, de coût, de qualité de service et de respect de l'environnement.

Keolis est un opérateur majeur du transport public de voyageurs en France, en Europe et dans le monde. Son ambition est de proposer à ses clients autorités organisatrices des transports (AOT) des solutions de mobilité sur mesure adaptées aux problématiques de chaque territoire et aux évolutions des habitudes de déplacement des voyageurs. Opérateur de tous les modes de transport collectifs et partagés, Keolis assure aux utilisateurs de ses transports des trajets toujours plus fluides et confortables.

Fort de la diversité de ses métiers, SNCF Logistics propose des solutions multimodales personnalisées qui allient performance économique, environnementale et sécurité. C'est le partenaire de la compétitivité des entreprises dans le monde entier.

SNCF Immobilier quant à lui, regroupe les savoir-faire de SNCF en matière de gestion et de valorisation immobilière, d'aménagement urbain et de logement. Le métier permet à SNCF de s'inscrire dans les objectifs poursuivis par le Gouvernement en faveur du logement. La priorité porte, notamment, sur :

- l'accélération des cessions/libérations de fonciers constructibles ;
- la mise en place de partenariats avec les villes et métropoles pour faciliter la programmation des opérations de logements ;
- le renforcement des synergies entre les différentes filiales immobilières spécialisées du groupe ferroviaire.

II- Le service Transparence & Open data

SNCF a conclu en juin 2014, pour trois (03) ans un partenariat avec Transparency International qui a permis de piloter le programme transparence SNCF. Dans ce cadre se sont tenues 7 réunions du comité de pilotage, de nombreux échanges et des recommandations (open

data, conflits d'intérêts, affaires publiques, achats, lanceurs d'alertes...etc) de la part de Transparency International.

2.1. La démarche d'ouverture des données de transports

La demande de transparence est de plus en plus forte, tant auprès des administrations qu'au niveau des entreprises². Le secteur du transport est particulièrement actif et est aujourd'hui amené à s'impliquer encore plus dans cette démarche du fait des textes législatifs récents. En 2015 et 2016, les lois Macron, Valter et Lemaire ont successivement contribué à établir le cadre législatif de l'open data en France. Du fait de leurs statuts et missions, les EPIC mais aussi les entités délégataires de service public du groupe (Keolis...) entrent directement dans le champ d'application de ces lois à différents titres.

Plus précisément:

La loi pour la croissance, l'activité et l'égalité des chances économiques (dite loi Macron) du 06 août 2015 influence les obligations du Groupe Public Ferroviaire et particulièrement son activité Voyageurs. Elle fait peser de nouvelles obligations sur les opérateurs et autorités organisatrices de transport, en venant consacrer l'obligation de diffusion libre, immédiate et gratuite des données des services réguliers de transport public de personnes et des services de mobilité.

La loi Valter du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public s'est attachée aux conditions de réutilisation des données en posant un principe de gratuité des données mais tout en listant un certain nombre d'exceptions qui ne seront pas applicables à SNCF au regard du dernier décret d'application.

La loi pour une République Numérique adoptée le 7 octobre 2016 dite "Loi Lemaire" vient achever le processus de transposition des textes européens en imposant une véritable obligation d'ouverture des données. Elle vient renforcer l'ouverture des données publiques, crée un service public à la donnée confié à l'Etat et traite de la diffusion des données d'intérêt générale en vue de leur réutilisation par tous et définit les biens communs de l'information. En amendant la loi dite "CADA" de 1978 (relative à l'accès aux documents administratifs), elle

² cf. Charte sur la transparence des données SNCF

oblige les EPIC du Groupe Public Ferroviaire à diffuser spontanément en ligne dans un standard ouvert certaines données entendu au sens très large (documents, codes, sources, bases de données et données) et supprime la possibilité pour ces derniers de s'opposer à la réutilisation des informations publiques (des exceptions sont maintenues ou introduites relatives à la sécurité, au secret en matière commerciale industrielle et aux droits de propriété intellectuelle des tiers).

2.2. La mise en œuvre de l'open data dans le contexte SNCF

Le groupe SNCF s'est engagé depuis 2012 dans une ambitieuse politique d'ouverture des données, accélérée en 2014 avec la mise en œuvre du programme Transparence. Cette politique s'est traduite par l'ouverture de près de 50 jeux de données Open data, et la publication des courriers du président, de nombreux rapports et archives, répond à plusieurs objectifs.

Tout d'abord, un devoir de redevabilité. En tant que grande entreprise publique, SNCF se doit de rendre compte aux citoyens et à ses clients de son activité. Cette démarche présente aussi l'intérêt pour l'entreprise de montrer l'étendue et la réalité de ses missions en luttant contre certaines idées préconçues.

En externe, cette démarche constitue aussi un accélérateur d'innovations. Le libre accès aux données permet de faire émerger de nouveaux services adaptés aux usagers. En interne, c'est aussi une véritable ressource pour SNCF tant pour les SI (Systèmes d'informations) que pour les métiers, qui peuvent partager leurs données. Cette diffusion plus large des données est un moyen pour l'entreprise d'initier un changement. Outre le fait de répondre aux obligations légales qui lui sont contraintes d'ouvrir ses données de transport, la SNCF va plus loin en ouvrant, de plein gré, des données supplémentaires, issues de divers métiers et regroupées par sujet, par thème ou par type. Ainsi, à l'heure actuelle, des données à caractère social, de sûreté, de sécurité ou encore concernant les ressources humaines sont accessibles sur le portail Open Data SNCF. Chaque ouverture résulte d'un processus plus ou moins long de négociations visant à convaincre les producteurs de données d'autoriser la mise en ligne et la réutilisation des données par tous.

L'ouverture des données se veut donc à la fois technique, juridique et contrôlée:

- Technique: l'ouverture et le partage des données doit être fait dans des formats exploitables et réutilisables. Autrement dit, la donnée est proposée dans un format technique compréhensible par du code informatique
- Juridique: l'ouverture est encadrée par des textes législatifs et évolue selon les nouvelles prescriptions. A ce jour, la donnée est réutilisable aux termes d'une

licence de type ODBL sans renoncement au droit de propriété intellectuelle. Suivant sa nature, la donnée est partagée librement sous un modèle freemium (protocole Macron)

2.3. Organisation et Missions

Le service dépend directement de la Direction de l'information et de la stratégie de communication (cf. **Annexe 1 - organigramme**), ses missions principales s'axent autour de l'animation interne et externe de l'open data et la promotion de la transparence des activités de l'entreprise.

Les actions menées au sein du service sont:

- définir la stratégie open data de l'entreprise et la gouvernance des données
- assurer l'implication des différents métiers dans la démarche open data, d'encourager la gestion autonome de la plateforme par chaque référent open data
- administrer de manière fonctionnelle et technique du catalogue de données SNCF: s'assurer du bon fonctionnement de la plateforme, tant au niveau de la production qu'au niveau de la gestion des utilisateurs
- suivre les mises à jour et les nouvelles créations: Mettre à disposition des données selon le programme transparence, les demandes internes et externes
- procéder au monitoring de la plateforme : surveillance et mesure des actions effectuées sur la plateforme
- assurer le reporting et l'amélioration qualité :
- accompagner les utilisateurs de l'open data, suivre les projets open data et les réutilisations: Faire une analyse des données SNCF publiées sur d'autres plateformes open data / Suivre des développements externes effectués à partir des données mis à disposition
- établir une veille juridique sur les sujets open data afin de se prémunir contre les risques juridiques liés aux sujets open data du domaine des transports

III- Mission et objectifs

La plateforme open data du groupe SNCF reflétant ainsi la vitrine de ses missions de transport public et un support de communication pour animer les opérations a besoin d'être gérée de manière interactive avec des ajouts réguliers de contenus. La mise à jour apparaît dès lors

primordiale pour donner de la visibilité et valoriser le patrimoine informationnel que constituent les données de transport SNCF. De plus, elle permet de fidéliser les utilisateurs, plus il y a du contenu, plus la plateforme cumule un grand nombre de requêtes, la mise à jour est également l'occasion pour le groupe d'afficher son expertise en la matière, de confirmer son positionnement et de répondre aux enjeux du mouvement qui émerge. C'est dans cette optique que la mission d'animation de la plateforme open data a été proposée comprenant ainsi des opérations transverses, entre autres la collecte de données, rapatriement de l'information, et valorisation de l'information.

Les objectifs assignés à cette mission sont de permettre à chaque producteur d'une part de centraliser ses données pour les retrouver facilement, de mettre à disposition des informations fiables et de qualité, ensuite d'homogénéiser la présentation et les métadonnées des jeux de données et enfin de montrer l'intérêt et la pertinence de la donnée publiée en lui donnant du sens.

Chapitre 2 : Mise à jour du catalogue data.sncf, optimisation des graphiques d'analyse et reporting de l'activité des domaines

La plupart des entreprises aujourd'hui se donnent pour défi d'assurer une meilleure gestion de leur information et de la rendre disponible pour servir des réutilisations utiles et profitables. Pour la mener à bien il importe de collecter et gérer cette information, depuis une ou plusieurs sources, et à la distribuer auprès de l'audience requise.

Pour mettre en œuvre sa politique open data, la SNCF dispose d'une plateforme fournie par le prestataire OpenDataSoft (ODS). ODS est une solution qui permet d'ouvrir ses données pour les partager entre collaborateurs pour faciliter la mutualisation des informations, c'est la possibilité de partager des données avec différents publics, et notamment le grand public. C'est aussi l'opportunité pour les utilisateurs de pouvoir voir toutes les données qui sont mises en ligne dans tout le groupe (*possibilité de voir le tableau qui contient toutes les données à plat*), et d'avoir un système de filtres assez simple pour optimiser la recherche des jeux de données. ODS est un moyen plus pratique d'accéder aux données brutes avec des métadonnées précises pour chaque jeu de données tout en respectant la logique "open" de base.

Notons la formule *Open = accessibilité + format + interopérable*, tout le monde peut donc utiliser sans besoin d'avoir recours à un logiciel spécifique, le fait est de pouvoir jouer avec la donnée et lui donner du sens à travers des visualisations. Une des règles de l'open data sur le plan juridique est l'accessibilité, et dans cette optique, l'authentification n'est pas toujours nécessaire (Toutefois, il est possible de pousser à l'authentification pour encourager une certaine relation avec des utilisateurs privilégiés qui aiment utiliser les données SNCF par le biais de l'information). Le processus d'authentification peut servir dans ce cas à pousser des informations aux utilisateurs, instaurer un climat de confiance, et les inviter si possible à des échanges dans l'ultime but de favoriser la création de services innovants. « *Ouvrir à tous mais identifier parmi ces tous ceux avec qui on pourrait avoir une relation privilégiée pour faire partie de notre écosystème plus ou moins lointain* »

Réutiliser revient à se connecter automatiquement aux API générées pour récupérer les données. Les formats de données supportés par la plateforme, c'est-à-dire desquels il est possible d'extraire de l'information sont: .csv, .dat, .json, .geojson, .jpeg et .png (qui permettent de créer une galerie d'images), .kml, .kmz, .ods, .osm, .tsv, .txt, .xml, .xls, .xlsx, .zip. (autres formats spécifiques). Toutefois, d'autres formats ne sont pas supportés car ils n'ont pas de vue tabulaire. Cependant, ils peuvent être téléchargés et attachés à un jeu (onglet information dans la vue édition du jeu), il s'agit des formats suivants: .pdf, .doc, .docx, .ppt, .pptx, .gifs.

Les plus utilisés sont:

- CSV : format le plus interopérable qui soit
- XLS : spécificités d'encoding pour la langue française
- JSON : format informatique, en lien avec le javascript et la plupart du temps exploitable par les différents systèmes d'exploitation.

La solution permet également de faire des branchements à des API (c'est-à-dire permettre une connexion automatique à ses données, qui alimente d'autres systèmes d'informations). L'avantage de l'API est l'interopérabilité, et donc l'échange entre les outils informatiques qui ont une très forte tendance à se multiplier.

ODS permet également de cadrer au maximum l'ouverture, d'avoir et de donner la maîtrise des jeux qu'on ouvre au moyen de comptes utilisateurs, et de politique de confidentialité.

L'entreprise SNCF souhaitait accentuer sa politique open data en ouvrant davantage des données, il a fallu réaliser les actions y relatives, c'est-à-dire collecter les données, les éditorialiser, les publier et réaliser des reporting. Nous avons donc pensé comment valoriser les données et la forme que pouvait avoir cette démarche de valorisation (optimisation des graphiques d'analyse-indicateurs, création des tableaux de bord et pages, projet data hebdo, mise à jour du dashboard interne).

I- Bien-fondé de la réalisation

Disposer de l'information pertinente et de qualité était autrefois dédié aux élites managériales, mais aujourd'hui tout le monde peut y accéder à un certain niveau, filtrer l'information, et la transformer. De nouveaux défis informationnels auxquels sont confrontées les entreprises.

La question centrale était surtout comment stimuler les acteurs à l'utilisation de la donnée? Comment apporter un aspect ludique à la donnée en lui donnant plus de sens? L'objectif principal étant de valoriser des données dynamiques et interactives. La particularité avec la mise à disposition et l'échange des jeux de données sur le portail open data est la réutilisation, notamment par exemple des réutilisateurs qui se servent des données pour créer d'autres sites d'information, des cartes par exemple (*Cf. Annexe 2*), la mise à jour de ces données alimente par exemple des sites qui vont intéresser des voyageurs à destination de la Bretagne qui vont pouvoir avoir une information sur mesure, s'assurer de la fiabilité et de la qualité puisque nous en sommes le producteur. Avoir l'information tous les jours en temps réel et en être certain puisqu'elle provient des producteurs identifiés de la donnée.

Il s'agit donc de la réutilisation et c'est le système de l'API qui le permet. L'open data c'est donc partager de la donnée, s'assurer de la fiabilité et de la qualité des informations qui circulent. Seulement, la mise à jour n'est pas automatisée pour la plupart des données, or une mise à jour non régulière et mal appliquée engendre de mauvaises interprétations et fausse les réutilisations.

Nous avons pensé dans un premier temps à mettre à jour des données et ensuite adapter les graphiques aux prochaines actualisations, cela permet de mettre en scène des données et de présenter la réalité des missions faites au service des usagers par exemple (services en gares, régularité des trains...).

II- Méthodologie

La réalisation des missions assignées a nécessité de suivre plusieurs étapes d'analyse afin de nous enquérir de l'existant en matière de process de mise à jour, d'outils de travail, et par la suite de proposer des solutions optimales. Les choix méthodologiques ont essentiellement porté sur des opérations d'audit, l'analyse de l'existant, et le benchmarking.

2.1. Analyse de l'existant et définition des besoins

Afin d'établir un inventaire analytique de l'ensemble des données publiées en open data, la phase d'audit a été échelonnée sur trois (03) semaines environ. Des échanges ont également eu lieu entre les référents open data, dans le but de valider les outils mis en place et d'approuver les

résultats de la phase d'étude. L'opération d'audit a été menée par le biais des outils existants intégrés dans la solution open data. Nous avons interrogé le catalogue complet des données présentes sur le site pour avoir de la visibilité sur certains manquements avant d'établir une politique de mise à jour des données.

1- Problématique et objectifs	2- Acteurs
<p>La mise à jour est une opération importante dans la démarche de transparence. Elle est un sujet très controversé tant en interne qu'en externe.</p> <p>Objectifs:</p> <ul style="list-style-type: none"> - Structurer la gouvernance de la plateforme - optimiser la gestion de flux de données - donner plus de visibilité à la plateforme 	<ul style="list-style-type: none"> - Guillaume LEBORGNE - Stella NGANDJOU - Léa DARMON - Stéphane HEQUET - Sandrine DELORME - Nicolas RASTOUIL

3- Bilan de l'existant (source : rapport OpenData fait par Sandrine DELORME et Stéphane HEQUET)

Etat des jeux de données:

Publics	Publics-non publics	Restreints	Restreints-non publics	Total
61	0	22	0	83

Thème	Nbre de jeux de données publics par thème	Nombre de jeux de données restreints	Total
Billetique	3	0	3
Comptage de flux	2	0	2
Développement durable	1	0	1
Gares et infrastructures	12	2	14
Moraires et itinéraires	10	1	11
Informations institutionnelles et financières	6	7	13
Qualité de service	7	6	13
Sécurité ferroviaire	10	2	12
Service voyageurs

Social et ressources humaines	18	0	18
Systèmes d'informations	0	0	0
Non renseigné	2	0	2
Total	83		

Nombre d'utilisateurs	6,9k
Appels d'API par utilisateur	92

4- Diagnostic

Ce bilan fait ressortir les défauts liés au processus de traitement et de valorisation de la donnée

Problème de mise à jour

Réutilisations minimales pour certains jeux de données

Problème de valorisation

Circuit de l'information data / production - publication - suivi

Automatisation des mises à jour

Optimisation des recherches (tags, RI...)

Jeux de données à mettre à disposition après évaluation des critères de recherche

5- Axes d'amélioration retenus

<i>Actions à court terme</i>	<i>Actions à long terme</i>
Petites modifications immédiates à faire selon les erreurs d'orthographe ou de logique dans la structuration	Mise en place du programme d'amélioration qualité
Créer, publier et mettre à jour les JD dont les informations sont à portée de main	Rédaction du manuel qualité / manuel des procédures
	Mise à jour continue de la plateforme

Tableau 1: Synthèse d'audit du catalogue data.sncf

- Analyse des besoins

Le principal livrable attendu était un catalogue mis à jour et exploitable.

2.2. Benchmarking

Afin d'améliorer l'existant, il nous a paru important d'évaluer d'autres plateformes open data afin de déceler les méthodes mises en place, s'en inspirer et d'en tirer le meilleur. Environ dix (10) sites ont été analysés. Le choix de ces sites a été fonction de la solution mise en place pour le contrôle de la création des données, qui était la même que celle du groupe SNCF, et de leur apparition dans des médias de communication intéressés par l'open data.

Cette première phase d'analyse portait sur la présentation des jeux de données et la manière de les interpréter et les valoriser.

Type de structure	Nombre	Dénomination
Organismes de transport	5	Transpole par le site open data de la MEL, KEOLIS Rennes par le site open data STAR, RATP, STIF
Organismes culturels	3	Ministère de la Culture et de la Communication, les archives d'Albert KAHN, inventaire de la bibliothèque d'André-Desguine
Plateformes régionales	2	Open data Ile de France, portail open data PACA

Tableau 2: Structures ayant servi au benchmarking

III- Réalisations et préconisations

Cette partie reprend pour chaque action à mener l'objectif escompté et les préconisations y relatives. A cela, s'ajoutent le plan d'actions et échéanciers établis après la phase d'audit.

3.1. Mise à jour du catalogue, méthodes et alertes

3.1.1. Objectif

L'objectif était de prévoir un plan de curation (plan d'actions + échéancier) du catalogue de données SNCF et un audit complétion des métadonnées. Les enjeux ici sont multiples, il est important de favoriser de bonnes conditions pour une meilleure indexation en mettant à jour les contenus proposés et privilégier l'expérience d'exploitation d'un internaute en s'adaptant à ses habitudes de recherche et en lui offrant des nouveautés régulièrement.

3.1.2. Préconisations et mise en œuvre:

- Le téléchargement du catalogue au format csv et l'exploitation du jeu de données sur la plateforme ODS

La plateforme ODS met à disposition un lien pour récupérer dans un format spécifique ou générer un fichier exploitable de l'ensemble des données SNCF, une fonctionnalité d'ailleurs utile pour optimiser la gestion et l'administration du portail. Cet export du catalogue permet

d'avoir une vue générale sur les données, leurs métadonnées et de vérifier à l'aide d'Excel ou tout autre tableur les informations à ajuster. Une fois importé, le catalogue est réinjecté et traité dans la plateforme comme un jeu de données au même titre que les autres.

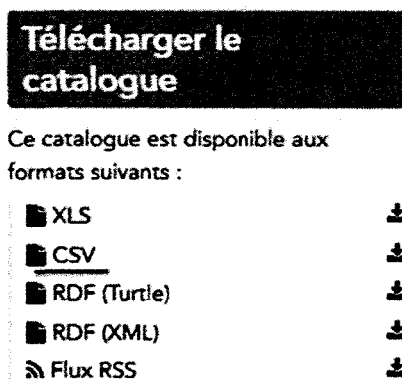


Figure 2: Procédure d'export du catalogue de données

- Filtrage, choix des facettes et requêtes (métadonnées DCAT: titre, producteur, licence, fréquence)

La consultation des données à l'aide d'un filtre constitue un moyen d'explorer rapidement un nombre important de données. L'utilisation de ces filtres ou requêtes permet d'afficher les données souhaitées et de masquer le reste, dans notre cas, il a fallu faire ressortir les cellules non renseignées afin d'évaluer le nombre de jeux de données sans les métadonnées indispensables ou encore les jeux de données non actualisés. Des requêtes spécifiques sont donc saisies dans le moteur de recherche de la plateforme avec l'attribut #null qui est un marqueur spécial qui permet d'indiquer si une information manque dans une base de données. Les informations qui nous paraissaient importantes à vérifier et à renseigner dans l'urgence étaient matérialisées par les requêtes: #null(publisher) ou #null (dcat_creator) pour vérifier si le producteur est renseigné à chaque fois, #null(identifiant technique du champ) pour l'identifiant technique du jeu utile pour les réutilisations, et parfois #null (dcat_accrualperiodicity) pour les dernières dates de mises à jour des données.

Un tableau de requêtes de recherche a été conçu pour le filtrage aisé des jeux de données et la conception des indicateurs.

Recher des enregistrements	REQUETE	EXEMPLE à partir du jeu ods-epi-monitoring	RESULTAT
qui contiennent une valeur	valeur	explore	tous les enregistrements contenant la chaîne de caractère explore
dont un champ contenant une valeur	id_champ valeur	action explore	tous les enregistrements contenant explore dans la colonne action
qui contiennent plusieurs valeurs	valeur1 AND valeur2	explore AND search	tous les enregistrements contenant à la fois explore et search
qui contiennent au moins une des valeurs	valeur1 OR valeur 2	action OR explore	tous les enregistrements contenant soit explore soit search
qui ne contiennent pas une valeur	NOT valeur	NOT anonymous	tous les enregistrements qui ne contiennent pas la chaîne de caractère anonymous
qui contiennent une valeur exacte	#exact(id_champ, « valeur »)	#exact(user_id, "anonymous"	tous les enregistrements contenant l'exacte chaîne de caractère anonymous dans la colonne user_id
si un champ est vide	#null(id_champ)	#null(referer)	tous les enregistrements qui ont une cellule vide dans la colonne referer
dont la date est antérieure à une valeur	id_champ date<=AAAA/MM/JJ	timestamp<=2016/09	tous les enregistrements ayant une date antérieure ou égale au mois de septembre
jusque maintenant	id_champ date<=#now(days=-valeur)	timestamp<=#now(days=-20)	tous les enregistrements ayant une date comprise dans les 7 derniers jours
depuis un point	#distance(latitude,longitude,distance en mètres)	#distance(48.866667,2.333333,1000)	tous les enregistrements géolocalisés à 1 km du centre de Paris

Tableau 3: Requêtes utilisées pour l'exploration des jeux de données

- Elaboration d'un plan d'actions: définition des objectifs, métadonnées obligatoires et calendrier de mise à jour avec indicateur quantitatif

Ce plan d'actions a été un outil de performance pour le projet d'actualisation ou de mise en qualité de la plateforme OPEN DATA. Il permettait de déterminer les opérations à effectuer par rapport aux écarts constatés, et de spécifier les moyens à utiliser afin d'atteindre les objectifs fixés.

							CODE COULEUR
Etablir un référentiel des bonnes pratiques	Identifier les acteurs / rôles / définir les procédures (production / publication / mise à jour) / Questionnaire / planification de la formation	Uniformiser les pratiques de gestion de la plateforme	Transparence & open data / copil open data	Réunions de travail	01/04/2017		Action indispensable pour la gestion optimale de la plateforme
Formations acteurs usagers	Questionnaire / planification de la formation	Management / Actualisation des usagers	Transparence & open data	Questionnaires / enquêtes	15/11/2016		La non-réalisation de cette opération pourrait impacter la gouvernance
Réorganiser groupes et usagers	Analyses et les utilisateurs selon leurs profils et déterminer pour chaque un groupe	Lancer les communications et assurer la visibilité des actions effectuées sur la plateforme / Partage des permissions sur la plateforme / structurer la gouvernance de la plateforme	Transparence & open data	réunions de travail	06/12/2016		
Refonte et mise à jour	Effectuer les mises à jour sur les jeux de données prioritaires (en fonction des JO les plus consultés)	Réunion et communication des utilisateurs / Mise à jour des données / logues de connectivité / transparence	Transparence & open data	réunions de travail / reporting			
Collecte	Identification des producteurs			rencontres de travail sur l'année / Temps pour interviewer les acteurs de la plateforme			Tâches incluses à la fois dans l'opération de mise à jour et la gestion quotidienne de la plateforme
Traitement	Structuration / Modélisation						

Tableau 4: Plan d'actions de mise à jour du catalogue data.sncf

- L'élaboration d'un référentiel de fréquences de mise à jour (par ex: trimestriel ou tous les trois mois, et en plus une catégorie pour les données froides à ne pas mettre à jour)

La mise à jour dépend des jeux de données concernés. Le référentiel de mise à jour précise pour chaque jeu de données:

- la périodicité de mise à jour (annuelle, trimestrielle...)
- le type de mise à jour
- les sources de données / producteurs
- le mode de réception

Titre	Domaine	Producteur	visibilité	Références MAJ			Type MA	Fréquence	Mode de réception
				Documents	Referents / interlocuteur				
JEP micro-services TER	StationCF	TER	visible						
Nombre de voyageurs en retard sur TER	StationCF	TER	restreint						
Répartition par genre de l'effectif	StationCF	Direction cohésion & ressources humaines	visible	Site web (page 9)	MALOUYRE Didier	Mensuel	Annuel	Site	
Accidents du travail	StationCF	Direction sécurité & Qualité de Service	restreint						
Accidents passagers depuis 2006	StationCF	Direction sécurité & Qualité de Service	visible	Doc annuel rapport de sécurité (page 9)	VENDREUX Yann	Mensuel	Annuel	Site	
Accompagnement de personnes à mobilité réduite dans les gares	StationCF	Games & Commissions	visible						
Accords collectifs nationaux	StationCF	Direction cohésion & ressources humaines	visible						
Adhérence	StationCF		restreint						
Adresses et géolocalisation des établissements SNCF	StationCF	Direction cohésion & ressources humaines	restreint	Fichier EXCEL	PATRICE BON-BRANDON CARNARO / Goussard de Curdi	Mensuel	Annuel	Site	
Agents SNCF ayant leur repos le dimanche	StationCF	Direction cohésion & ressources humaines	visible	Site web (page 47)	MALOUYRE Didier	Mensuel	Annuel	Site	
Agents SNCF de nationalité étrangère	StationCF	Direction cohésion & ressources humaines	visible	Site web (page 12)	MALOUYRE Didier	Mensuel	Annuel	Site	
Agents SNCF en situation de handicap	StationCF	Direction cohésion & ressources humaines	visible	Site web (page 29)	MALOUYRE Didier	Mensuel	Annuel	Site	
JEP Micro-services Transition	StationCF	Transition	restreint						
JEP Mobilité	StationCF	CanalTP	visible						
JEP Temps réel Transition	StationCF	Transition	visible						
JEP Théorique et Temps réel	StationCF	SNCF	visible						
Archives SNCF	StationCF		visible						
JT Ligne	StationCF		restreint						
JT TGV	StationCF		restreint	Fichier CSV					
Requêtes MATÉRIEL - RRD	StationCF	ODJLR	restreint						
Sauvegarde satisfaction clients en gare	StationCF	Games & Commissions	visible						
Cartographie OpenStreetMap des gares SNCF en	StationCF		restreint						

Tableau 5: extrait du référentiel de mise à jour

- La mise en qualité de l'existant : Cette étape a consisté à renseigner les jeux sans information et en particulier, dans la vue édition d'un jeu, sous "l'onglet information".
- Collecter les autres données estimées importantes pour la publication

L'optimisation du catalogue ne saurait se limiter à l'actualisation de l'existant mais aussi à son enrichissement par de nouvelles entrées. Il apparaît alors évident de prendre en compte les requêtes non abouties des internautes afin d'anticiper sur les données à publier. Cela a été possible grâce aux reports des requêtes sans résultats du portail open data. Ces résultats sont visibles dans le tableau de suivi intégré à la solution open data.

C'est dans cette optique qu'ont été créés en plus près de 30 jeux de données sur les tarifs, abonnements, statistiques, menus restauration et bien d'autres. Ainsi, les utilisateurs entrent dans le processus de publication de l'information, ce qui permet à l'entreprise de rester au plus près de leurs attentes.

3.2. Moissonnage des données externes et interopérabilité

Le moissonnage³ permet aux administrateurs de facilement créer et mettre à jour un grand nombre de jeux de données en les important depuis une source externe telle que, par exemple, un catalogue CSW ou un service ArcGIS.

Les deux principaux usages des moissonneurs sont :

- Amorcer votre domaine avec des jeux de données provenant d'un portail existant
- Maintenir vos jeux de données synchronisés avec un service externe

Les moissonneurs créent les jeux de données, mettent à jour leurs métadonnées et ressources, les maintiennent à jour, et les publient.

C'est un procédé qui peut être très long sur les services externes qui contiennent de nombreux ou de gros jeux de données. Il est donc séparé en deux phases.

- Premièrement, le moissonneur se connecte au service externe et découvre tous les jeux de données qu'il contient. Il crée ensuite un jeu de données non publié pour chaque jeu de données distant qu'il trouve. Ces jeux de données contiennent toutes les métadonnées et ressources disponibles.
- Ensuite, le moissonneur traite et publie tous les jeux de données moissonnés.

3.2.1. La fédération des jeux de données externes, le cas de KEOLIS

La fédération permet de collecter des sources de données externes sur son propre domaine. Dans la logique de centraliser les données du grand groupe SNCF, le service open data a souhaité référencer les données appartenant à KEOLIS, une de ses principales filiales.

a. Objectif

Cadrer le projet après une première estimation technique pour récupérer les données avec test sur la sandbox (environnement bac à sable).

b. Résumé de l'estimation,

Plusieurs cas de figure nécessitaient d'être pris en compte dans le projet, ont été identifiés:

- 6 sources OpenDataSoft (Rennes, Angers, Ile-de-France, St Malo, Lille et Dijon)

³ **Source:** *OpenDataSoft*

- 3 sources avec API à disposition sur des plateformes autres qu' OpenDataSoft (Bordeaux, Brest, Lyon), téléchargeables et exploitables après avoir créé un compte. Pour exemple, 87 jeux de données récupérés depuis l'API Cub de Bordeaux, et disponibles dans le bac à sable.
- 4 sources plus complexes (Aix-en-Provence, Metz, Pau et Strasbourg) en raison des formats. Des connecteurs spécifiques sont à prévoir pour la récupération automatique des formats GTFS.

c. Mise en œuvre

La mise en œuvre de cette tâche a nécessité de passer par un certain nombre d'étapes, à savoir:

- La création d'une cartographie des jeux de données disponibles (fédération ODS, lien API WFS, connecteurs spécifiques ou index de liens), pour avoir une vision on ne peut plus claire et exhaustive de tout ce qui est publié comme donnée dans le cadre des activités de Keolis.

Après un test dans l'environnement bac à sable, on a pu se rendre compte que l'opération de fédération des jeux de données sur la plateforme pouvait entraîner quelques contraintes notamment:

- L'augmentation des appels API, avec pour conséquence directe une surcharge dans les requêtes
- Des requêtes trop importantes (à moins de limiter le nombre d'appels API par jeu de données, mais cette solution n'est pas très adéquate)

	Récupération manuelle	Récupération automatique
Risques	Contraintes de coût et de temps (tant pour les prestataires que pour le personnel du service)	Limitation de quotas et entrave à la gestion optimale de la plateforme

Tableau 6: Analyse des risques liés au projet de fédération des jeux de données KEOLIS

Ces difficultés amènent à se poser des questions sur les éléments précis à publier (surtout pour le contenu API) et comment cela pourrait impacter les quotas ?

Pour les jeux de données complexes à récupérer, nous avons pu parvenir à une solution de référencement des sources (la création des jeux de données et le pointage vers les pages sources via des liens).

- la mise en production

Les actions prévues pour cette étape se déclinent comme suit:

- Recenser les liens sources des jeux de données sélectionnés pour ensuite créer de nouveaux jeux de données dans la sandbox et reprendre les informations descriptives de base, cette opération implique de travailler dans un premier temps dans l'environnement de test afin d'effectuer tous les tests possibles et d'effectuer les vérifications nécessaires.
- la migration des tests vers le domaine principal: C'est le passage à la production proprement dit, les jeux de données sont reportés sur l'espace principal dédié à la communication des jeux de données.

c- Résultats



Figure 4: Résultat de la fédération des jeux de données KEOLIS

3.2.2. Le référencement des archives sur Gallica par le biais de l'inventaire des archives en open data

Le référencement des archives SNCF sur Gallica est une opération qui va permettre de multiplier l'accès au patrimoine archivistique à travers l'échange et l'interopérabilité des métadonnées disponibles.

a. Objectif

Le but était de reconstituer le corpus numérisé sur le site Gallica afin de lui apporter plus de visibilité.

b. Mise en œuvre et préconisations:

- La réalisation d'un mapping entre les métadonnées des archives SNCF et les attributs du Dublin Core.

La SNCF dispose d'un mur d'archives virtuel "Open Archives"⁴, un projet abouti en 2016. Le but de ce portail est de communiquer des archives historiques sous différents supports: textes, vidéos, sons. Il vise à faciliter l'accès aux documents conservés dans différents centres d'archives en proposant des versions numérisées. En effet, les archives SNCF sont des archives publiques et à ce titre elles peuvent être, suivant certaines règles, consultées par tous. Le site s'adresse donc autant aux spécialistes, universitaires, ferroviphiles⁵ ainsi qu'à un public plus large. Au niveau de la réalisation, « Open Archives » est le résultat de la réutilisation d'un jeu de données recensant l'ensemble des archives avec un certain nombre de métadonnées et rendus disponibles en open data sur Open Data SNCF⁶, relayée sur un site web grand public présentant une interface plus facile d'accès.

Le référencement des archives sur Gallica exige de respecter les éléments du format Dublin Core pour faciliter l'interopérabilité. Or, la description actuelle des archives SNCF n'a pas été réalisée selon les exigences définies. Il y avait donc une nécessité de faire correspondre les métadonnées existantes avec les éléments du Dublin Core, c'est-à-dire de réaliser avant tout une revue des catégories type Gallica BNF (*image ci-dessous*) dans un fichier de correspondance pour ne faire le travail qu'une fois dans le jeu de données.

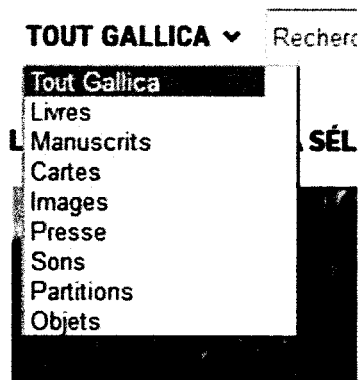


Figure 5: Typologie Gallica

⁴ <http://www.sncf.com/fr/transparence/open-archives>

⁵ Amateur passionné de chemins de fer (wiktionnaire)

⁶ <https://datasncf.opendatasoft.com/explore/dataset/archives-sncf-new/?sort=thematique>

- La traduction ou la conversion (via une moulinette informatique / un batch) des informations (les données en dublin core) dans une base de données avec un modèle orienté OAI (Open Archives Initiative).

Le protocole utilisé par la BNF pour assurer l'interopérabilité est l'OAI/PMH. OAI-PMH⁷ est le sigle d'Open archives initiative - protocol for metadata harvesting, ce qui signifie "protocole pour la collecte de métadonnées de l'Initiative pour les Archives ouvertes". C'est un moyen d'échanger sur Internet des métadonnées entre plusieurs institutions, afin de multiplier les accès aux documents numériques. Il permet d'accroître la visibilité des collections numériques sur Internet, de reconstituer virtuellement des corpus à partir de ressources accessibles sur différents sites, d'alimenter des portails thématiques.

L'OAI-PMH définit deux types d'acteurs : les fournisseurs de données, qui déposent leurs métadonnées sur un serveur web appelé "entrepôt", et les fournisseurs de service qui collectent (on dit aussi "moissonnent") ces données pour les intégrer à l'index de leurs propres bibliothèques numériques. Un même établissement peut jouer les deux rôles, diffuser ses métadonnées et collecter celles des autres.

Le fonctionnement de base du protocole OAI-PMH repose sur une communication de client à serveur. Le client envoie des requêtes au serveur en http, le serveur répond par un flux de données en XML.

N'étant pas du ressort du service Transparence et Open Data, le projet a fait l'objet d'une réflexion en vue de l'automatisation des actions à faire, d'où l'intervention du prestataire OpenDataSoft sur les actions plus développées. Pour réaliser le connecteur OAI-PMH, opération purement technique, un point API sera conçu selon les types de sources Gallica (on en compte 9 au total dans le menu déroulant à gauche de la recherche sur la page d'accueil), en attente des correspondances avec les 84 catégories de la typologie SNCF (Cf. **annexe 4**). Depuis le jeu existant, l'idée serait de dupliquer la colonne typologie et de remplacer les catégories par leur correspondante BNF-Gallica mais cela implique 84 processeurs « remplacer le texte ». Il semblerait plus simple de télécharger le fichier en csv, de modifier la donnée source dans Excel par exemple en ajoutant la colonne des catégories Gallica, puis de remplacer la source par ce nouveau jeu et de sélectionner en facette cette nouvelle colonne.

⁷ http://www.bnf.fr/documents/Guide_oaipmh.pdf

- le lancement des tests et la mise en production (À voir avec ODS)

Une fois le fichier de correspondances établi, la tâche revenait à ODS de faire des tests de conversion des données pour faire correspondre le format à celui requis. L'état d'avancement du livrable lorsque nous quittions l'entreprise était estimé à 40 %.

3.3. Génération des graphiques d'analyse statistique, et décryptage des données

La plupart des plateformes open data pour valoriser leurs données utilisent des graphiques d'analyse et des cartes statiques. La consultation des graphiques permet de visualiser rapidement l'information ou les informations principales que peuvent révéler un jeu de données, de faciliter la comparaison des chiffres s'il y a lieu et d'envisager les possibilités de réutilisation en fonction de la nature des données. Les représentations graphiques ont ceci de particulier qu'elles peuvent influencer sur l'efficacité et la crédibilité du message, dans la perspective de communication au grand public, elles restent une simplification de la réalité.

a. Objectif

Optimiser l'analyse et valoriser le contenu des jeux de données

b. Mise en œuvre:

Mettre en place des actions selon les objectifs spécifiés a nécessité de passer par bon nombre d'étapes :

- L'exploration du catalogue de données pour déterminer dans un premier temps les données les plus importantes et pour lesquelles un décryptage pourrait être utile.
- Ensuite parmi ces données le repérage à l'aide de la fonction Statistiques les jeux de données les plus consultés
- La vérification et l'ajustement des éléments souches du graphique d'analyse
- L'élaboration des vues personnalisées pour certains jeux de données

c. Résultats

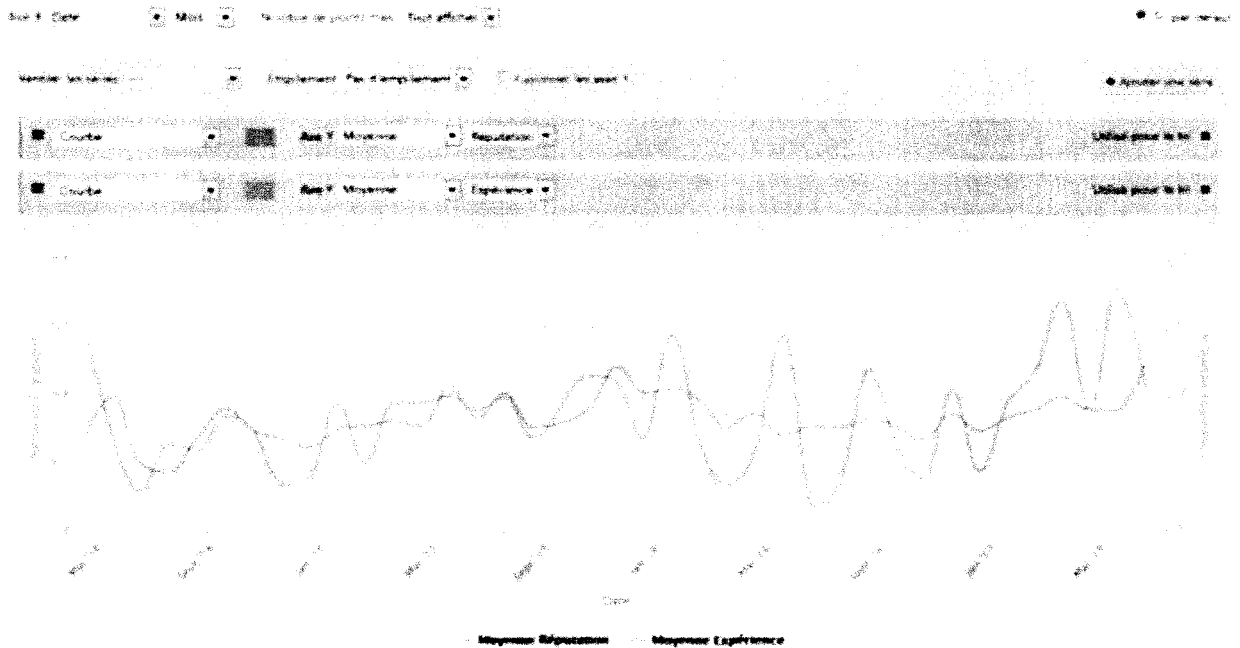


Figure 6: Représentation graphique du jeu de données “Baromètre d’opinions SNCF”

Procès-verbal de recensement des travaux d'installation d'un dortoir et d'un réfectoire pour femmes pilotes en gare de Paris-Ivry

1961/02/1916

Ministère des Travaux Publics, des Postes et des Télégraphes, Contrôle et surveillance des Chemins de fer, Réseau d'Orléans

Le projet de construction d'un dortoir et d'un réfectoire, pour les femmes nouvellement embauchées au sein de la Compagnie des Chemins de fer de Paris à Orléans, ayant été approuvé par décision ministérielle.

Description technique : A4, papier, 3 pages

[ACCÈS EN LIGNE](#) [L'ÉTAT](#) [TOUTES LES DONNÉES](#)

[CONSULTER L'ARCHIVE](#)

Figure 7: Exemple de notice élaborée à partir des données des archives historiques

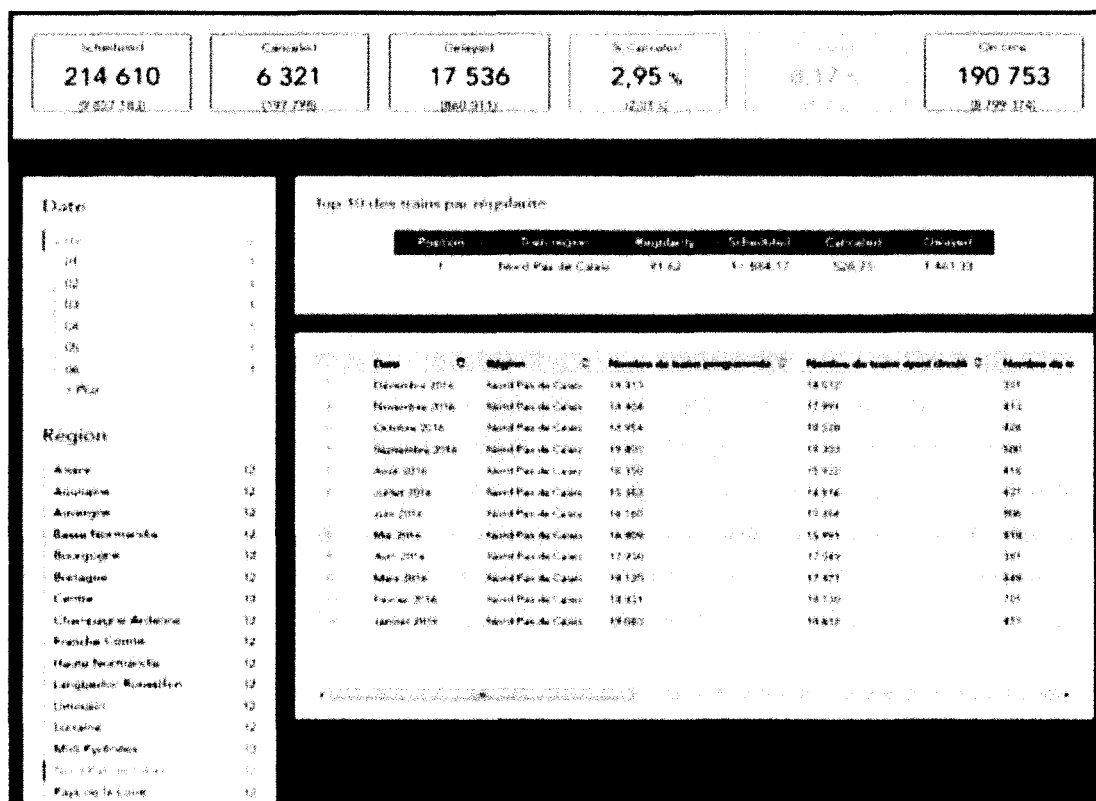


Figure 8: Analyse des données de régularité TER

3.4. Activité de reporting: DÉFINIR LES MÉTRIQUES DE SUIVI DES USAGES et création des dashboard

Qui sont les personnes qui utilisent les données et avec quelles intensités ? Quelle est la répartition de la popularité parmi nos jeux de données ? Que cherchent à faire les utilisateurs avec les données ? Ce sont des questions qui sont posées bien régulièrement par les producteurs des données.

Il est possible aujourd’hui de consulter les statistiques et de définir les métriques de suivi des usages dans le back-office de la plateforme mais pour l’ensemble des jeux de données. Mais les demandes reçues très souvent sont distinctes et particulières. Chaque acteur souhaiterait connaître pour chaque donnée qu’il publie les statistiques d’usage (consultations simples, appels API et téléchargements) avec des indicateurs précis.

Quelques indicateurs ont été définis pour analyser l'usage des données, entre autres:

- **l'activité des jeux de données:** Nombre ou moyenne mensuelle de téléchargements et d'appels API des jeux de données les plus utilisés.
- **Jeux de données les plus populaires :** Les 5 ou 10 jeux de données ayant reçu le plus d'appels API sont considérés comme étant les jeux de données moteurs du public cible.
- **Nombre d'utilisateurs :** Ensemble d'utilisateurs ayant interagi avec le jeu de données.

Des tableaux de bord ont été conçus à base de code HTML et CSS (cf.annexe 5) pour analyser les usages des données de SNCF Réseau et de Gares & Connexions.

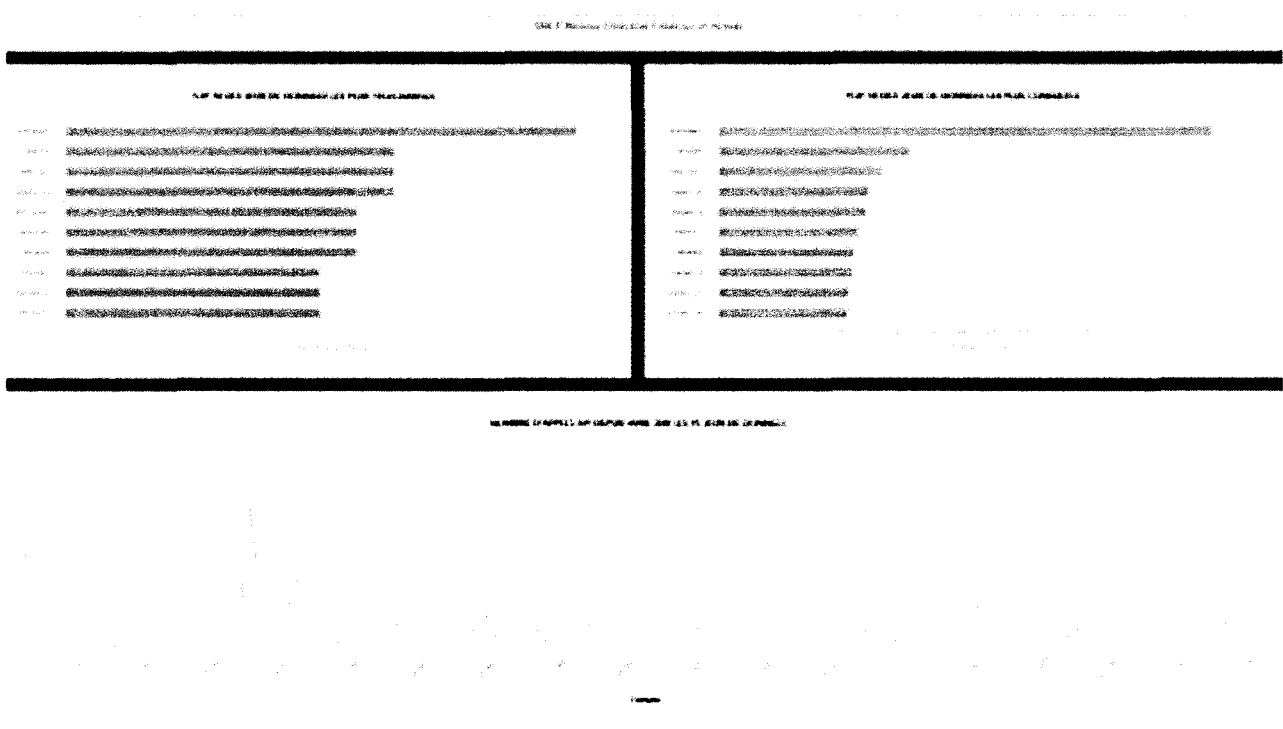


Figure 9: Extrait du tableau de bord / Analyse des usages des données SNCF Réseau

Chapitre 3 : Discussion du travail réalisé

Aucune entreprise, quels que soient la taille et son type d'activité ne peut fonctionner sans information et communication. Dans le contexte de l'ouverture des données, du fait que la valeur de la donnée croît exponentiellement avec la création d'un écosystème au sein duquel de nombreux flux de données se combinent, avec pour conséquence le besoin d'accord de partage de données entre différents acteurs, parfois concurrents. Il est aussi vital que cette ouverture soit encadrée et administrée, que le portail soit à jour, intégrant les dernières actualisations et évolutions, tout en éliminant les éléments obsolètes ou ne présentant plus aucun intérêt. En résumé, les données doivent être à jour pour prétendre à une exploitation et à la création de services secondaires fiables. En étant ainsi mises à jour, les "Data" se défont des informations erronées. Ce qui réduit fortement les risques de commettre des erreurs, de transmettre des données fausses aux nouveaux collaborateurs et, de manière plus générale, de se retrouver avec des informations vides de sens. En cela, l'administration fonctionnelle du portail open data joue un rôle déterminant dans l'amélioration continue, puisqu'elle s'attaque aux facteurs pouvant affecter la qualité.

Pour mener à bien ces missions assez contraignantes sur la durée du contrat, il a fallu une organisation assez méthodique et rigoureuse. Néanmoins, nous pouvons nous réjouir de l'aboutissement.

La réalisation des missions sus-décrites fera certainement avancer le service dans ses objectifs de performance. Toutefois, le suivi et l'aboutissement des projets tout au long de la période de professionnalisation ont entraîné de nouvelles visions. Outre la démarche de valorisation des données par leur publication sur le portail, la démarche open data peut être revue au sens d'organisation des événements de type hackathon⁸ ou open data dans le but de dynamiser les données. Aller au-delà de la simple publication, développer un écosystème à travers des animations, à travers des relations qui se mettent en place. Les concours, les appels à projet sont un moyen d'inciter à la réutilisation et d'assurer le développement. Des événements qui favorisent la créativité numérique et une acculturation à l'open data. Les hackathons réunissent

⁸ <https://fr.wikipedia.org/wiki/Hackathon>

des équipes pluridisciplinaires de développeurs, porteurs de projets, graphistes et autres profils en fonction de la thématique de l'évènement autour d'un objectif : réaliser un prototype d'application à partir des jeux de données publics. La difficulté aujourd'hui ce n'est pas de créer de l'information mais d'animer, de toucher des cibles, parce qu'elles sont éparses.

La mise en œuvre nécessiterait donc une réflexion basée sur différents scénarii de réalisation, d'abord au niveau planification, ensuite fonctionnel, entre autres:

- Préciser les objectifs et enjeux de l'évènement
- Définir les règles du jeu
- Choisir l'espace adéquat pour la phase avant et post-concours
- Choisir une récompense
- Assurer la communication régulière pour être sûr de toucher un large public
- Constituer une équipe de votants ou alors le jury
- Penser à la phase de clôture de l'édition (célébration de l'équipe gagnante, cérémonie de fermeture...)

L'aboutissement du projet proposé passera notamment par l'évaluation de la pertinence des scénarios proposés et par l'application de ceux-ci.

Dans cette perspective, les objectifs visés permettront d'apporter davantage une plus-value à l'open data SNCF. Les données constituent, et vont constituer davantage encore durant la décennie à venir, un levier essentiel de la transformation numérique des entreprises et de leurs conditions de création de valeur.

Conclusion

Les premières semaines de notre parcours de professionnalisation nous ont paru un peu difficiles dès lors qu'il nous fallait nous familiariser avec un nouvel environnement qui impose ses contraintes, et nous intégrer dans une équipe nouvelle.

En effet, la ville de Paris qui nous a accueilli n'était pas notre résidence habituelle, ajouté au fait que la SNCF était une entreprise que nous connaissions encore peu. Or, il s'avère que pour traiter des données spécifiques, il est judicieux de maîtriser d'entrée de jeu, le cadre dans lequel les techniques professionnelles vont être déployées.

Au terme d'un séjour de 12 mois passés au service Transparence & Open data de la SNCF, nous avons pu mettre en pratique des connaissances théoriques acquises durant deux années de dur labeur passées à l'Université de Lille 3. En plus de nous confronter aux difficultés réelles et incontournables du monde du travail et de la gestion de projet. L'occasion nous a été donnée de réaliser plusieurs tâches qui ont constitué une mission globale de stage. Chacune de ces tâches, s'est en effet inscrite dans la stratégie open data définie par l'entreprise. Nous gardons de ces moments, un excellent souvenir qui constitue désormais une expérience professionnelle valorisante et encourageante pour notre avenir. Ce stage de professionnalisation nous aura permis d'apprendre à gérer le stress dans des situations complexes et d'accroître notre faculté de polyvalence.

Nous pensons que cette expérience en entreprise nous a offert une bonne préparation à l'insertion professionnelle car elle fut pour nous une expérience enrichissante et complète qui conforte notre désir d'exercer dans le domaine de la gestion des flux d'information. Nous n'avons véritablement pas éprouvé de difficultés particulières concernant le travail effectué, dès lors que le personnel de l'entreprise et plus particulièrement le Directeur de la transparence et de l'open data, nous a parfaitement intégrés dans l'environnement professionnel. Nous avons pu travailler dans des conditions matérielles idoines.

Qu'il nous soit permis de déclarer sans risques de nous tromper, que riche d'enseignement, notre stage a largement comblé nos attentes, si l'on s'en tient à la confrontation du bilan évoqué plus haut aux objectifs de stage fixés par Lille 3. Nous aurions en toute honnêteté

aimé que notre séjour durât plus longtemps au sein de la SNCF pour améliorer encore selon les manquements observés certaines aptitudes et parfaire encore notre sens du professionnalisme.

En définitive, nous retiendrons de ces missions que l'open data croît de plus belle, et que le processus de valorisation des données en entreprise concerne tant les acteurs que les utilisateurs visés. En plus de cette valorisation, nous avons observé l'aspect fréquentation mais nous n'avons pas effectué d'étude d'usages, qui semble importante pour évaluer tous les contours de la réutilisation ou les cas de réutilisations de ces données. Les notions d'usage et de fréquentation ne se recouvrent pas totalement. Ce sont deux notions complémentaires, deux manières de traiter la question des publics. L'étude des usages apparaît comme un sous-ensemble de celle de la fréquentation. Au-delà de la masse de données disponibles, la véritable nouveauté réside dans la démultiplication des potentiels de réutilisation et de croisement de données, dans cette perspective de création de valeur. De plus, dans un monde en pleine accélération, l'usage en temps réel ou non des données devient une variable essentielle de la prise de décision opérationnelle⁹.

⁹ <http://www.cigref.fr/wp/wp-content/uploads/2016/11/CIGREF-Valorisation-des-donnees-Pratiques-Modele-2016.pdf>

DEUXIEME PARTIE: RAPPORT DE RECHERCHE

L'OPEN DATA DES ENTREPRISES: QUELS USAGES ?

Introduction

Les données numériques se veulent de plus en plus présentes dans divers domaines d'activités du fait qu'elles constituent des briques de base de la société de l'information. Leur création et croissance exponentielles débouchent sur de grands entrepôts, elles sont générées tant par les administrations publiques que par les entreprises privées ou encore par les établissements ou associations à visée scientifique. Le croisement des données et leur analyse peuvent permettre à chacun de contribuer d'une manière ou d'une autre à la vie publique. Nous avons observé que l'open data est un enjeu pour l'usage, à travers l'open data des entreprises, nous découvrons qu'on va au-delà de l'open data administratif, gouvernemental, où on peut comprendre que l'intérêt public est mis en avant par la mise à disposition à tous des données. L'aspect juridique n'est pas en reste, ces appréhensions reflètent une problématique globale. Des gisements d'entreprises sont aussi considérés d'intérêt public et c'est une nouvelle fonction que se donnent les entreprises, qui est de s'intéresser au caractère public que peuvent revêtir leurs données, de considérer que leurs informations sont des biens publics. On passe d'une étape où c'était essentiellement l'état, l'administration (notion de bien public) à une imprégnation progressive des entreprises. A quels usages sont destinées les données? Quelle est la valeur ajoutée?

Les usages font polémique, nombreux sont ceux qui considèrent que ce sont des données publiques qui n'ont pas à être privatisées, pour d'autres, le fait que les données d'entreprises soient publiques permet d'encourager l'innovation et d'informer au mieux sur les missions du service public.

Cerner ces nouvelles allures technologiques, ou mieux les contours de l'assimilation des Nouvelles Technologies de l'Information et de la Communication (NTIC) est une démarche cruciale dans les recherches engagées au croisement de l'informatique et des Sciences Humaines et Sociales. Elle avise d'une certaine manière l'adaptabilité des technologies au contexte sans cesse croissant que revêtent les activités en appréhendant les usages et innovations qui peuvent conduire à des changements sociétaux majeurs. Les méthodes d'analyse d'usage sont employées pour tenter d'évaluer l'utilisabilité et l'utilité des outils conçus à court et à long terme.

L'usage est aussi un indicateur important pour les SHS dans la mesure où il se montre déterminant dans la compréhension des étapes de mise en place des innovations technologiques et des construits culturels au niveau micro et macrosocial.

Cinq (5) objectifs de l'analyse des usages ont été spécifiés par Jauréguiberry & Proulx en 2011 dans leur ouvrage "Usages et enjeux des technologies de l'information"¹⁰:

- l'analyse de l'interaction dialogique entre l'utilisateur et le dispositif technique.
- l'analyse de la coordination entre l'usager et le concepteur du dispositif.
- l'analyse de la situation de l'usage dans un contexte de pratiques.
- l'inscription des dimensions politiques et morales dans la conception de l'objet technique et dans la configuration de l'usager.
- l'ancrage social et historique des usages dans un ensemble de macrostructures.

Afin d'appréhender au mieux les contours du sujet, n'est-il pas judicieux de faire l'état des lieux de la recherche à ce sujet pour en définir de manière précise l'objet et le cadre théorique dans lequel elle s'inscrit ? Par ailleurs, ne faudrait-il pas également effectuer une description de la problématique et des questionnements qui sous-tendent ce projet, sans oublier bien évidemment, les axes de recherche, le dispositif méthodologique devant permettre de le mener à bon port et enfin les intérêts qu'il présente pour les entreprises qui se lancent dans le mouvement d'ouverture des données aussi pour le monde de la recherche scientifique ?

Notre travail s'inscrit dans un premier niveau d'études consistant à faire un état de l'art analytique. Nous examinerons dans ce mémoire comment la question de l'usage des données ouvertes est abordée dans la littérature autrement que par les seules techniques de visualisation, quelle est la valeur ajoutée apportée par l'exploitation des données ouvertes.

¹⁰ <https://www.cairn.info/usages-et-enjeux-des-technologies-de-communication--9782749214405-p-32.htm>

I- Définition de la problématique et présentation de son positionnement dans la littérature

Toute donnée lorsqu'elle est mise à disposition sur une quelconque plateforme de diffusion, fournie par un établissement public ou privé est considérée comme publique. Comme définition claire de la donnée publique, nous dirons qu'il s'agit de "Toute donnée collectée dans le cadre d'une mission de service public. Elle concerne toutes les thématiques d'action importantes telles que: le tourisme, l'éducation, les finances et le transport...".

A part les infographes et les journalistes, l'Open data intéresse potentiellement tout le monde, du chercheur qui veut collecter des sources pour sa recherche jusqu'au journaliste qui va pouvoir accéder à certaines données, en faire de l'exploitation directe qui va déboucher sur des visualisations, les développeurs qui vont pouvoir concevoir des applications ou des associations et habitants qui souhaiteraient être mis au courant des projets en cours dans leur ville (infrastructures, services, qualité de l'air et bien d'autres...). Clairement, l'open data intéresse ceux qui aimeraient avoir l'information et révèle un potentiel économique non négligeable.

1.1. Panorama de la littérature scientifique

Les usages de l'open data sont étudiés dans la littérature scientifique sous plusieurs angles, à travers les notions d'innovation, d'amélioration des services via des flux ascendants et descendants et enfin d'enjeux économiques.

Le phénomène d'ouverture des données tend à se généraliser. Les entreprises s'impliquent de plus en plus dans la démarche, en agissant soit en délégation de service public soumises de ce fait aux obligations légales d'ouverture des données (Suez environnement par exemple...), soit volontairement après analyse des opportunités y découlant (SNCF).

L'open data est une "philosophie ouverte des ressources" (Noyer, Carmes, 2013, p.137-138), une approche innovante de communication pour les entreprises car il ne s'agit pas seulement de médiatiser les données mais de les présenter dans le but de les valoriser et espérer un retour utilisateur sous forme de valeur ajoutée pour faire évoluer les formats produits ou certaines procédures en interne impliquant plusieurs acteurs.

Elle est perçue au vu des possibilités de réutilisations comme substance du crowdsourcing, une forme d'externalisation voire de collaboration possible avec des individus à

l'extérieur de l'entreprise. Il constitue une alternative aux autres formes de production comme la réalisation en interne, l'externalisation classique, ou l'innovation en réseaux¹¹ (Pisano et Verganti, 2008). L'un des objectifs principaux des données ouvertes est la réutilisation. Une possibilité pour les développeurs informatiques et autres citoyens de s'inventer de nouveaux usages. Divers et variés, ils sont fonction du réutilisateur. On assiste donc à un engrenage technologique attisant ainsi fortement les attentes et les créations en matière d'innovation.

L'entreprise met à disposition de tout éventuel usager des données, on peut distinguer à partir des usages les experts des amateurs. Les experts forment des acteurs majeurs dans la communauté de réutilisateurs, puisqu'ils sortent du cadre de la consultation simple et se démarquent par leur capacité à développer de nouveaux produits qui peuvent être intégrés dans l'offre des services de l'entreprise s'il y a du succès et mener même inconsciemment à des idées d'amélioration.

Lobre Katia dans son article "Open data: nouvelle pratique managériale" paru en 2012 donne une définition des réutilisateurs, "ce sont des personnes physiques ou morales, qui développent de nouveaux usages à partir des données ouvertes". Ils referment le cycle de distribution des données ouvertes¹². Il peut s'agir d'un simple citoyen qui souhaiterait obtenir dans le cadre des programmes transparence des données sur la gestion ou la gouvernance des affaires publiques. D'autres acteurs peuvent également être intéressés dans un cadre plus entrepreneurial, pour créer de nouveaux produits et générer de la valeur économique autour de la donnée. On observe cependant deux niveaux d'usages.

La réutilisation des données ouvertes est cependant ouverte à tout le monde quel que soit le but recherché. La relation entre le réutilisateur et les données est en quelque sorte ambiguë, il se retrouve confronté à certaines contraintes, les mêmes parfois que rencontrent les producteurs (Licences - richesse - fiabilité et exhaustivité de la donnée...). Tout cela dépend des buts recherchés (lancement des applications spécifiques à des jeux de données unitaires ou croisement de plusieurs sources de données) mais aussi de l'échelle de «l'industrialisation» des solutions qu'il développe ou propose.

¹¹ <https://hbr.org/2008/12/which-kind-of-collaboration-is-right-for-you>

¹²

https://www.ish-lyon.cnrs.fr/sites/www.ish-lyon.cnrs.fr/files/page/fichier/ISH_201506_IRF_LivreBlancDonneesOuvertes_BMeszaros.pdf

Romain Lacombe & Al distinguent dans leur étude¹³ cinq grands acteurs de la réutilisation des données ouvertes, parmi lesquelles les citoyens, les chercheurs, les acteurs associatifs, les acteurs économiques, la presse et médias. On leur attribue donc des usages divers¹⁴, entre autres,

- Applications internet / Mobile : La partie la plus visible du grand public

Il s'agit d'un quelconque logiciel dont les automatismes sont destinés à assister un utilisateur dans une de ses activités. Issu généralement de la transformation et le jumelage d'informations partagées, il permet généralement la résolution d'un problème spécifique.

- Consultation ANALYTICS : Aide à la prise de décision (évaluation, tableau de bord...)
- MÉDIATION (Education) : Visualisation de données complexes pour une meilleure compréhension, le fait de « Faire parler les données »
- Réutilisations spécialisées : Usages spécifiques dans un cadre personnel ou pour des organisations

Les réutilisateurs de manière générale traduisent ces données partagées sous forme brute en un énoncé compréhensible dans les services ou applications. Dans le cadre des données produites par la SNCF par exemple, de nouvelles ressources sont créées par des "experts" à partir d'agrégats de données pour faciliter la mobilité des citoyens. A leur tour, les citoyens intègrent le processus via l'analyse de leurs retours utilisateurs ou l'analyse de leurs usages effectifs. Une analyse brève a été faite sur les motivations communautaires liées aux usages de l'open data, il semblerait selon Bruno Asdourian et Virginie Zimmerli que l'aspect communautaire est au centre des motivations des experts qui contribuent à la création d'idées. Ces motivations peuvent être à la fois intrinsèques et extrinsèques. Les éléments qui les poussent à participer peuvent être liés au capital culturel (esprit geek et apprentissage des compétences transversales) d'une part et au capital social d'autre part (communautarisme, organisation en réseau des acteurs). Les motivations extrinsèques quant à elles sont plutôt en lien avec la gestion de carrière, les récompenses ou la reconnaissance par l'entreprise et par ses pairs. Quant aux citoyens, leurs motivations envers l'usage des outils numériques proposés sont principalement orientées vers les aspects récréatifs, utilitaires ou en lien avec l'aspect culturel fort des outils libres.

¹³ <http://www.enssib.fr/bibliotheque-numerique/documents/57094-pour-une-politique-ambitieuse-des-donnees-publiques.pdf>

¹⁴ http://www.hauts-de-seine.fr/fileadmin/PDF/Cadredevie/Gestion_de_l_environnement/C2D/C2D_InterventionFrancoisLEPIVAIN.pdf, page 8.

Toutefois, il y a de toute évidence des trésors d'efficacité dans l'usage des données ouvertes par les entreprises. L'ouverture peut être également un stimulus puissant pour l'économie. Il existe plusieurs manières de partager des données gratuites, mais toutes partent de l'idée qu'elles produiront de la valeur grâce aux réutilisateurs.

L'avènement de l'open data constitue sans conteste, l'une des innovations majeures de ces dix dernières années. L'ouverture des données rencontre aujourd'hui un intérêt de plus en plus large. De nombreux acteurs lui attribuent des enjeux économiques. On constate un assouplissement des frontières entre les entreprises et le public¹⁵ (B Asdourian, 2015). En effet, les informations autrefois réservées uniquement qu'à l'interne sont rendues disponibles au grand public, spécifiées pour la plupart comme "professionnels-amateurs". Ce faisant, les entreprises escomptent un retour sous forme de valeur ajoutée.

Ce mouvement occasionne la création de nouveaux produits et services afin d'accroître l'innovation et également d'élargir les marchés externes d'utilisation des innovations. Ainsi, alors que l'innovation était autrefois l'apanage des départements de R&D, certaines entreprises se rendent compte aujourd'hui de l'importance de la mise à disposition des données à un plus large public avec pour conséquence directe et positive l'interaction et la collaboration à partir des créations découlant de l'exploitation des données.

Nous pouvons donc dire que l'open data actionne la création de valeur à partir du brassage des usages collectifs.

Le site pro-01.net évoquerait un marché de l'open data en France à près de 50 milliards d'euros (Alain Van Cuyck, 2014)¹⁶. Convaincus du pouvoir d'aide à la prise de décision que pourrait avoir l'open data, les entreprises s'insèrent de plus en plus dans ce marché aux contours innovants. Leurs usages apporteraient des plus-values importantes notamment en terme de veille, de stratégies, de tendances du marché, de géolocalisation des données, de stratégies de proximité et de services. L'open data pourrait ainsi générer d'importantes retombées économiques sur le marché de l'information,

¹⁵ <https://rfsic.revues.org/1316>

¹⁶ <https://hal.archives-ouvertes.fr/hal-00963337/document>

Toutefois le modèle économique n'est pas clairement fixé, les licences aussi sur le plan juridique. La mise en place des services de données ouvertes est un important marché économique pour l'entreprise et ses acteurs en termes d'usages, d'aide à la prise de décision, et au développement innovant.

1.2. Panorama de la littérature professionnelle

La littérature professionnelle quant à elle apporte un éclairage sur la valeur des usages possibles de l'open data pour l'entreprise en tant que consommatrice et productrice.

La plupart des entreprises suivent une démarche volontariste d'ouverture de leurs données et considèrent que cela s'inscrit purement dans une stratégie d'innovation. Les données ouvertes servent d'abord à améliorer les procédures en interne. Le bon déroulé d'une démarche open data implique un maillage d'acteurs divers pour assurer le partage politique, l'assurance juridique, l'expertise technique, la prise au sérieux hiérarchique...etc. L'aspect marketing n'est pas en reste puisque le but ultime pour certains reste la satisfaction des clients¹⁷. Les réutilisations sont généralement accessoires aux offres que l'entreprise propose en amont. Bien que le principe de l'open data est par définition la "gratuité", les jeux de données peuvent être monétisés. *"Si la commercialisation de données, dont certaines entreprises sont spécialistes, ne s'inscrit pas à proprement parler dans le mouvement Open Data, c'est une opportunité que l'entreprise ne peut négliger"* (Bluenove, 2014).

Le rapport Bluenove sur l'ouverture des données des entreprises recense comme usages principaux:

- Le développement des applications
- L'accentuation du trafic internet
- L'interprétation des données au moyen des visualisations

Si les usages de l'opendata dans les sphères publiques et parapubliques ne font plus de mystère, les usages que l'entreprise peut avoir de l'open data sont plus obscurs¹⁸ (Yannick Thebault, 2015). Il existe de nombreux usages possibles des données ouvertes d'entreprises, les usages "consommateurs" étant dissociés des usages "producteurs".

¹⁷ http://www.bluenove.com/bluenove/wp-content/uploads/2014/06/Livre_blanc_bluenove_-_Open_Data_pour_lentreprise.pdf, page 27

¹⁸ <http://www.astrakhan.fr/LeLab/article/Id/OpenData---quels-usages-pour-l-entreprise---9>

Quelques axes pour les usages sont répertoriés, entre autres marketing, datavisualisation, recherche, économie, culture et technique.

L'usage d'enrichissement de la recherche consiste pour une entreprise fournissant un service de recherche à ses partenaires qu'ils soient clients, fournisseurs ou employés à augmenter la pertinence du résultat avec des données ouvertes.

La Datavisualisation (ou Dataviz) est selon wikipedia la « un résumé visuel de données statistiques chiffrées ». Les datavisualisations ne sont pas des données: elles sont des graphiques utilisant les données pour les “mettre en scène” et permettre leur interprétation. En cela, elles sont des créations dérivées basées sur les potentiels d'exploitation de données libérées. Les données peuvent avoir plusieurs représentations selon qu'elles soient temporelles, géolocalisées, proportionnelles, issues d'une distribution statistique, arborescentes, de flux ou de réseaux. Ce mode d'exploitation de la donnée est fortement utilisé par les journalistes et les communicants, c'est un moyen de faire parler les données et de leur attribuer un sens. Dans l'entreprise, la datavisualisation est employée pour représenter les données de la manière la plus simple, le plus souvent des données de reporting pour les métiers finances, ressources humaines et bien d'autres.

L'open data dans un “*esprit marketing*” vise prioritairement l'opportunité, par exemple assembler un certain nombre d'indicateurs qui servirait à mesurer les activités. Dans le domaine marketing, les usages de l'opendata par l'entreprise se concentrent essentiellement autour de l'opportunité à disposer d'indicateurs complémentaires à ceux déjà utilisés tels que des données sociodémographiques, de trafic routier ou de transport, de délinquance ou de classification de la concurrence (Yannick Thebault, 2015).

La réutilisation des données publiques présente un potentiel commercial. En 2012, la commission européenne a estimé que le marché de l'information pourrait atteindre les 140 milliards d'euros par an, contre 30 milliards actuellement.

Avec la révolution numérique, les données prennent par ailleurs une place centrale dans l'économie. Ouvrir et partager les données publiques, c'est organiser la mise en ligne de données essentielles, qui vont enrichir les analyses de nombreux décideurs, permettre de nombreuses économies de temps de travail ou permettre des prises de décisions mieux informées. C'est créer de grands référentiels partagés par tous les acteurs et encourager le développement de nombreux services à forte valeur ajoutée. C'est donc à la fois une stratégie de souveraineté et un fort levier

de développement économique. <https://www.coheris.com/relation-client-data/solutions/analytics/open-data-quels-impacts-pour-lentreprise-et-sa-bi/>

1.3. Observations InSitu

Pour compléter cette phase exploratoire, nous avons également observé des usages très divers, lors du dernier Barcamp¹⁹ organisé sur le projet de création d'une plateforme open data des transports, beaucoup parlaient des possibilités de création de nouvelles applications combinées et sites web. Nous avons voulu savoir en quoi consistaient l'usage et les perspectives d'usage de ces données. Le sujet est important à étudier car il ne suffit pas seulement de dire qu'on va ouvrir les données mais il convient de savoir à quelles fins et à qui elles devront servir. Des start-up ? Des personnes qui souhaiteraient monter des affaires, ou commercialiser les données (revendre les données ?).

II- Pistes à étudier pour une recherche approfondie du sujet + méthodologie d'étude

Cette analyse de la littérature montre que l'aspect commercialisation des données ou des flux secondaires (données croisées, produit final... etc.) est à creuser, les auteurs font état de ce qui existe actuellement comme usages des données ouvertes d'entreprises mais nous n'en savons pas plus sur ce que produit l'open data en termes de coûts pour les réutilisateurs, ni de quelle manière la technologie open data est appréhendée chez les usagers et selon les usages (qualité de la donnée-exhaustivité...). Par ailleurs, nous avons aussi procédé à des observations In situ. Les grands sujets qui semblent intéressants du point de vue de la littérature sont l'appréhension de la technologie open data et l'assimilation des usages, Du point de vue de nos observations, l'aspect motivations des usagers est aussi à creuser.

On pourrait compléter cette première phase d'analyse documentaire par une étude de cas. Ce choix méthodologique est justifié par l'ambition de vouloir comprendre, expliquer et avoir une vision holistique de l'ancrage de l'open data chez les usagers. Le choix de la démarche a été fait sur la base du cadre conceptuel qui laisse une marge de liberté quant à l'apparition de nouveaux thèmes ou de nouvelles réalités et sur le caractère exploratoire du terrain de recherche puisque peu d'études approfondissent le sujet.

¹⁹ Journée d'échanges avec les acteurs du numérique et des transports, pour définir la future plateforme des données de mobilité

L'étude de cas en sciences humaines et sociales (SHS), renvoie à une méthode d'investigation à visée d'analyse et de compréhension qui consiste à étudier en détail l'ensemble des caractéristiques d'un problème ou d'un phénomène restreint et précis tel qu'il s'est déroulé dans une situation particulière, réelle ou reconstituée, jugée représentative de l'objet à étudier²⁰. À caractère exploratoire, cette approche pourrait porter sur des usages ou non-usages des données, des comportements, des systèmes de représentation des individus par rapport à un lieu, une institution, une profession... Les résultats qui découlent de cette méthode sont des données textuelles à caractère qualitatif (données d'observation, entretiens, questions ouvertes longues des questionnaires). Nous pourrions chercher à comprendre pourquoi telle ou telle catégorie d'individus utilise tel ou tel type de données, et ce que cela révèle en terme de représentations qu'il s'agisse des relations sociales, des pratiques culturelles ou encore de facteurs personnels. Le périmètre de la recherche sera limité à quelques entreprises lancées à ce jour dans la démarche d'ouverture des données. Ces entreprises seront choisies en fonction de la spécificité ou de la particularité des données qu'elles donnent à consulter. Les terrains plus favorables pour ce type de recherche seraient les hackathon, les open lab...Etc.

Cette approche de recherche pourrait également comporter un aspect quantitatif, le but étant de chercher à mesurer des différences entre les types de réutilisateurs, à travers des statistiques de fréquentation (nombre de visites), d'actions effectuées sur les ressources (types de données recherchés, thématiques prioritaires...) ou de satisfaction (échelles...). Les études quantitatives s'appuient généralement sur des questionnaires passés auprès d'échantillons représentatifs d'une population avec des caractéristiques bien définies. Les données chiffrées obtenues à l'issue de l'étude quantitative permettront de mesurer, d'interpréter des écarts, de rechercher des attractions et des répulsions entre les catégories de réutilisateurs ayant apporté des réponses différentes dans le questionnaire.

²⁰ <http://hal.univ-brest.fr/hal-00579008/document>, page 3

Conclusion

L'open data des entreprises gagne de plus en plus en valeur et permet de nombreux usages. De sa gestion rationnelle dépendra la qualité des services et produits tributaires. Une chaîne de valeur qui définit bien les règles de la mutation digitale observée aujourd'hui. La présente étude avait pour principal objectif une revue et étude documentaire afin de mieux situer l'objet de recherche. L'ensemble des analyses menées a permis de dégager des questions centrales auxquelles une recherche minutieuse tentera d'apporter des réponses.

Néanmoins, nous constatons que notre recherche sur les usages de l'open data se heurte à quelques difficultés, le travail que nous avons mené se limite à une synthèse des travaux existants sur les usages de l'open data. Il faudrait que nous regardions de plus près les ouvrages spécialisés que nous n'avons pas véritablement pu explorer. Toutefois il y a un intérêt à ce travail, c'est justement de constater que la question de l'usage est bien un sujet de recherche, le fait de connaître et comprendre ce qui conduit les usagers dans l'exploitation de cette mine de données ouvertes et comment utilisent-ils ce qui leur est proposé.

Bibliographie

I- Sources primaires

A- Sources documentaires

- SNCF, Les Infos. «SNCF à travers le monde». La Plaine Saint-Denis, 2016.
- SNCF, Direction de la Communication et de l'Information. «Profil groupe 2016». la Plaine Saint-Denis, 2016

B- Sources orales

Personnes ressources	Fonction	Date et lieu de l'entretien
Guillaume LEBORGNE	Directeur de la Transparence et de l'open data	Tout le long du stage
Léa Darmon Raphoz	Chargée de mission transparence et patrimoine	Tout le long du stage
Nicolas Rastouil	Chef de projets SI	Tout le long du stage
Ysée Rogé	Customer Success Manager chez ODS	Tout le long du stage
Benjamin Duhamel	Expert fonctionnel chez BNF	07 juillet 2017
Sonia METCHE	Directrice Etudes chez Kantar Media	15 décembre 2016

C- Textes réglementaires

- Charte des données ouvertes de la SNCF
- LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique. LEGIFRANCE, mis à jour le 23 août 2017. Disponible sur le web <https://www.legifrance.gouv.fr/affichLoiPubliee.do?idDocument=JORFDOLE000031589829&type=general&legislature=14> (consulté le 12 janvier 2017)

II- Sources bibliographiques

A- Ouvrages généraux

- Pierre Colletette, «L'étude de cas au service de la recherche: Méthodologie». Université du Québec, 1997. Disponible sur le web: <http://fulltext.bdsp.ehesp.fr/Rsi/50/81.pdf> (consulté le 26 août 2017)
- Said El Moustafid, «Méthodologie de recherche par étude de cas dans les sciences de gestion au Maroc». Groupe ISCAE, 2014. Disponible sur le web: <http://fulltext.bdsp.ehesp.fr/Rsi/50/81.pdf> (consulté le 26 août 2017)
- SEURRAT, Aude. « Ecrire un mémoire en sciences de l'information et de la communication. Récits de cas, démarches et méthodes ». France : éd. Presses Sorbonne Nouvelle, 1980, extrait Google books.

B- Ouvrages spécialisés

- Bluenove, «Open Data : quels enjeux et opportunités pour l'entreprise ?», Livre blanc, 2011. Disponible sur le web: http://www.bluenove.com/bluenove/wp-content/uploads/2014/06/Livre_blanc_bluenove_-_Open_Data_pour_lentreprise.pdf (consulté le 10 août 2017)
- Christine Michel. «Analyse des usages des plateformes de construction de connaissances par des méthodes mixtes et réflexives pour l'amélioration de l'appropriation et de la structuration de l'information : SECONDE PARTIE : Synthèse et bilan des recherches». Ordinateur et société. Université Lyon 1, 2015. Disponible sur le web: <https://hal.archives-ouvertes.fr/tel-01212203/document> (consulté le 27 juillet 2017)
- Romain Lacombe, pierre-henri Bertin et Al. , «Les données publiques au service de l'innovation et de la transparence». Ecole des ponts ParisTech, 2011. Disponible sur le web: <http://www.enssib.fr/bibliotheque-numerique/notices/57094-pour-une-politique-ambitieuse-des-donnees-publiques> (consulté le 27 juillet 2017)
- Simon CHIGNARD, « L'Open data : Comprendre l'ouverture des données publiques », Entreprendre. Fyp Editions, mars 2012

C- Articles de périodiques et mémoires

- Alain Van Cuyck. «L'open data comme nouvelle forme de gouvernance numérique : enjeux, marches, modèles, idéologies..» XIX colloque international franco-roumain "Culture et Responsabilité sociale dans la communication des organisations", Mar 2013, Bucarest, Roumanie. Disponible sur le web: <https://hal.archives-ouvertes.fr/hal-00963337/document> (consulté le 27 juin 2017)
- ARMAND AMATO Etienne, « Enjeux et opportunités de la data visualisation : interagir avec les données », *I2D-Information, données & documents* 2015/2, (Vol.52), p.34-35. Disponible sur le web: https://www.cairn.info/load_pdf.php?ID_ARTICLE=I2D_152_0034 (consulté le 20 juin 2017)
- Benyayer Louis-David, Chignard Simon, « Focus - Les enjeux économiques de l'ouverture des données : pas de marché, pas de valeur », *Informations sociales*, 2015/5 (n° 191), p. 36-39. Disponible sur le web : <http://www.cairn.info/revue-informations-sociales-2015-5-page-36.htm> (consulté le 03 juillet 2017)
- Bruno Asdourian et Virginie Zimmerli, « *Open data*, usagers du numérique et entreprises publiques », *Revue française des sciences de l'information et de la communication*, 2015. Disponible sur le web : <http://rfsic.revues.org/1316> (consulté le 10 juillet 2017)
- Danièle Bourcier, Primavera De Filippi. «L'Open Data : universalité du principe et diversité des expériences ?». *La semaine juridique - édition générale*, LexisNexis, 2013, p.47. Disponible sur le web: <https://hal.archives-ouvertes.fr/hal-01026106> (consulté le 27 juillet 2017)
- David Gross-Amblard, Hélène Jaudoin, «L'Open Data, l'ouverture des données pour de nouveaux usages». *Interstices.info*, 2015. Disponible sur le web: https://interstices.info/jcms/int_71721/l-open-data-l-ouverture-des-donnees-pour-de-nouveaux-usages (consulté le 27 juillet 2017)
- Franck Debos, « *Open data* et culture de la donnée : le cas *OpeNRJ* », *Revue française des sciences de l'information et de la communication*, 2017. Disponible sur le web: <https://rfsic.revues.org/2639#quotation> (consulté le 12 août 2017)
- Guillaume Leborgne, « Diffuser le patrimoine informationnel d'une entreprise, c'est lui donner de la valeur », *I2D-Information, données & documents* 2017/2, p.12-13

- Henri VERDIER, « Au-delà de l'ouverture des données, ce qui est en jeu, c'est l'ouverture de la décision », *Informations sociales*, 2015/5 (n° 191), p. 20-25. Disponible sur le web : <http://www.cairn.info/revue-informations-sociales-2015-5-page-20.htm> (consulté le 03 juillet 2017)
- Institut régional d'administration, «L'Open Data, un service aux usagers?». 2016. Disponible sur le web: <http://www.ira-nantes.gouv.fr/seminaire/open-data/ii-lopen-data-un-service-aux-usagers/> (consulté le 27 juillet 2017)
- Jean-Michel SALAÜN. «La réutilisation et l'exploitation des données ouvertes favorisent-elles leur économie?», *économie du document*, 2014. Disponible sur le web: <https://archinfo24.hypotheses.org/2426> (consulté le 15 juin 2017)
- Josiane Ducournau, « L'open data: fiche synthétique», *Savoirs CDI*, 2016. Disponible sur le web: <https://www.reseau-canope.fr/savoircdi/societe-de-linformation/tic-et-documentation/veille-technologique/lopen-data-fiche-synthetique.html> (consulté le 03 juillet 2017)
- Katia Lobre, Lebraty Jean-Fabrice, « L'open Data : nouvelle pratique managériale risquée ? », *Gestion 2000*, 2012/4 (Volume 29), p. 103-116. Disponible sur le web : <http://www.cairn.info/revue-gestion-2000-2012-4-page-103.htm> (consulté le 23 juin 2017)
- Scala Dream Team, «7 exemples concrets d'utilisation pour mieux comprendre à quoi sert l'open data», *l'exploitation des données*, 2015. Disponible sur le web: <http://www.groupe-scala.com/open-data-02-7-exemples-concrets-dutilisation-pour-mieux-comprendre-a-quoi-sert-lopen-data/> (consulté le 27 juillet 2017)
- Simon CHIGNARD. «Comment diffuser et valoriser des données d'entreprise?», 2012. Disponible sur le web: <https://donneesouvertes.info/2012/05/14/comment-diffuser-et-valoriser-des-donnees-dentreprises/> (consulté le 15 juin 2017)
- Simon CHIGNARD. «Où est le marché de l'open data», 2012. Disponible sur le web: <https://donneesouvertes.info/2012/03/12/ou-est-le-marche-de-lopen-data/> (consulté le 15 juin 2017)
- Yannick Thebault, «OpenData : quels usages pour l'entreprise ?». Astrakhan innovation management, 2015. Disponible sur le web:

<http://www.astrakhan.fr/LeLab/article/Id/OpenData---quels-usages-pour-l-entreprise---9>
(consulté le 27 juillet 2017)

III- Documents de référence

- CRISCO - dictionnaire électronique des synonymes. Disponible sur le web: <http://www.crisco.unicaen.fr/des/>, Université CAEN Normandie.
- Wiktionnaire - Le dictionnaire libre. Disponible sur le web: https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil, Wikimedia foundation.

IV- Sitographie

- <http://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>
- <https://data.sncf.com/>
- <https://www.opendatasoft.fr/>

Annexes

Annexe 1: Organigramme

Annexe 2: Cartographie OpenStreetMap des gares SNCF Transilien en Ile-de-France

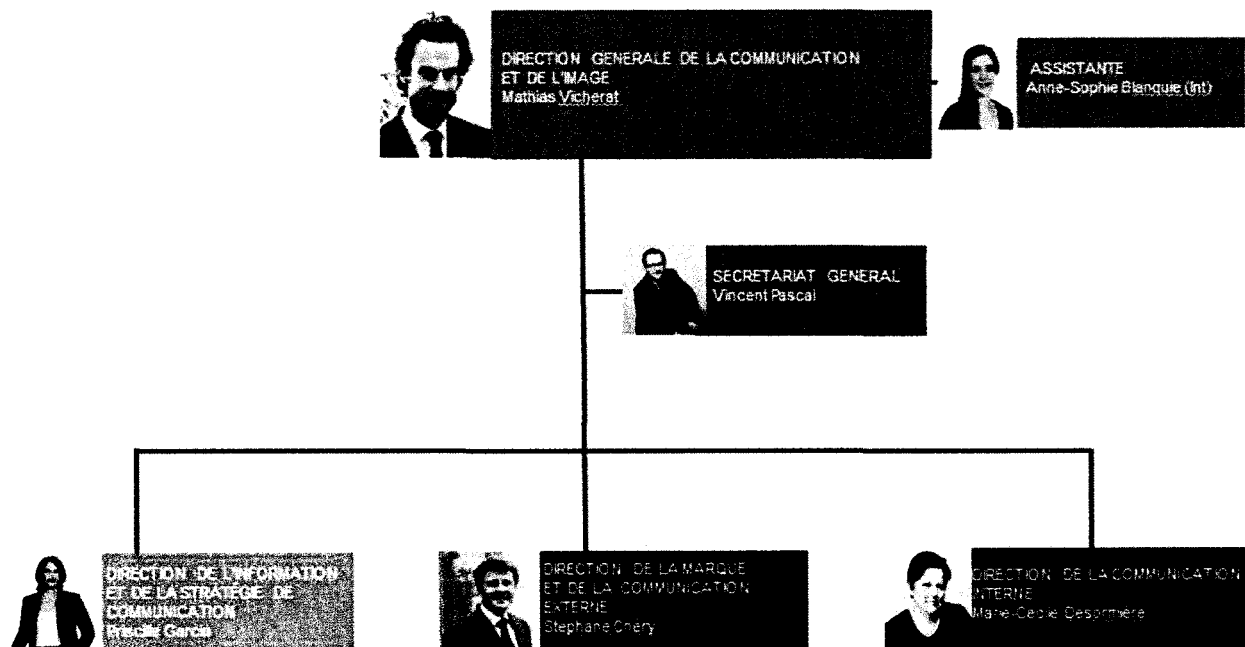
Annexe 3: Charte Open data SNCF

Annexe 4: Tableau de correspondances-catégories Gallica

Annexe 5: Extrait du code de conception des tableaux de bord et graphiques

Annexel

Organigramme

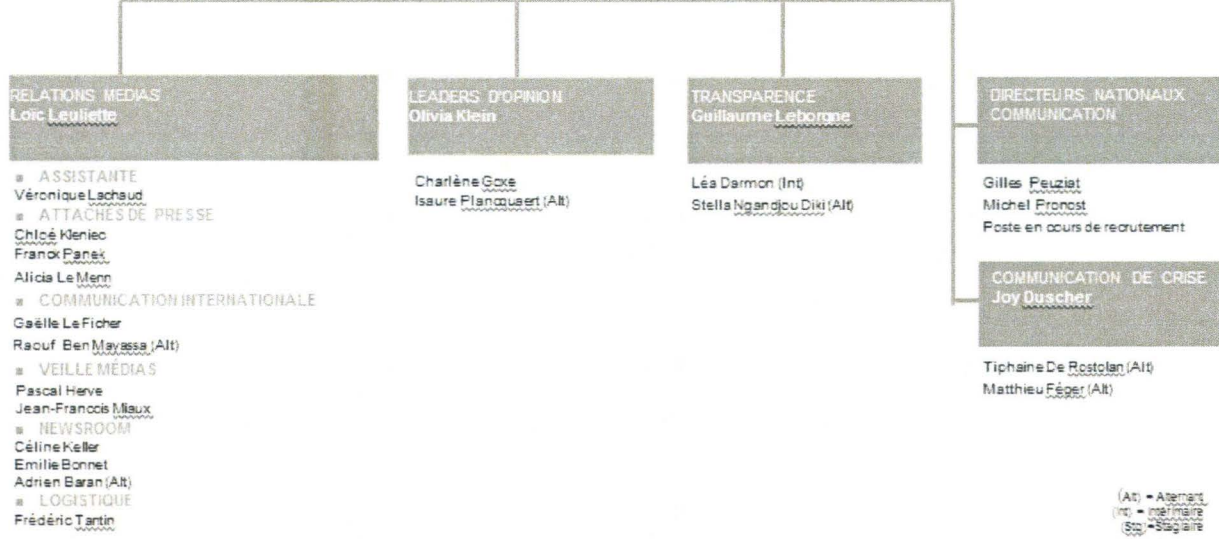


DIRECTION DE L'INFORMATION ET DE LA STRATÉGIE DE COMMUNICATION



DIRECTRICE DE L'INFORMATION ET DE LA STRATÉGIE DE COMMUNICATION
Priscille Garcin

ASSISTANTE
Sabrina Bouzidi-Boucherba

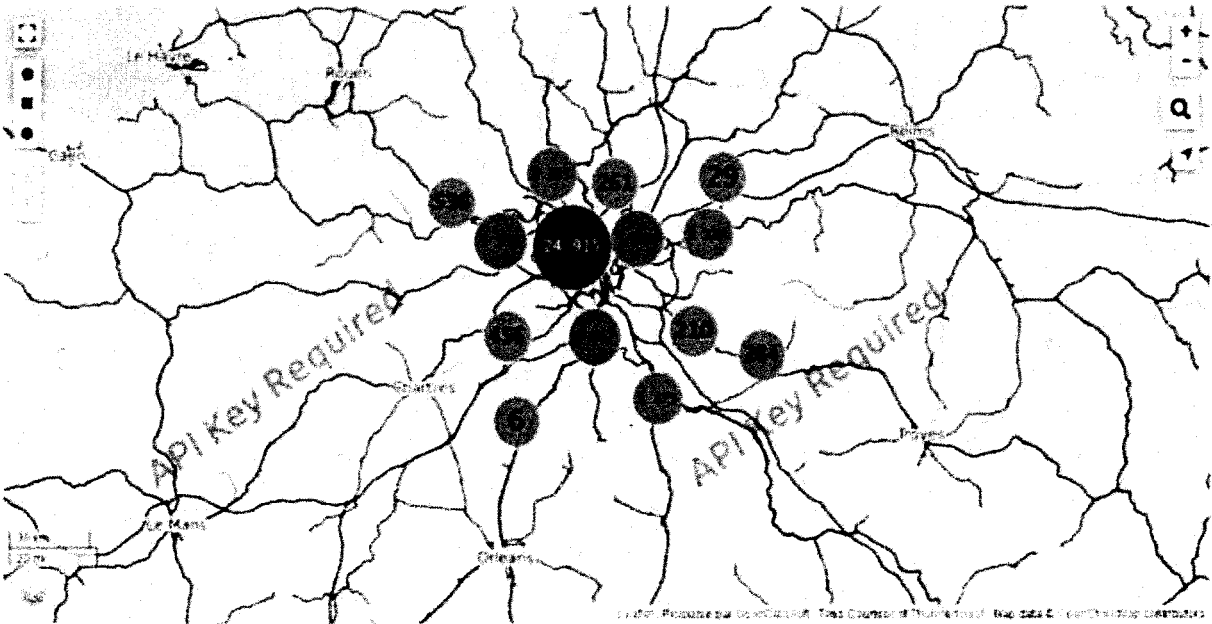


(Alt) - Alternant
(Int) - Intérimaire
(Sp) - Spécialiste
(Ext) - Extérieur

Annexe2

Cartographie OpenStreetMap des gares SNCF

Transilien en Ile-de-France



Annexe3

Charte Open data SNCF

Charte sur la transparence des données SNCF

Préambule

L'accès aux données publiques, favorisé par les outils numériques, est une nouvelle exigence des citoyens vis-à-vis des administrations et entreprises gestionnaires de services publics. Anticipant cette attente devenue loi, le Groupe Public Ferroviaire (désigné comme SNCF dans la suite du document) s'est positionné, en France, comme une entreprise pionnière en matière de transparence en mettant en œuvre dès 2012 une politique d'ouverture des données.

Cette transparence s'incarne à travers 6 principes mis en œuvre au sein de SNCF :

Principe 1 : SNCF a un devoir de redevabilité en matière de données publiques

Dans l'exercice de ses missions de service public, SNCF a pour ambition d'ouvrir le plus grand nombre possible de ses données et documents d'entreprise. SNCF le fait par esprit de redevabilité de grande entreprise publique, pour montrer la réalité de son fonctionnement et partager des ressources utiles au développement d'idées et de nouveaux services par des tiers.

Principe 2 : L'ouverture des données est envisagée par défaut

Toute donnée mérite d'être exposée en interne et/ou en externe. La règle est d'envisager par défaut l'ouverture des données.

Principe 3 : Le degré d'ouverture est fonction de la criticité

L'examen de la criticité des données permet de déterminer le niveau d'ouverture (granularité, périodicité) et éventuellement de justifier leur non publication.

Il existe des motifs légitimes justifiant que certaines données puissent ne pas être diffusées : juridique, financier, sécurité ferroviaire et des systèmes d'information, sûreté des personnes, secret industriel et commercial, et données à caractère personnel.

Dès lors que la donnée est diffusée sur un support électronique par SNCF ou un tiers (Autorités responsables de transports, partenaires) alors la donnée doit être disponible en Open Data

Principe 4 : Le propriétaire interne de la donnée est son producteur.

En interne, le propriétaire de la donnée est son producteur. En accord avec le cadre légal et la politique de Transparence de l'entreprise, il définit le potentiel de valeur de la donnée, ainsi que le niveau de protection associé et donne un avis sur sa criticité. Lorsque la donnée est exposée, il est responsable de sa qualité.

Principe 5 : Les données rendues disponibles sont de qualité

Avant d'être diffusées, les données doivent être mises en qualité et documentées. Elles doivent le rester et faire l'objet de mises à jour régulières.

Principe 6 : Les données sont proposées dans un format ouvert et sous licence

SNCF Open Data

Le format sous lequel est délivrée la donnée doit être ouvert et permettre la réutilisation. Selon le format technique, la licence SNCF Open Data ou les CGU de l'API déterminent les droits et devoirs des réutilisateurs de données. Ces derniers sont libres d'utiliser, de partager et de modifier les données ouvertes par SNCF sous condition de diffuser les modifications apportées aux données dans les mêmes termes.

Annexe4

Tableau de correspondances-catégories Gallica

TYPOLOGIE open Archives	Concordance Gallica
Affiche	images
Album photographique	images
Article	sans objet (archives)
Avis	sans objet (archives)
Avis général	sans objet (archives)
Billet de train	objets
Brevet	sans objet (archives)
Brochure	livres
Bulletin d'information	livres
Cahier des charges	sans objet (archives)
Carte	cartes
Carte, brochure	cartes
Cartes postale	images
Certificat	sans objet (archives)
Circulaire	sans objet (archives)
Communiqué de presse	livres
Compte-rendu	sans objet (archives)
Contrat	sans objet (archives)
Convention	sans objet (archives)
Correspondance	sans objet (archives)
Correspondance ; déclaration ; plan	sans objet (archives)
Correspondance, Note	sans objet (archives)
Correspondance, plan	sans objet (archives)
Débat	sans objet (archives)
Délibération	sans objet (archives)
Dessin	cartes
Devis	sans objet (archives)
Diagramme	diagramme
Discours	sans objet (archives)
Dispositions générales	sans objet (archives)

Documentation	sans objet (archives)
Dossier d'approbation ministérielle	sans objet (archives)
Dossier de presse	livres
Dossier de projet	sans objet (archives)
Exposé	sans objet (archives)
Fiche	sans objet (archives)
Fiche documentaire	livres
Indicateur horaire	livres
Notice, Plan	cartes
Ordre de service	sans objet (archives)
Ordre du jour	sans objet (archives)
Ordre général	sans objet (archives)
Ouvrage	livres
Photographie	images
Photographie, plan	images
Photographie, tableau	images
Photographies	images
Plan	cartes
Plan, Carte, Graphique	cartes
Plaquette	livres
Procès-verbal	sans objet (archives)
Procès-verbal de récolement	sans objet (archives)
Procès-verbal, note, plan	sans objet (archives)
Projet	sans objet (archives)
Protocole	sans objet (archives)
Rapport	sans objet (archives)
Rapport, Plan, Tableau	sans objet (archives)
Registre	sans objet (archives)
Règlementation	sans objet (archives)
Répertoire	sans objet (archives)
son	sons
Tableau	sans objet (archives)
Télégramme	sans objet (archives)
Texte réglementaire	sans objet (archives)
Traité	sans objet (archives)
Traité, Plan	sans objet (archives)
vidéo	images

Annexe5

Extrait du code de conception des tableaux de bord et graphiques

```
1 <div class="container">
2
3 <ods-dataset-context
4     context="regularitemensuelleter,regularitemensuelleterfiltered"
5     regularitemensuelleter-dataset="regularite-mensuelle-ter"
6     regularitemensuelleterfiltered-dataset="regularite-mensuelle-ter">
7
8 <div class="row">
9     <div ods-aggregation="total"
10         ods-aggregation-context="regularitemensuelleter"
11         ods-aggregation-function="SUM"
12         ods-aggregation-expression="nombre_de_trains_programmes">
13     <div ods-aggregation="canceled"
14         ods-aggregation-context="regularitemensuelleter"
15         ods-aggregation-function="SUM"
16         ods-aggregation-expression="nombre_de_trains_annules">
17     <div ods-aggregation="delayed"
18         ods-aggregation-context="regularitemensuelleter"
19         ods-aggregation-function="SUM"
20         ods-aggregation-expression="nombre_de_trains_en_retard_a_1_arrivee">
21
22     <div ods-aggregation="totalfiltered"
23         ods-aggregation-context="regularitemensuelleterfiltered"
24         ods-aggregation-function="SUM"
25         ods-aggregation-expression="nombre_de_trains_programmes">
26     <div ods-aggregation="canceledfiltered"
27         ods-aggregation-context="regularitemensuelleterfiltered"
28         ods-aggregation-function="SUM"
29         ods-aggregation-expression="nombre_de_trains_annules">
30     <div ods-aggregation="delayedfiltered"
31         ods-aggregation-context="regularitemensuelleterfiltered"
32         ods-aggregation-function="SUM"
33         ods-aggregation-expression="nombre_de_trains_en_retard_a_1_arrivee">
34
35
36 <div class="ods-box" >
37     <div class="row">
38         <div class="col-md-2 col-sm-3 col-xs-4">
39             <div class="kpi">
40                 <div class="kpi-title">
41                     Scheduled
42                 </div>
43                 <div class="kpi-value">
44                     {{ totalfiltered | number }}
45                 </div>
46                 <div class="kpi-value-reference">
47                     {{{ total | number }}}
48                 </div>
49             </div>
50         </div>
51     <div class="col-md-2 col-sm-3 col-xs-4">
52         <div class="kpi">
```

