



**HAL**  
open science

## L'impact économique des données ouvertes

Ariane Gueke Kike Ayedoun

► **To cite this version:**

Ariane Gueke Kike Ayedoun. L'impact économique des données ouvertes. Sciences de l'information et de la communication. 2023. dumas-04235677

**HAL Id: dumas-04235677**

**<https://dumas.ccsd.cnrs.fr/dumas-04235677>**

Submitted on 10 Oct 2023

**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.

**Ariane Gueke Kike AYEDOUN**

**MASTER Information documentation  
Première année**

**Mémoire de stage**

**Mission effectuée du 17 mars au 7 Juillet 2023**

**au  
Laboratoire Gériico**

**Villeneuve-d'ascq France**

**Sujet : L'impact économique des données ouvertes**

**Sous la direction de :**

**M. Stéphane CHAUDIRON (Tuteur pédagogique)**

**et**

**M. Eric KERGOSIEN (Tuteur professionnel)**

**Soutenu le 03 Juillet 2023 à l'UFR DECCID-SID  
Université de Lille, Sciences Sociales et Humaines  
BP 60 149, 59 653 Villeneuve d'Ascq Cedex**

**Année Universitaire 2022/2023**



## Table des matières

Remerciements .....	3
Liste des sigles et abréviations .....	4
Résumé .....	5
Abstract .....	5
Introduction .....	6
Partie I : Comprendre les données ouvertes .....	9
1 - Le contexte actuel de l'ouverture des données .....	10
1.1- L'évolution des politiques et cadres réglementaires .....	10
1.2- L'influence des mouvements open source et open archive initiative.....	12
2- Mise en œuvre de la vision politique dans l'ouverture des données .....	13
2.1- Le cadre législatif et cadre juridique à l'international et en Europe .....	14
2.2- Création d'infrastructures technologiques pour les données ouvertes .....	16
3- Limites des données ouvertes.....	18
3.1- Protection des données sensibles .....	19
3.2- Accès inégal aux données ouvertes.....	21
4- Enjeux économiques des données ouvertes .....	22
4.1-Innovation et croissance économique .....	23
4.2-La création d'emplois et de nouvelles opportunités professionnelles.....	25
Partie II : Méthodologie (Analyse de Corpus).....	27
1-Présentation du terrain d'étude.....	27
1.1- Revue de la littérature .....	28
1.2-Présentation des travaux antérieurs existants sur la méthode .....	28
1.3- Les principaux concepts .....	30
2- Constitution et préparation du corpus.....	31
3- Analyse des données .....	33
3.1- Catégorisation.....	33
3.2- Analyse lexicale .....	36
Partie III : Exploration des résultats.....	38
1- Interprétation des résultats .....	38
2- Réflexion critique sur les résultats .....	39
2.1- Discussion des biais potentiels et des limites de l'interprétation.....	40
2.2- Suggestions pour de futures recherches ou approfondissements .....	41
Conclusion .....	43
Bibliographie .....	44
<b>Sitographie .....</b>	<b>46</b>
Annexes.....	49
Résumé .....	58

## Remerciements

- *A Monsieur Eric Kergosien, maître de conférences et directeur adjoint du laboratoire Gériico pour son accueil, sa sympathie, sa disponibilité, son accompagnement et pour tous les précieux conseils tout au long de cette année et pendant ce stage.*
  - *A Monsieur Stéphane Chaudiron, professeur à l'université de Lille, pour sa disponibilité, ses enseignements tout au long de l'année, sa sympathie et son aide précieuse dans la réalisation de ce mémoire.*
- *Au personnel administratif du laboratoire Gériico pour leur sympathie et chaleureux accueil.*
- *A mes chers parents, pour leur amour, leur encouragement et soutien indéfectible tout au long de mes études.*
- *A toutes ces personnes formidables trop nombreuses pour les citer, qui ont été d'un grand soutien depuis le début de cette année.*

## Liste des sigles et abréviations

API - Interface de programmation d'application (Application Programming Interface)

ASEDIE - Association des éditeurs de logiciels et solutions Internet pour les entreprises

BOAI - Initiative pour l'accès ouvert de Budapest (Budapest Open Access Initiative)

CADA - Commission d'accès aux documents administratifs

CKAN - Comprehensive Knowledge Archive Network

CNIL - Commission nationale de l'informatique et des libertés

CSV - Comma-Separated Values

FOIA - Freedom of Information Act

JSON - JavaScript Object Notation

OAI - Open Archives Initiative

OAI-PMH - Protocol for Metadata Harvesting

PSI - Public Sector Information

RDF - Resource Description Framework

RGPD - Règlement général sur la protection des données

RATP - Régie autonome des transports parisiens

UNESCO - United Nations Educational, Scientific and Cultural Organization

XML - eXtensible Markup Language

## Résumé

Les données ouvertes sont considérées comme un mouvement visant à rendre accessibles au public des ensembles de données de manière libre et gratuite. En comprenant les principes et les valeurs fondamentales du mouvement des données ouvertes, il devient aisé de saisir pleinement son potentiel et les implications qu'il peut avoir dans divers domaines. Cette pratique revêt une grande importance pour les organisations, qu'elles soient publiques ou privées, car la gestion efficace des données est devenue une préoccupation majeure à mesure que leur valeur économique est reconnue. A travers une approche d'analyse de corpus de la question de l'impact économique des données ouvertes, trois hypothèses sont examinées : l'ouverture des données favorise l'innovation et la création d'entreprises, l'accès aux données ouvertes améliore l'évaluation des risques et la prise de décision, et la disponibilité des données encourage la collaboration entre les acteurs économiques. Les résultats de cette recherche permettent de mieux comprendre comment l'open data peut stimuler l'économie et comment elle se manifeste.

**Mots-clés :** données ouvertes, économie, réutilisation, innovation

## Abstract

Open data refers to the concept of making sets of data freely and openly accessible to the public. Understanding the principles and core values underlying the open data movement enables us to fully grasp its potential and implications in various fields. This practice holds significant importance for both public and private organizations, as efficient data management has become a major concern as the economic value of data is recognized. Through a corpus analysis approach to the question of the economic impact of open data, three hypotheses are examined: open data fosters innovation and entrepreneurship, access to open data enhances risk assessment and decision-making, and data availability promotes collaboration among economic actors. The findings of this research provide a better understanding of how open data can stimulate the economy and manifest itself in practice.

**Keywords :** open data, economy, reuse, innovation

# Introduction

L'*open data* ou encore les données ouvertes désignent l'ensemble des données numériques dont l'accès, la réutilisation, la redistribution et l'usage sont de nature libre. En effet, les usagers grâce à l'*open data* peuvent accéder librement aux données mises en ligne par les institutions publiques. De ce fait, on pourrait s'attendre à ce que l'ouverture des données soit exclusivement une pratique réservée au secteur public. Cependant, la réalité est tout autre, quelques entreprises dans le monde aujourd'hui se démarquent en mettant également certaines de leurs données en *Open Data*. La volonté derrière ce mouvement est celle de la transparence et de la participation citoyenne. Les huit principes qui régissent l'*open data* ont été listés aux États-Unis en décembre 2007 lors d'une rencontre à Sebastopol en Californie. « À l'issue de deux jours de rencontre, les 30 participants ont défini ensemble huit critères qui permettent de qualifier une donnée publique d'ouverte :

*-complètes : toutes les données publiques doivent être rendues disponibles dans les limites légales liées à la vie privée ou la sécurité ;*

*-primaires : les données ouvertes sont telles que collectées à la source, non-agrégées avec le plus haut niveau de granularité ;*

*-fraîches : les données doivent être disponibles dès qu'elles sont produites ;*

*-accessibles : les données doivent être utilisables par le plus grand nombre d'utilisateurs potentiels ;*

*-lisibles par les machines : les données peuvent faire l'objet d'un traitement automatisé par les machines ;*

*-non discriminatoires : elles peuvent être utilisées par tous sans réclamer un enregistrement préalable ;*

*-dans un format ouvert : ce format ne doit pas être la propriété d'une organisation en particulier et doit faire l'objet d'une gouvernance commune par ses usagers ;*

*-avec une licence ouverte : les principes de Sebastopol vont plus loin que l'Open Definition en demandant que les données soient placées dans le domaine public (De l'Ardèche, 2019)<sup>1</sup>."*

---

<sup>1</sup> De l'Ardèche, S. G. P. L. C. D. (2019, 15 décembre). *Chapitre 1 Les grands principes d'une démarche d'open data | Rapport sur l'open data.*  
<https://dataactivist.coop/ardeche/rapport/partie1.html>



*“En 2013, verra le jour la charte internationale de l’open data qui fixe cinq principes qui vont compléter celles précédentes et créer un cadre pour les objectifs politiques :*

*-Pratique par défaut : L’objectif est de faire de l’open data la norme pour les administrations, tout en respectant les lois sur la propriété intellectuelle et la vie privée.*

*-Temps opportun et exhaustivité : Les données doivent être publiées rapidement et de manière complète, en fournissant des informations détaillées et de qualité.*

*-Accessibilité et utilisabilité : Les données doivent être accessibles sans enregistrement préalable, via des portails centraux, et être accompagnées d’une licence ouverte.*

*-Comparabilité et interopérabilité : Les données doivent être comparables et interopérables, fournies dans des formats structurés et normalisés qui favorisent leur réutilisation et leur compatibilité avec d’autres sources de données.*

*-Amélioration de la gouvernance et participation citoyenne : L’open data vise à améliorer la gouvernance en encourageant la transparence et la participation des citoyens dans la gestion des gouvernements et des institutions.*

*-Développement inclusif et innovation : Les données ouvertes sont perçues comme un moyen de favoriser le développement inclusif et l’innovation en donnant à tous les acteurs les moyens d’agir et de créer de la valeur ».*

Cette charte internationale complète les définitions précédentes de l’open data et établit des objectifs politiques ambitieux pour la publication des données ouvertes. Cependant, elle est moins opérationnelle pour déterminer si une donnée spécifique est considérée comme ouverte ou non (De l’Ardèche, 2019)<sup>2</sup>.

Pour codifier ces actions, la loi portant pour une république numérique sera promulguée en octobre 2016. Avec cette loi, l’open data devient non plus uniquement un choix de quelques-uns mais est désormais une obligation pour l’ensemble des acteurs investis d’une mission de service public (Agueke, 2023)<sup>3</sup>.

Dans le cadre de mon stage au laboratoire de recherche Geriico, j’ai pu m’intéresser de près à l’analyse des pratiques et usages des plateformes destinées au partage des données au niveau

---

<sup>2</sup> De l’Ardèche, S. G. P. L. C. D. (2019, 15 décembre). *Chapitre 1 Les grands principes d’une démarche d’open data | Rapport sur l’open data.*

<https://dataactivist.coop/ardeche/rapport/partie1.html>

<sup>3</sup> Agueke. (2023, 5 mai). *Données ouvertes institutionnelles.* ODEP.

<https://odep.hypotheses.org/789>

institutionnel en France. Les institutions sur lesquelles nous avons travaillé sont les ministères, les régions, les métropoles, les communes, les communautés de communes, les unités de recherche de l'université de Lille.

L'*open data* encore appeler données ouvertes désigne la pratique visant à rendre accessible au public les données en ligne. La gestion efficace des données est devenue une préoccupation importante pour les organisations qu'elles soient publiques ou privées. Quels sont alors les enjeux économiques de l'*open data* ? Dispose-t-on au sein des institutions, des ressources humaines et financières nécessaires pour garantir une mise en place efficace de l'ouverture des données ? Quelle est la qualité des données disponible sur les plateformes ? Comment la mesurer ? La création de valeur, qui se manifeste par un processus au cours duquel une organisation génère un avantage ou une amélioration dépassant le coût initial, est-elle exclusivement liée à la monétisation ? Ou bien la création de valeur peut-elle prendre d'autres formes ? Comment se manifeste la réutilisation par les tiers des données mises en ligne ? Quel est le modèle économique de l'ouverture des données ? L'ouverture de la donnée a-t-elle fait émerger un nouveau marché économique ? Quels sont les acteurs de cette filière ?

Pour y parvenir, nous avons identifié plusieurs hypothèses essentielles qui orienteront notre réflexion et notre analyse de l'impact économique de l'*open data*. Ces hypothèses servent de point de départ et seront évaluées grâce à une analyse de corpus. D'abord, l'ouverture des données peut stimuler l'innovation et la création d'entreprises : On peut supposer que la disponibilité des données ouvertes favorise l'émergence de nouvelles idées, produits et services et de même conduire à de nouvelles formes de revenus et de modèles économiques. Ensuite, l'accès aux données ouvertes peut vraisemblablement favoriser une évaluation plus précise des risques, une planification plus rigoureuse et une prise de décision mieux éclairée. Enfin, on pourrait supposer que la mise à disposition des données ouvertes encourage la collaboration entre les acteurs du secteur public et ceux du secteur privé et ainsi favoriser l'émergence de partenariats, d'initiatives conjointes et de synergies qui ont un impact économique positif. Dans la première partie de ce travail, nous procéderons à une revue de la littérature pour faire l'état de l'art sur le sujet. Dans la deuxième partie, nous aborderons la méthodologie d'analyse de corpus utilisée dans cette étude. Dans la troisième partie, nous présenterons les résultats de notre exploration.

# Partie I : Comprendre les données ouvertes

Comprendre les données ouvertes est essentiel pour saisir pleinement l'importance et les implications de ce mouvement. Les données ouvertes représentent un concept clé dans l'ère numérique, où la disponibilité et l'accessibilité des données jouent un rôle croissant dans la société et l'économie. Le mouvement des données ouvertes vise à rendre les informations et les ressources numériques accessibles au public, sans restriction ou avec des restrictions minimales, favorisant ainsi la transparence, l'innovation et la collaboration. Dans ce contexte, il est crucial de comprendre les principes fondamentaux des données ouvertes, tels que la libre disponibilité, l'interopérabilité, la réutilisation et la participation. Les données ouvertes sont généralement mises à la disposition du public dans des formats numériques lisibles par machine, ce qui facilite leur utilisation et leur traitement par des outils d'analyse automatisés. Cela permet aux individus, aux entreprises et aux organismes gouvernementaux de tirer parti de ces données pour développer de nouvelles idées, améliorer les services existants et prendre des décisions éclairées. En comprenant les principes et les valeurs sous-jacents du mouvement des données ouvertes, nous sommes en mesure d'apprécier pleinement son potentiel et ses implications dans divers domaines. Cela inclut des secteurs tels que l'économie, l'administration publique, la recherche scientifique, la santé, l'environnement et bien d'autres encore. En examinant les exemples réussis de l'utilisation des données ouvertes, nous pouvons identifier les opportunités, les défis et les meilleures pratiques pour exploiter pleinement le potentiel des données ouvertes dans notre société. Cette section se concentre sur la présentation des concepts et des principes fondamentaux du mouvement des données ouvertes, ainsi que sur l'exploration de son impact économique, social et technologique.

# 1 - Le contexte actuel de l'ouverture des données

Le contexte actuel de l'*open data* est marqué par une adoption croissante à l'échelle mondiale, stimulée par plusieurs facteurs clés. Tout d'abord, la demande de transparence gouvernementale pousse les administrations à rendre leurs données accessibles au public, favorisant ainsi la responsabilité et la confiance. Ensuite, les données ouvertes offrent de vastes opportunités d'innovation et de création de valeur pour les entreprises et les individus, stimulant ainsi le développement économique et technologique. De plus, l'engagement citoyen est renforcé grâce à l'accès aux informations publiques, permettant aux citoyens de participer activement à la prise de décision et à la résolution des problèmes sociétaux. Enfin, la collaboration entre les différentes parties prenantes, notamment les gouvernements, les entreprises, les chercheurs et la société civile, est favorisée par l'échange de données ouvertes. Cependant, deux paramètres essentiels doivent être pris en compte dans la conception actuelle des données ouvertes. D'une part, l'évolution des politiques et cadres réglementaires encadre la manière dont les données sont collectées, gérées et partagées, garantissant ainsi leur intégrité et leur confidentialité. D'autre part, l'influence grandissante des mouvements *open source* et *open science* incite à la collaboration, à la transparence et au partage des connaissances, créant ainsi un environnement propice à l'innovation et à l'avancement scientifique. Dans le développement qui suivra, nous examinerons plus en détail ces deux paramètres clés de l'*open data*, en mettant en évidence leur impact sur la société et les défis qu'ils soulèvent.

## 1.1- L'évolution des politiques et cadres réglementaires

Bien avant la loi du numérique de 2016, l'*open data* ou encore l'ouverture des données ont laissé des traces. Avant l'avènement du numérique, les prémices qui devaient servir plus tard de fondement à cette loi avaient déjà vu le jour. Ainsi, la déclaration des droits de l'homme et du citoyen de 1789 disposait déjà en son Article 15 : « *la Société a le droit de demander compte à tout Agent public de son administration* ». Cet article représente en effet l'un des premiers piliers juridiques de l'accès aux données publiques. Dans la continuité de cette volonté de transparence, une autre loi sera adoptée le 25 juin 1794. L'idée derrière cette loi était celle de permettre l'ouverture des dépôts d'archives aux citoyens. Plus tard, en 1978, deux lois sont promulguées. Il s'agit notamment de la loi informatique et libertés créant par ailleurs la

commission nationale de l'informatique et des libertés (CNIL) et de la loi mettant en place la commission d'accès aux documents administratifs (CADA).

Ces deux lois ont grandement permis l'établissement d'un cadre éthique et légal pour la protection de la vie privée des individus. Elles ont grandement contribué à la promotion de la transparence tout en respectant la protection des données personnelles. Dès 1997, le gouvernement au travers de son administration met en place la diffusion de nombreuses données textuelles dans le but de renforcer la politique de transparence. Des années après avec l'avènement du numérique sont lancés des portails de l'administration française respectivement : le portail [service-public.fr](http://service-public.fr) en 2000 et le portail [vie publique française](http://vie-publique.fr) en 2002. Ces portails ont permis à l'administration de créer une infrastructure solide de diffusion et de réutilisation des données publiques. En 2003, avec la directive européenne concernant la réutilisation des informations du secteur public, les conditions de la réutilisation à but lucratif ou non sont définies. Cette directive a été transposée dans la législation nationale des Etats membres de l'Union Européenne. En France, l'année 2005 sera celle de cette transposition. Il faudra attendre 2010 pour que les acteurs publics et notamment ceux des métropoles se lancent dans l'aventure des données ouvertes.

Le premier portail *Open Data* est de ce fait créé. C'est la métropole rennais qui se lance la première dans l'aventure avec la création du portail [data.rennesmetropole.fr](http://data.rennesmetropole.fr). Ensuite, [opendata.paris.fr](http://opendata.paris.fr) voit également le jour en 2011. Dès lors, les services publics commencent à mettre sur pied des portails leur permettant de pouvoir partager leurs données. En 2011 est créée la mission Etalab rattachée au secrétariat général pour la modernisation de l'action publique. Cette mission chargée de l'élaboration du portail interministériel des données avait pour but de promouvoir l'ouverture des données et leur réutilisation. C'est également en 2011, qu'est établie par le décret 2011-577 le droit de réutilisation des données sans frais par le public. Afin de mieux superviser la politique d'ouverture des données, l'État créé le poste d'administrateur général des données par un décret du 16 septembre 2014. Placé sous l'autorité du premier ministre, l'administrateur général des données est chargé de surveiller et d'améliorer l'usage des données par les acteurs impliqués. La loi portant pour une république numérique de 2016 viendra mieux encadrer les choses tout en contraignant tous les agents du service publics à l'ouverture des données produites préalablement dans le cadre de leurs missions. L'*open data* devient non plus uniquement le choix de quelques-uns mais désormais une obligation pour l'ensemble des acteurs investis d'une mission de service public.

## 1.2- L'influence des mouvements open source et open archive initiative

Dans les années 1990, l'émergence de plusieurs mouvements *Open* a fortement influencé l'ouverture des données telle que nous la connaissons aujourd'hui. L'*open source* ou encore l'accès libre est un logiciel libre. Richard Stallman alors programmeur et militant en est à l'origine. L'idée est de faire des utilisateurs plus que des consommateurs et de les impliquer à part entière dans le développement et l'édition du logiciel. Son idée sera concrétisée en 1997, par Eric Raymond, un hacker qui fera naître le mouvement open source. On entend par *open source*, l'accès au code du logiciel afin que des tiers puissent en disposer et faire évoluer le logiciel. Le mouvement *open source*, qui promeut le partage libre et ouvert du *code source* des logiciels, a jeté les bases d'une culture de collaboration et de transparence.

Les principes fondamentaux de l'open source, tels que l'accès libre, la possibilité de modifier et de distribuer le *code source*, ont inspiré de nombreuses initiatives d'ouverture des données. Les pratiques et les valeurs de l'*open source* ont été adoptées pour la publication et la réutilisation des ensembles de données, créant ainsi un environnement propice à la libre circulation de l'information. Dans le contexte des données ouvertes, le mouvement *open source* joue un rôle essentiel dans le développement d'outils, de plateformes et de normes qui facilitent la collecte, la gestion et la diffusion des données. Des solutions *open source* telles que CKAN, Data.gov, et OpenRefine ont été développées pour faciliter la publication et la réutilisation des données ouvertes, offrant des infrastructures robustes et évolutives.

L'*Open Archives Initiative* (OAI) est une initiative visant à faciliter l'échange et la valorisation des archives numériques et ayant vu le jour en 1999. Avant la création de ce projet, les abonnements aux revues papier et électroniques avaient connu une augmentation des prix en raison de la position dominante de certains éditeurs, résultant de mouvements de consolidation dans l'édition scientifique. Cela a créé un quasi-monopole et rendu difficile pour les bibliothécaires de proposer une offre documentaire complète et cohérente. Parallèlement, le développement et la démocratisation d'Internet ont coïncidé avec l'émergence des premiers dépôts d'archives ouvertes. Cette conjonction a joué un rôle significatif dans la prise de conscience des scientifiques de l'impact potentiel d'une publication diffusée par le biais de ce média. Elle vise à promouvoir l'accès libre et gratuit aux publications scientifiques, permettant

ainsi à la communauté scientifique mondiale d'accéder aux résultats de la recherche de manière plus ouverte et collaborative. L'OAI a développé le protocole OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), qui permet l'interopérabilité entre les archives ouvertes, facilitant ainsi la recherche et la récupération des publications scientifiques.

Ces mouvements ont une influence considérable sur l'ouverture des données en créant des normes, des pratiques et des communautés dédiées à la libre circulation de l'information. Ils encouragent les principes de transparence, de collaboration et de partage des connaissances, qui sont essentiels pour l'évolution des données ouvertes. Grâce à l'*open source* et à l'OAI, de nombreux acteurs, qu'il s'agisse de gouvernements, de chercheurs, d'entreprises ou de la société civile, ont pu bénéficier d'outils et de ressources qui facilitent l'accès, la réutilisation et l'exploitation des données, contribuant ainsi à l'émergence d'une culture de l'ouverture des données à l'échelle mondiale.

## 2- Mise en œuvre de la vision politique dans l'ouverture des données

La mise en œuvre de la vision politique en matière d'ouverture des données nécessite la combinaison d'un cadre législatif solide, d'infrastructures technologiques adaptées et d'une sensibilisation accrue des acteurs concernés. Cette vision politique repose sur la volonté des gouvernements et des institutions de promouvoir l'accès libre et ouvert aux données, de faciliter leur réutilisation et de favoriser l'innovation.

Pour mettre en œuvre cette vision, plusieurs actions clés doivent être entreprises. Tout d'abord, il est essentiel d'établir un cadre législatif solide qui définit les droits et les obligations liés à l'ouverture des données. Cela implique d'adopter des lois et des politiques qui obligent les institutions publiques à publier activement des ensembles de données dans des formats ouverts, accessibles et réutilisables par tous. Ce cadre législatif peut également inclure des mécanismes de protection de la vie privée et de sécurité des données, afin de garantir une utilisation responsable des informations publiques. En parallèle, la mise en place d'infrastructures technologiques adaptées est essentielle pour faciliter la collecte, la gestion et la diffusion des données ouvertes. Cela implique la création de portails de données ouvertes, de plateformes de partage et de collaboration, ainsi que de normes et de protocoles techniques pour assurer l'interopérabilité des données. Ces infrastructures technologiques doivent être évolutives,

sécurisées et conviviales, afin de faciliter l'accès et la réutilisation des données par un large éventail d'utilisateurs, qu'il s'agisse de chercheurs, d'entreprises ou de la société civile.

## 2.1- Le cadre législatif et cadre juridique à l'international et en Europe

À l'échelle mondiale, diverses mesures ont été instaurées afin de faciliter la mise à disposition des données. Parmi ces mesures figure la Déclaration de l'Organisation des Nations Unies sur l'accès à l'information, qui garantit le droit d'accéder aux informations détenues par les gouvernements et met en évidence l'importance de l'ouverture des données en termes de transparence, de responsabilité et de participation des citoyens. « *L'UNESCO défend l'accès à l'information en tant que liberté fondamentale et pilier essentiel de l'édification de sociétés du savoir inclusives. L'organisation promeut les droits et les valeurs énoncés dans l'article 19 de la Déclaration universelle des droits de l'homme* ». <sup>4</sup>

La déclaration en reconnaissant le droit fondamental des individus à accéder aux informations détenues par les gouvernements, crée une base solide pour promouvoir l'ouverture des données et demande aux gouvernements de faciliter cet accès de manière proactive. On peut y voir l'intention de créer un environnement propice à l'ouverture de la donnée.

Au niveau européen, la Directive européenne sur la réutilisation des informations du secteur public, également connue sous le nom de Directive PSI, joue un rôle essentiel dans la promotion de l'ouverture des données. Cette directive établit des règles encadrant la réutilisation des informations provenant du secteur public, en exigeant des États membres de l'Union européenne qu'ils rendent ces informations facilement accessibles et réutilisables. Elle met également en place des principes de non-discrimination, de transparence et de tarification équitable pour la réutilisation des données publiques. Le 4 juillet 1966, le président des États-Unis Lyndon B. Johnson a promulgué la loi sur la liberté d'information, connue sous le nom de *Freedom of Information Act* (FOIA). Cette loi a été adoptée en réponse au mécontentement des citoyens concernant la gestion de la guerre du Vietnam. Elle oblige les agences fédérales à

---

<sup>4</sup> Droit à l'information. (2022). *Articles*. <https://www.unesco.org/fr/right-information#:~:text=L'UNESCO%20d%C3%A9fend%20l'acc%C3%A8s,des%20droits%20de%20l'homme>.



rendre leurs documents administratifs disponibles à toute personne qui en fait la demande. Cependant, ce n'est qu'en 1995 que le concept même de l'open data a fait son apparition. Cela s'est produit lorsque le Conseil national de la recherche aux États-Unis a exprimé sa volonté de promouvoir les échanges ouverts et complets de données scientifiques. Il convient toutefois de souligner que cette idée n'était pas vraiment nouvelle, car dès 1942, Robert King Merton appelait les scientifiques à abandonner la logique basée sur les droits de propriété intellectuelle et à partager librement les connaissances, considérant cela comme une condition essentielle pour une recherche véritablement efficace.<sup>5</sup> (Meszaros, 2015)

*« Les premiers résultats tangibles sont apparus en décembre 2001 avec le lancement de l'Open Access Initiative (BOAI), également connu sous le nom de Budapest Open Access Initiative. Ce projet visait à établir la première archive ouverte, en imposant notamment l'utilisation de formats d'échange ouverts tels que le XML, même si à l'époque on parlait davantage de libre accès ou de libre diffusion. Par la suite, les choses ont légèrement accéléré. En 2002, Lawrence Lessig crée la licence Creative Commons. En 2004, l'Open Knowledge Foundation est entrée en scène, ayant pour objectif de fournir des logiciels adaptés pour faciliter la mise à disposition des données et de développer le cadre juridique nécessaire à ces données. Ainsi, en 2007, la licence ODbL, principalement destinée aux bases de données, a été créée. La même année, la conférence sur le gouvernement ouvert s'est tenue à Sebastopol en Californie, où ont été élaborés les 8 critères fondamentaux qui définissent les données ouvertes. Toujours en 2007, l'amendement Honest Leadership and Open Government Act, également connu sous le nom de l'OPEN Government Act of 2007, a été adopté. En 2009, le jour même de son investiture en tant que président des États-Unis, Barack Obama a lancé le projet Open Government Initiative, un pas supplémentaire vers la transparence gouvernementale, qui a abouti à l'ouverture du premier site dédié principalement aux données ouvertes, le portail data.gov. En 2009, le gouvernement britannique a également lancé le portail data.gov.uk, complété un an plus tard, en septembre 2010, par la licence Open Government Licence. Pour renforcer cette logique d'ouverture et lui donner plus de légitimité et de cadre pratique, les dirigeants du G8 ont signé en juin 2013, lors du sommet de Lough Erne en Irlande du Nord,*

---

<sup>5</sup> Meszaros, B. (2015, 10 juin). *Livre blanc sur les données ouvertes*.  
<https://inria.hal.science/AO-DROIT/halshs-01162692v1>

*une charte pour l'ouverture des données publiques, affirmant ainsi clairement la place que les données ouvertes ont su conquérir et doivent occuper au sein de nos sociétés. »<sup>6</sup>*

## 2.2- Création d'infrastructures technologiques pour les données ouvertes

La création d'infrastructures technologiques pour l'open data, implique de mettre l'accent sur les plateformes de données ouvertes, les formats de données ouverts, les normes et métadonnées, les API et les politiques de gouvernance des données. Leur rôle demeure crucial dans le soutien à l'ouverture des données et pourrait avoir un impact sur la création d'un écosystème d'ouverture des données plus dynamique et collaboratif. Les plateformes de données ouvertes sont ainsi des systèmes en ligne qui permettent de stocker, gérer et publier des ensembles de données dans des formats ouverts et interopérables. Elles offrent aux gouvernements, aux organisations et aux individus un espace centralisé pour partager et accéder à des données de manière transparente. Au nombre de ces plateformes, nous pouvons citer à titre d'exemple :

- Data.gouv.fr : Cette plateforme est l'initiative Open Data du gouvernement français. Elle offre un large éventail de jeux de données provenant de différentes administrations, organismes publics et collectivités territoriales. Les données couvrent divers sujets tels que les finances publiques, la démographie, l'environnement, etc.
- European Data Portal : Il s'agit de la plateforme Open Data de l'Union européenne. Elle rassemble les données provenant des différents États membres de l'UE et offre un accès centralisé à un large éventail de jeux de données sur divers sujets, y compris l'économie, l'agriculture, l'énergie, etc.
- OpenDataSoft : C'est une plateforme Open Data qui permet aux organisations et aux administrations de publier, partager et visualiser leurs données de manière conviviale. Elle offre également des fonctionnalités de visualisation et de création de tableaux de bord interactifs.

---

<sup>6</sup> Ibid

« Une API est un ensemble d'outils et de protocoles facilitant l'intégration de services et de logiciels d'applications. Elle permet à vos produits et services de communiquer avec d'autres, sans avoir à renforcer votre infrastructure de connectivité. Une API peut être privée (interne), partenaire (partagée avec des partenaires spécifiques pour générer des revenus supplémentaires) ou publique (en interaction avec des applications tierces pour favoriser l'innovation) ». <sup>7</sup> Le partage d'API offre divers avantages tels que la création de nouveaux revenus, l'extension de la portée de votre marque, et la stimulation de l'innovation et de l'efficacité grâce à la collaboration externe. Elles facilitent l'intégration des données dans des applications tierces et encouragent le développement d'applications innovantes qui utilisent les données ouvertes. Les API permettent également la mise à jour en temps réel des données et favorisent la réutilisation dynamique des informations. <sup>8</sup>

Concernant les formats de données ouverts, ils sont essentiels pour l'interprétation et la réutilisation des données. Ils garantissent que les données peuvent être lues et utilisées par différents systèmes et applications, sans restriction techniques ou propriétaires. Des formats tels que CSV, JSON, XML et RDF sont couramment utilisés pour l'*open data*, favorisant ainsi l'interopérabilité et la réutilisation des données. <sup>9</sup>

---

<sup>7</sup> Comprendre ce qu'est une API. (s. d.-b). <https://www.redhat.com/fr/topics/api>

<sup>8</sup> Comprendre ce qu'est une API. (s. d.). <https://www.redhat.com/fr/topics/api>

<sup>9</sup> Dymytrova, V. (2020, 5 août ). *Du traitement des données à la création de valeur : comprendre les pratiques professionnelles des réutilisateurs des données ouvertes.* <https://hal.science/hal-02913346>

### 3- Limites des données ouvertes

L'ouverture des données présente de nombreux avantages, mais il est également important de reconnaître les limites et les défis associés à cette pratique. Sont exclus du droit d'accès les documents administratifs pouvant porter atteinte à l'exercice des activités régaliennes de l'État et à l'intérêt général. Plusieurs secrets sont protégés notamment concernant les délibérations du Gouvernement, le secret de la défense nationale, la sécurité publique ou encore le déroulement des procédures juridictionnelles (De l'Ardèche, 2019)<sup>10</sup>. La protection des données sensibles constitue ainsi l'une des principales préoccupations en matière d'ouverture des données. Alors que la transparence et l'accès aux informations sont valorisés, certaines données, notamment celles liées à la sécurité et à la défense, doivent être protégées. La divulgation de telles données pourrait compromettre la sécurité nationale, la vie privée des individus ou les intérêts stratégiques. Par conséquent, il est essentiel d'établir des mécanismes appropriés pour garantir la sécurité et la confidentialité de ces données sensibles tout en promouvant l'ouverture des autres données non sensibles. Un autre défi majeur réside dans l'accès inégal aux données ouvertes. Bien que l'objectif de l'ouverture des données soit de rendre les informations accessibles à tous, il existe une fracture numérique qui limite la capacité de certaines populations à bénéficier pleinement des données ouvertes.

---

<sup>10</sup> De l'Ardèche, S. G. P. L. C. D. (2019, 15 décembre). *Chapitre 1 Les grands principes d'une démarche d'open data | Rapport sur l'open data.*  
<https://dataactivist.coop/ardeche/rapport/partie1.html>

### 3.1- Protection des données sensibles

Suite à la loi Lemaire de 2017, le gouvernement français a précisé que l'ouverture des données publiques ne concerne ni les informations personnelles, ni celles liées à la sécurité nationale, ni celles couvertes par les différents secrets légaux<sup>11</sup>. Le Règlement général sur la protection des données (RGPD) est une réglementation européenne qui vise à protéger les droits et la vie privée des individus en ce qui concerne le traitement de leurs données personnelles. Bien que l'ouverture des données favorise la transparence et la collaboration, elle doit également respecter les principes et les obligations du RGPD<sup>12</sup>. Pour une question d'éthique, il est donc nécessaire de mettre en place des mesures de sécurité et de confidentialité appropriées pour protéger les données personnelles, ainsi que d'obtenir le consentement des individus concernés avant leur traitement. Ainsi, le RGPD joue un rôle essentiel dans la limitation de l'ouverture des données, en imposant des obligations de conformité et en mettant l'accent sur la protection des droits des individus en matière de confidentialité et de contrôle de leurs données personnelles.

La sécurité nationale n'est qu'un aspect parmi d'autres justifiant le refus de communication des données publiques. La protection du secret de la défense nationale constitue également un motif complexe, notamment lorsque des impératifs de transparence, tels que ceux liés à l'environnement ou au nucléaire, entrent en conflit. Ainsi, les impératifs sécuritaires peuvent limiter considérablement le mouvement de l'open data, bien qu'un équilibre puisse être trouvé pour inclure les missions régaliennes de l'État dans le périmètre de l'ouverture des données, tout en préservant les avantages mutuels que l'autorité publique et les citoyens devraient en retirer.<sup>13</sup> D'après l'article 6 de la loi 17 juillet 1978, aujourd'hui codifié à l'article L 311-5 CRPA, exclut le partage au grand public des données qui relèvent du : « *secret des délibérations du gouvernement et des autorités responsables relevant du pouvoir exécutif, (2o a) ou par le « secret de la défense nationale , (2o b), celles qui concernent la*

---

<sup>11</sup> <https://cfdc2023.univ-tln.fr/wp-content/uploads/Atelier-6-AGUILON-Claire.pdf>

<sup>12</sup> *Le règlement général sur la protection des données - RGPD*. (s. d.). CNIL.  
<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

<sup>13</sup> Warusfel, B. (2018b). Enjeux et limites de l'ouverture des données en matière de sécurité et de défense. *Revue française d'administration publique*, 167(3), 551.  
<https://doi.org/10.3917/rfap.167.0551>

*conduite de la politique extérieure de la France , (2o c) ou encore qui sont relatives à la sûreté de l'État, à la sécurité publique ou à la sécurité des personnes (2o d) (Open Data à Paris : et maintenant ? , s. d.-b) ».*<sup>14</sup>

La Commission d'accès aux documents administratifs (CADA) et la jurisprudence administrative ont identifié différents types de documents administratifs jugés non communicables pour des raisons de sécurité. Par exemple, les documents révélant la position de la France dans une négociation ou pouvant affecter les négociations internationales en cours ne sont pas communicables. En ce qui concerne la sécurité publique, la CADA refuse la communication des documents qui pourraient compromettre la protection des personnes et des biens ou entraver l'action publique en matière de maintien de l'ordre et de sécurité publique. Par exemple, la localisation des caméras de vidéosurveillance d'une commune, un plan détaillé d'une installation de stockage d'hydrocarbures incluant des informations sur la sécurité, ou une base de données recensant les incidents sur le réseau ferroviaire ne sont pas communicables.

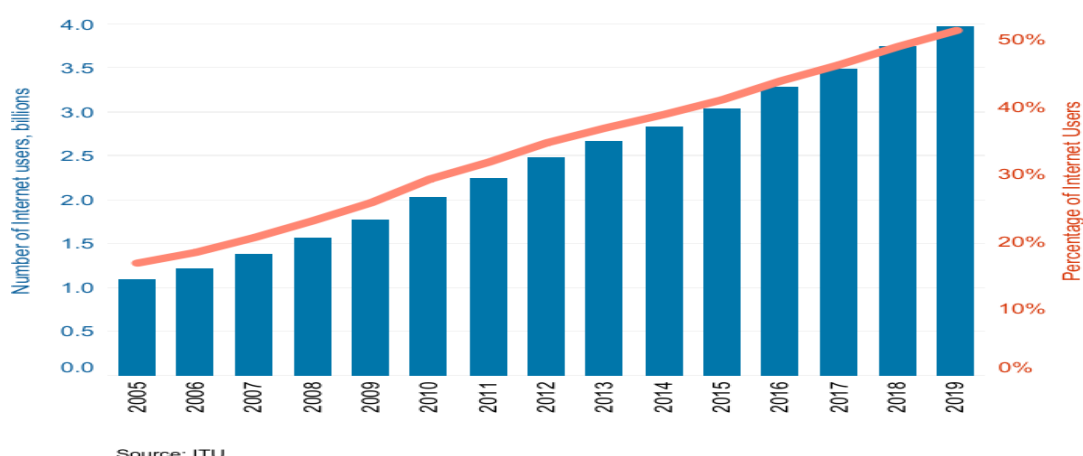
La question de l'accès aux données ouvertes peut également poser des problèmes lorsqu'on en vient à des domaines sensibles tel que le nucléaire. Malgré le droit des individus à obtenir des informations sur les risques environnementaux que pourraient produire la radioactivité, certaines informations peuvent ne pas être communiquées en raison de leur caractère sensible et donc relevant de la sécurité nationale. En effet, la lutte contre le terrorisme peut entraver la transparence gouvernementale et l'*open data*, car elle concerne à la fois la criminalité et la violence politique, impliquant les exceptions sécuritaires liées à la sécurité publique, aux procédures judiciaires et à la sécurité nationale. Cela peut conduire au refus de divulguer des informations essentielles pour évaluer les risques sanitaires ou environnementaux au nom de la prévention du terrorisme (Curtin, 2003, 101).

---

<sup>14</sup> Ibid

## 3.2- Accès inégal aux données ouvertes

La fracture numérique est un concept complexe qui englobe plusieurs aspects. Elle se réfère au fossé entre les personnes qui ont accès à Internet et celles qui en sont exclues. Cependant, la fracture numérique ne se limite pas uniquement à l'accès, mais comprend également des aspects tels que l'abordabilité, la qualité et la pertinence de la connexion. Comme souligné par Michael Kende, la fracture numérique n'est pas une notion binaire. Les disparités d'accès à Internet sont causées par divers facteurs. Ces facteurs sont la disponibilité, l'accessibilité financière, la qualité du service, la pertinence, l'interconnectivité, la culture numérique et l'accès aux équipements. Selon la Commission des Nations unies sur le haut débit pour le développement durable, environ 3,6 milliards de personnes, soit près de 53,6% de la population mondiale, ne sont pas connectées à Internet (De Vasconcelos Aguiar & Muller, 2022).



**Figure 1 : Taux de connectivité à Internet dans le monde par l'agence spécialisée des Nations Unies chargée des technologies de l'information et de la communication**

Il y aurait actuellement environ 11 millions de Français (soit un peu plus de 20% de la population) qui sont touchés par l'« illettrisme », selon le Syndicat de la presse sociale (SPS) dans son livre blanc. Parmi les adultes en situation de précarité, 18% rencontrent des difficultés liées aux dysfonctionnements, aux problèmes techniques et aux délais d'attente, ce qui les bloque dans leurs démarches. Environ un adulte sur cinq éprouve des difficultés pour accéder aux outils numériques. La fracture numérique générationnelle est également observée, avec les personnes âgées étant les plus touchées. 27% des personnes âgées de 60 ans ne sont pas

connectées à Internet, et ce pourcentage atteint 42% chez les plus de 80 ans. Cependant, il serait incorrect de se concentrer uniquement sur la fracture numérique générationnelle, car il est également constaté que les jeunes souffrent parfois d'illectronisme.

Selon un rapport de l'Insee en 2019, un peu plus de 21% de la population âgée de plus de 15 ans éprouve des difficultés à communiquer via Internet. Il convient de noter que ces données concernent les personnes âgées de 15 ans et plus en France (hors Mayotte), bien que l'illectronisme soit plus répandu chez les retraités (Akbaraly, s. d.)<sup>15</sup>.

## 4- Enjeux économiques des données ouvertes

Les données ouvertes, la pratique consistant à rendre les données et informations publiques accessibles et utilisables par tous, ont radicalement changé le paysage économique. En permettant la libre circulation des données, les données ouvertes peuvent favoriser l'innovation, stimuler la croissance économique et créer de nouvelles opportunités commerciales. En moyenne, 62 % des États membres de l'UE estiment que l'utilisation des données ouvertes a un impact sur l'emploi, la technologie et l'innovation, ainsi que sur la création d'entreprises<sup>16</sup>. Les gouvernements, les entreprises et les particuliers ont désormais accès à de vastes quantités d'informations précieuses, allant des données démographiques aux statistiques économiques, en passant par les données environnementales et géographiques. L'accessibilité accrue aux données a le potentiel d'améliorer l'efficacité opérationnelle, de faciliter la prise de décision éclairée et d'accroître la transparence.

---

<sup>15</sup> Akbaraly, M. (s. d.). Fracture numérique : Quelles Conséquences sur l'Accessibilité ? <https://blog.ipedis.com/fracture-numerique-queelles-consequences-sur-laccessibilite>

<sup>16</sup> Publications Office of the EU. (2022, 14 décembre). The Open Data Maturity Report 2022 is out ! *Publications Office of the EU*. <https://data.europa.eu/en/news-events/news/open-data-maturity-report-2022-out>



## 4.1-Innovation et croissance économique

En plus de répondre au besoin croissant de responsabilité sociale et environnementale, les données ouvertes peuvent avoir un impact significatif sur l'économie en améliorant les services et en augmentant la transparence. Bien que les modèles économiques et les licences ne soient pas totalement stabilisés, le développement des applications mobiles et de la géolocalisation devrait profiter à de nombreuses applications potentielles. L'accessibilité et la réutilisation des données ouvertes encouragent l'innovation en fournissant un référentiel de connaissances et d'informations sur lequel les entrepreneurs et les développeurs peuvent s'appuyer pour créer de nouvelles applications et solutions. En tirant parti des données ouvertes, les entrepreneurs peuvent également découvrir de nouvelles opportunités commerciales et créer des modèles commerciaux innovants. La réduction des temps de trajet et des émissions de gaz à effet de serre peut être obtenue grâce à des applications de planification d'itinéraire plus efficaces, utilisant les données des transports publics. Il est également possible de comprendre la consommation d'énergie et de réduire l'empreinte carbone grâce à des applications de gestion de l'énergie alimentées par des données énergétiques.

Une application de planification d'itinéraires urbains, citymapper, utilise des données ouvertes sur les transports publics comme base. Présente dans 41 villes d'Europe, d'Asie et des États-Unis, la startup créée par Azmat Yusuf, 39 ans, ancien employé de Google. La startup utilise les données des autorités de transport pour fournir des horaires de bus et de train à jour, des itinéraires en direct, ainsi que des détails sur les perturbations (Dor, 2020). En utilisant divers modes de transport, les utilisateurs peuvent utiliser les itinéraires disponibles pour planifier efficacement leurs déplacements grâce aux données. Cependant, Citymapper a rencontré des difficultés pour le déployer en France lorsqu'un désaccord est survenu en mars 2016 entre la Régie autonome des transports parisiens (RATP) et Citymapper en Île-de-France. Citymapper a allégué que la RATP les empêchait d'accéder aux données en temps réel des horaires de métro et de bus en interdisant l'adresse IP. En 2016, l'abondance des demandes de services de Citymapper a mis la RATP en difficulté. Pour y remédier, la RATP divulguait ses données en janvier de l'année suivante, mais elle imposait également une redevance sur les

demandes dépassant un certain plafond (30 millions de requêtes par mois) <sup>17</sup>. Sur leur site internet, citymapper s'exprime en ces termes : « *Il est essentiel de rappeler les bénéfices que l'Open Data va apporter aux Franciliens : des services innovants, une meilleure information pour l'utilisateur et une fréquentation accrue des transports en commun. Le tout grâce à des applis développées sans frais pour le contribuable. Notre mission chez Citymapper est de rendre les villes plus faciles à vivre. Nous sommes convaincus que le futur de la ville passe par une collaboration avec les entreprises innovantes et souhaitons être un partenaire des autorités publiques* »<sup>18</sup>.

Ainsi, les données ouvertes créent de la valeur économique grâce à la monétisation.

En République tchèque, l'organisation "Czechitas" organise le projet "Digital Academy", qui se concentre spécifiquement sur l'analyse des données à l'aide de données ouvertes. Cela facilite la formation informatique des femmes et leur intégration ultérieure dans le marché du travail informatique. En France, Open Data for Businesses a contribué au développement d'un annuaire national des entreprises basé sur une API ouverte de données d'entreprise ouvertes pour faciliter la fourniture de services aux entreprises. Un autre exemple d'utilisation des numéros d'identification d'entreprise est le service "My Business", qui fournit des simulateurs, des calculatrices et des outils pour aider les entrepreneurs (Office des publications de l'UE, 2022).

---

<sup>17</sup> Jacqué, P. (2016, 9 avril). Citymapper, cette start-up qui agace la RATP. *Le Monde.fr*. [https://www.lemonde.fr/economie/article/2016/04/09/citymapper-cette-start-up-qui-agace-la-ratp\\_4899122\\_3234.html](https://www.lemonde.fr/economie/article/2016/04/09/citymapper-cette-start-up-qui-agace-la-ratp_4899122_3234.html)

<sup>18</sup> *Open Data à Paris : et maintenant ?* (s. d.). Citymapper. <https://citymapper.com/news/1307/open-data-a-paris-et-maintenant>

## 4.2-La création d'emplois et de nouvelles opportunités professionnelles

L'ouverture des données publiques, souvent appelées données ouvertes, joue un rôle essentiel dans la création d'emplois et de nouvelles opportunités de carrière. En permettant un accès ouvert et gratuit à une variété d'ensembles de données, les données ouvertes favorisent l'émergence de *start-ups*, d'entreprises innovantes et de projets entrepreneuriaux axés sur l'exploitation de ces données. Ces initiatives contribuent à la croissance économique en créant des emplois dans des domaines tels que l'analyse de données, le développement d'applications, la visualisation de données et la conception de produits et de services. Elle favorise également le développement des compétences numériques et l'acquisition d'expertises. Les personnes intéressées par l'analyse de données, la science des données, la programmation et d'autres domaines liés à l'exploration de données peuvent bénéficier d'une formation, de programmes d'apprentissage et de programmes éducatifs axés sur les données ouvertes. Cela leur permet d'acquérir les compétences dont ils ont besoin pour travailler dans ces industries en expansion et de saisir de nouvelles opportunités de carrière.

En Lituanie, les solutions de neurotechnologie collectent automatiquement des données librement disponibles pour tirer des conclusions sur les conditions du marché du travail et les besoins futurs. La solution peut prévoir de manière exhaustive et systématique les besoins du marché du travail dans la région, afin d'organiser de manière ciblée les services locaux d'orientation professionnelle, de formation et de reconversion. La solution nécessite moins de ressources humaines et financières. En Italie, les données ouvertes facilitent l'analyse et l'identification de mesures appropriées pour promouvoir une croissance économique durable, inclusive et durable, l'emploi productif et un travail décent pour tous. Selon les prévisions du scénario optimiste, le rapport slovène sur l'impact économique et social des données ouvertes dans le pays montre que le potentiel de croissance du marché des données ouvertes est estimé à 293 millions d'euros d'ici 2025. On estime que 7 441 employés travailleront dans des données ouvertes d'ici 2025.

En Espagne, l'association d'information intersectorielle ASEDIE analyse chaque année la valeur économique et sociale des entreprises qui réutilisent les données du secteur public (et privé) pour développer des produits à valeur ajoutée. Selon l'édition 2022 de son rapport, grâce à l'exploitation de ces données, ces entreprises ont généré un chiffre d'affaires de plus de 2

milliards d'euros et ont fait travailler près de 23 000 professionnels. En Allemagne, le ministère fédéral allemand de l'Économie et du Climat a publié une étude qui examinait, entre autres, l'impact économique des ensembles de données à haute valeur ajoutée sur l'économie allemande (Office des publications de l'UE, 2022)<sup>19</sup>.

*« Pour les clients historiques de ce marché de la donnée matière première, ceux qui s'acquittaient autrefois de redevances pour réutiliser les données, l'effet est multiple. Ils ont tout d'abord réalisé une économie nette, la ressource devenant gratuite. Mais certains d'entre eux ont pu par ailleurs regretter que la relation avec les producteurs se soit banalisée et qu'en perdant la relation commerciale ils perdaient aussi la capacité à agir sur la qualité des données fournies, par exemple. Enfin, les redevances agissaient aussi comme une barrière à l'entrée: la donnée devenant gratuite, de nouveaux réutilisateurs sont apparus, parfois en concurrence directe avec les clients historiques du marché »<sup>20</sup>.*

---

<sup>19</sup> Publications Office of the EU. (2022, 14 décembre). The Open Data Maturity Report 2022 is out ! *Publications Office of the EU*. <https://data.europa.eu/en/news-events/news/open-data-maturity-report-2022-out>

<sup>20</sup> <https://fing.org/wp-content/uploads/2020/02/Open-Data-impact-bilan-2010-2018-de-l-open-data.pdf>

## Partie II : Méthodologie (Analyse de Corpus)

La méthodologie de recherche joue un rôle fondamental dans notre étude car elle définit l'approche utilisée pour collecter, analyser et interpréter les données. Dans cette section, nous présenterons en détail la méthodologie adoptée pour mener notre étude. Cette dernière porte sur l'analyse de corpus, qui est une approche méthodologique utilisée en linguistique pour analyser de grands corps de texte. Nous avons choisi cette méthodologie car elle permet un examen complet des modèles et des structures linguistiques dans un corpus donné. Pour mener notre analyse, il est important que nous présentions notre terrain d'étude, la revue littéraire sur le sujet et l'analyse de notre corpus.

### 1-Présentation du terrain d'étude

Lors de mon stage au sein du laboratoire de recherche Gériico, j'ai eu l'opportunité de travailler sur les usages des plateformes de données ouvertes des institutions françaises. Ma tâche principale était la revalorisation des études que les étudiants ont menées précédemment. Au fur et à mesure que je me familiarisais avec cet environnement, des préoccupations sont nées. Parmi elles, une a particulièrement retenu mon attention : il s'agit de l'impact économique des données ouvertes. Le choix d'étudier l'impact économique des données ouvertes repose sur plusieurs facteurs clés. Tout d'abord, les données sont devenues une ressource précieuse dans l'économie numérique, et la collecte de données massives est omniprésente. Ainsi, il devient crucial d'explorer comment utiliser efficacement ces données pour stimuler la croissance économique. L'open data offre un énorme potentiel d'innovation, de création de valeur et de développement économique. Il est alors essentiel de comprendre comment les données ouvertes peuvent faciliter l'activité économique, encourager l'innovation et générer de nouvelles opportunités commerciales.

Dans le même temps, les données ouvertes peuvent améliorer la transparence et la gouvernance, en renforçant la responsabilité dans des domaines aussi divers que les finances publiques, les marchés ou la santé. En analysant l'impact économique des données ouvertes, nous pouvons évaluer comment la disponibilité des données affecte la gouvernance, l'efficacité des politiques publiques et la promotion d'un environnement économique plus juste et équitable. Ces réflexions s'entremêlent naturellement avec ma volonté de réévaluer les recherches antérieures des étudiants, en mettant l'accent sur l'impact économique des données ouvertes.

Enfin, il existe un besoin croissant d'informations fiables pour prendre des décisions éclairées, développer de nouvelles solutions et innover. Les données ouvertes répondent à ce besoin en donnant accès aux données, ce qui peut favoriser la création d'entreprise, augmenter la productivité et encourager l'entrepreneuriat. Cette dimension pratique et entrepreneuriale de l'*open data* a été le moteur de ma décision d'approfondir la recherche sur son impact économique.

## 1.1- Revue de la littérature

La pratique de l'analyse de contenu ne se limite pas aux linguistes, contrairement à ce que l'on pourrait penser. Elle est également utilisée dans d'autres domaines tels que le droit et les sciences politiques, où elle sert d'outil d'analyse. L'analyse de contenu peut prendre différentes formes en fonction du domaine d'application, et il n'existe donc pas de méthode unique pour la réaliser. Chaque approche dépend du type de support étudié. De plus, la définition de l'analyse de contenu a évolué au fil du temps et varie en fonction des auteurs. Dans les paragraphes suivants, nous allons présenter les principaux concepts entourant l'analyse de contenu de manière générale, puis nous aborderons les travaux antérieurs existants sur ce sujet.

## 1.2-Présentation des travaux antérieurs existants sur la méthode

L'analyse de contenu permet de distinguer l'étude de la teneur et celle du style dans les communications. L'étude de la teneur concerne les thèmes et les opinions exprimés dans les discours, les éditoriaux, etc. Elle permet de décrire les différentes sortes de nouvelles diffusées ou la nature des opinions émises par les commentateurs. En revanche, le problème du "style" se concentre sur la structure particulière d'un message donné, en utilisant des symboles et des signes pour transmettre le sens (Harold Dwight Lasswell (1902-1978), s. d.). Harold Dwight Lasswell<sup>21</sup>, un politologue et théoricien de la communication, a influencé le développement de l'analyse de contenu en proposant une approche centrée sur l'étude du contenu des messages médiatiques. Il a encouragé l'analyse des thèmes, des symboles et des images pour comprendre

---

<sup>21</sup> *Harold Dwight Lasswell (1902-1978)*. (s. d.). <https://www.reseau-canope.fr/savoirscdi/societe-de-linformation/le-monde-du-livre-et-de-la-presse/histoire-du-livre-et-de-la-documentation/biographies/harold-dwight-lasswell-1902-1978.html>

l'influence des médias sur l'opinion publique et les processus politiques. Son ouvrage pertinent dans ce domaine est intitulé "The Analysis of Political Behavior: An Empirical Approach" (L'analyse du comportement politique : une approche empirique), publié en 1948 (Kariel, 1956).

D'après Bardin, l'analyse de contenu est « un ensemble de techniques d'analyse des communications visant, par des procédures systématiques et objectives de description du contenu des énoncés, à obtenir des indicateurs (quantitatifs ou non) permettant l'inférence de connaissances relatives aux conditions de production/réception (variables inférées) de ces énoncés » (Bardin, 1977, p. 43).<sup>24</sup> Selon : « Ghiglione et Beauvois (1980) « L'analyse de contenu n'est pas une pratique théorique et technique fermée sur elle-même, il s'agit d'un outil qui permet d'apporter des éléments de réponse aux problèmes qui ont suscité sa mise en œuvre et qui répond à des objectifs d'étude bien déterminés ». Pour A. Muchielli, l'analyse de contenu correspond à « un terme générique désignant l'ensemble des méthodes d'analyse de documents, le plus souvent textuel, permettant d'explicitier le ou les sens qui sont contenues et/ou le ou les manières dont ils parviennent à faire effet de sens » (Muchielli 1996, p. 36) ».<sup>25</sup>

L'analyse thématique, selon Lilian Negura dans son ouvrage de 2006 intitulé : « L'analyse de contenu dans l'étude des représentations sociales », vise à repérer les unités sémantiques qui composent le discours, puis à les catégoriser. Une fois cette catégorisation effectuée, il est important de prendre en compte la composante affective de ces unités sémantiques. Ainsi, l'analyse thématique peut être considérée comme un outil pour analyser les unités de base qui peuvent ensuite être classifiées en opinions, attitudes et stéréotypes.

---

<sup>24</sup> [https://theses.hal.science/tel-01144410/file/these\\_A\\_ZERAI\\_Zeineb\\_2011.pdf](https://theses.hal.science/tel-01144410/file/these_A_ZERAI_Zeineb_2011.pdf)

<sup>25</sup> Contributeurs aux projets Wikimedia. (2022). Analyse de contenu. *fr.wikipedia.org*. [https://fr.wikipedia.org/wiki/Analyse\\_de\\_contenu](https://fr.wikipedia.org/wiki/Analyse_de_contenu)

### 1.3- Les principaux concepts

**-Corpus** : Un corpus est un ensemble de pièces ou de documents liés à un domaine spécifique. Dans les sciences du langage, un corpus est utilisé comme base d'étude d'un phénomène linguistique. Il peut s'agir d'une collection de textes ayant une unité de genre ou d'époque. Le terme "corpus" est également utilisé dans d'autres domaines tels que l'histoire, la philologie et le droit, où des chercheurs travaillent sur des collections similaires pour leurs propres études. Un corpus peut être constitué de divers éléments tels que des textes, des mots ou des témoignages oraux enregistrés ou transcrits (Dalbera, 2002)<sup>26</sup>.

**-Occurrence et co-occurrence** : Dans l'analyse de corpus, on peut examiner les occurrences ou encore les fréquences d'un mot ou d'une expression pour étudier leur utilisation, leur contexte d'apparition, leurs co-occurrences avec d'autres mots, ou encore pour identifier des schémas ou des tendances linguistiques. L'analyse des co-occurrences cherche à extraire du texte des relations entre éléments du message, ou plus précisément elle s'attache à remarquer les présences simultanées (co-occurrence ou relation d'association) de deux ou plusieurs éléments dans une même unité de contexte, c'est-à-dire dans un fragment de message préalablement défini (Bardin, 2013)<sup>27</sup>.

**-Catégorisation** : « *Les catégories sont des rubriques ou classes qui rassemblent un groupe d'éléments (unités d'enregistrement dans le cas de l'analyse de contenu) sous un titre générique, rassemblement effectué en raison des caractères communs de ces éléments. Le critère de catégorisation peut être sémantique (catégories thématiques : par exemple, tous les thèmes signifiant l'anxiété seront rassemblés dans la catégorie « anxiété », alors que ceux qui*

---

<sup>26</sup> Dalbera, J. (2002). Le corpus entre données, analyse et théorie. *Corpus*, 1. <https://doi.org/10.4000/corpus.10>

<sup>27</sup> Bardin, L. (2013). Chapitre VI. L'analyse des relations. Dans : , L. Bardin, *L'analyse de contenu* (pp. 268-288). Paris cedex 14: Presses Universitaires de France.



*signifient la détente seront regroupés sous le titre conceptuel « détente »), syntaxique (les verbes, les adjectifs), lexicaux (classement des mots selon leur sens avec appariement des synonymes et des sens proches), expressifs (par exemple, catégories classant les divers troubles de langage), (Bardin, 2013) »<sup>28</sup>.*

## 2- Constitution et préparation du corpus

Dans le cadre de mon travail d'analyse de contenu portant sur la question de l'impact économique des données ouvertes J'ai eu à exploiter un corpus composite. Mon corpus est essentiellement composé des avis des experts, des avis des acteurs du domaine sur le volet économique des données ouvertes, de bilans sur l'ouverture des données, et de quelques entretiens faits avec les agents de services internes de la métropole lilloise. Nos motivations pour le choix d'un corpus composite sont :

Tout d'abord, un corpus composite permet d'obtenir une représentativité plus complète en rassemblant des données provenant de différentes sources ou domaines. Cela permet de prendre en compte une variété de perspectives et de points de vue, offrant ainsi une vision plus globale du sujet étudié. De plus, en incluant des données variées telles que des articles de presse, des rapports, des études et des témoignages, un corpus composite enrichit l'analyse en fournissant différents types d'informations et de perspectives sur le sujet. Un autre avantage de notre choix du corpus composite est la possibilité de valider les résultats à travers différentes sources et types de données. En comparant les informations provenant de différentes sources, il est possible de confirmer, nuancer ou remettre en question les résultats obtenus, renforçant ainsi la robustesse de l'analyse. De plus, un corpus composite permet d'explorer des connexions, des relations et des intersections entre différents domaines ou thématiques, favorisant une compréhension plus holistique et approfondie du sujet étudié. En incluant des données provenant de diverses disciplines ou secteurs, il devient possible d'aborder de manière plus complète des problématiques qui nécessitent une approche globale.

Pour constituer le corpus, j'ai choisi d'utiliser le moteur de recherche Google. Sur ce dernier, nous avons émis des requêtes que nous avons trouvées pertinentes et en lien avec notre problématique et nos hypothèses de départ. Nous avons émis trois requêtes :

- Avis d'expert impact économiques des données  
ouvertes
- Bilan données ouvertes

---

<sup>28</sup> Ibid

- Réutilisation des données ouvertes : état actuel

Dans le cadre de la dernière requête nous avons appliqué le filtre actualité. Les documents que nous considérons alors comme pertinents étaient les bilans, articles de presse, et éventuellement des articles scientifiques. Cette phase de recherche nous a conduit vers des documents qui n'entrent pas dans le cadre de notre étude. Nous avons parallèlement trouvé des articles qui cadrent avec notre seuil de pertinence, cependant l'accès au contenu en entier n'était pas possible. A titre d'illustration nous pouvons citer le journal "La gazette".

Les documents que nous avons jugés pertinents de notre recherche sur Google et inclus dans notre corpus sont :

- Avis d'expert : L'open data privé, le nouvel eldorado des entreprises ?
- Open Data : comment créer de la valeur à partir d'une donnée « gratuite »
- Open Data : un deal parfait ?
- Open Data définition : qu'est-ce que c'est ? À quoi ça sert ?
- Open data : de nouveaux jeux de données adaptés aux attentes des utilisateurs
- L'impact économique des données ouvertes : que savons-nous déjà ?
- Open Data définition : qu'est-ce que c'est ? À quoi ça sert ?
- La data au coeur d'un nouvel équilibre dans les échanges public-privé
- Coopération public – privé et données : comment et pourquoi les collectivités doivent reprendre la main
- Dix ans d'ouverture des données publiques : un bilan critique

S'ajoutent à ces documents, les travaux des étudiants du master 2 information documentation et la transcription de trois entretiens qualitatifs que nous avons menés dans le cadre de mon stage à Geriico. Ces entretiens avaient pour but d'étudier l'usage de la plateforme Open Data Mel par les services internes de la métropole Lilloise (MEL). Deux de ces entretiens ont été l'occasion pour nous, d'aborder d'une part les questions touchant à l'usage de la plateforme et d'autre part, les questions sur le volet économique notamment la réutilisation des données ouvertes. Le troisième entretien était strictement axé sur les enjeux économiques de l'open Data. En effet, l'interviewé est chef d'équipe du service public métropolitain de la donnée au sein de la MEL. Ainsi nous avons jugé bon d'avoir un entretien unique avec lui afin de mieux approfondir certaines questions relatives à l'impact économique des données ouvertes.

### 3- Analyse des données

L'analyse des données joue un rôle central dans notre étude, permettant de dévoiler des insights précieux et d'apporter une compréhension approfondie de notre corpus. Dans cette section, nous procéderons à une catégorisation rigoureuse et à une analyse lexicale approfondie de nos données. La première étape de notre analyse consistera en une catégorisation systématique de nos données. Nous examinerons attentivement les différents éléments du corpus afin d'identifier les thèmes, les motifs récurrents et les structures sous-jacentes. Cette catégorisation nous permettra de regrouper les données en fonction de leurs caractéristiques communes, facilitant ainsi l'identification des tendances et des relations entre les différents éléments. Une fois que nous aurons effectué la catégorisation, nous procéderons à une analyse lexicale approfondie de chaque catégorie. En examinant de près les mots et les expressions utilisés dans notre corpus, nous pourrions détecter les nuances sémantiques, les opinions implicites, les connotations et les liens conceptuels. Cette analyse lexicale nous permettra de saisir les différents discours, les points de vue et les cadres de référence présents dans nos données.

#### 3.1- Catégorisation

Dans notre travail, nous avons identifié certaines catégories afin de traiter cette problématique à partir de nos recherches et connaissances sur le domaine. Les catégories choisies sont pertinentes et illustrent parfaitement la notion d'impact économique des données ouvertes. Nous allons expliquer les tenants et aboutissants afin de justifier nos choix. En considérant le volet économique de l'open data, la qualité des jeux de données disponibles sur les plateformes institutionnelles joue un rôle essentiel. Une bonne qualité de données est essentielle et primordiale dans la réutilisation des données. Voici quelques critères à utiliser pour évaluer la qualité des jeux de données :

-Certains formats de fichier sont plus utilisés que d'autres. Les formats CSV (Comma-Separated Values) et JSON (JavaScript Object Notation) sont particulièrement répandus. Ces formats permettent de récupérer facilement les données et de les réutiliser selon les besoins. Ils sont préférables aux formats propriétaires tels que Excel. Il est donc nécessaire de veiller à une harmonisation des formats afin de faciliter la réutilisation.

-La structure d'un fichier est également importante. Les noms des propriétés doivent être explicites, faciles à comprendre et interprétables par les machines. Il est déconseillé d'utiliser des espaces, des accents ou des majuscules dans les noms des propriétés.

-Le contenu des données doit également être soigneusement considéré. Chaque donnée devrait être épurée autant que possible, avec un type de donnée simple (nombre, pourcentage, chaîne de caractères, date, etc.) et une signification claire en termes métier.

- Le respect des standards, des référentiels et des schémas déjà établis.: la présence de données et de colonnes pivot permettant de lier les données à un référentiel (par exemple, le code SIRET ou SIREN)<sup>29</sup>.

La valeur produite par les données et leur exploitation peut également impacter de façon significative la manière dont cette dernière peut ou sera utilisée. En favorisant la réutilisation des données, les initiatives d'open data visent à promouvoir la création de valeur pour les utilisateurs et la fourniture de services aux citoyens. Il existe trois formes de valeur des données : matière première, levier et actif stratégique. La valeur monétaire des données en tant que matière première est annulée avec les données ouvertes. La disponibilité gratuite de nombreuses sources de données réduit leur rareté et leur valeur monétaire. Les acteurs de l'open data se sont tournés vers des modèles économiques axés sur les services plutôt que sur la vente

---

<sup>29</sup> *Nos réflexions sur la qualité des données* - [data.gouv.fr](https://www.data.gouv.fr/fr/posts/nos-reflexions-sur-la-qualite-des-donnees/). (s. d.).  
<https://www.data.gouv.fr/fr/posts/nos-reflexions-sur-la-qualite-des-donnees/>

de données brutes. Les données en tant que levier offrent une valeur significative à travers l'amélioration des performances organisationnelles (Benyayer & Chignard, 2015)<sup>30</sup>.

La réutilisation est également l'un des aspects du volet économique à prendre en compte. En effet, selon la jurisprudence française, la réutilisation fait référence à l'utilisation « à des fins autres que celles de la mission de service public pour lesquelles les documents ont été élaborés ou détenus ». Le terme "réutilisation" est devenu courant dans l'écosystème français avec la directive européenne sur la réutilisation des informations détenues par le secteur public. Cette directive définit, entre autres, les domaines dans lesquels la réutilisation des informations publiques est autorisée, ainsi que ses conditions, notamment l'obligation de transparence quant au calcul des redevances. La présence d'une redevance doit être justifiée par les coûts liés à la recherche et au traitement des données engagés par l'administration.

*« Le secteur privé est l'un de ceux qui bénéficient de la donnée ouverte. Des entrepreneurs créent de la valeur en utilisant des données à vocation publique de la même façon qu'ils le font avec d'autres types de données. Premièrement, les entreprises utilisent les données à vocation publique pour améliorer leur fonctionnement. Par exemple, aux États-Unis, des commerçants combinent des données tirées de l'enquête sur les communautés américaines (American Community Survey) à leurs propres données commerciales pour répartir leurs stocks au niveau régional. Pour ce qui est de leur fonctionnement, les entreprises utilisent notamment des données sur les prix pour établir les salaires et les indemnités (World Development Report 2021 team, s. d.) »<sup>31</sup>.*

---

<sup>30</sup> Benyayer, L., & Chignard, S. (2015). Focus - Les enjeux économiques de l'ouverture des données : pas de marché, pas de valeur. *Informations sociales*, n° 191(5), 36. <https://doi.org/10.3917/inso.191.0036>

<sup>31</sup> World Development Report 2021 team. (s. d.). *Rapport sur le développement dans le monde 2021 : Des données au service d'une vie meilleure | Le potentiel des données ouvertes au service des entreprises*. <https://wdr2021.worldbank.org/fr/spotlights/the-potential-of-open-data-to-help-businesses/>

Catégories	Sous catégories
Qualité	Standard, format, structure
Valeur	Économie, marché
Réutilisation	Usage, utilisation
Secteur privé	Innovation, entrepreneuriat, applications

**Tableau 1 : Tableau récapitulatif des différents jeux de catégories**

## 3.2- Analyse lexicale

Une fois les catégories établies, nous avons travaillé d'une part avec un outil en ligne de lexicométrie nommé Ana Text 2.3. Il permet d'obtenir une représentation du vocabulaire d'un texte, trié par ordre de fréquence décroissante, et catégorisé en fonction des principales parties du discours (*Site du laboratoire LIDILEM - Université Grenoble Alpes - Plateformes*, s. d.)<sup>32</sup>. C'est ainsi que nous avons pu récupérer d'une part le corpus lemmatisé qui est disponible dans l'annexe de ce document. D'autre part, nous avons combiné l'analyse réalisée à l'aide de l'outil à celle faite à la main. Grâce à cette manière de procéder, nous avons pu poursuivre aisément l'analyse qui était la nôtre.

---

<sup>32</sup> *Site du laboratoire LIDILEM - Université Grenoble Alpes - Plateformes*. (s. d).  
<https://lidilem.univ-grenoble-alpes.fr/ressources/plateformes>

Rang	Lemme	Fréquence	CorpusRef (par million)	LogLike (spécificité)
1	donnée	919	18.595	12083.611
2	open	180	0.64	3150.293
3	dater	298	17.29	3044.757
4	collectivité	90	1.515	1254.949
5	plateforme	73	0.59	1157.555
6	ouverture	133	22.65	1046.694
7	entreprise	112	30.9	770.267
8	service	154	167.955	640.481
9	acteur	114	62.325	627.590
10	économique	72	13.865	548.179
11	usage	79	31.575	484.148
12	impact	53	6.58	452.223
13	numérique	46	3.37	445.296
14	usager	35	0.955	425.054
15	information	82	56.95	412.703
16	ouvert	89	84.77	393.228
17	public	92	99.875	383.323
18	permettre	105	165.86	362.904
19	citoyen	54	20.5	336.484
20	utilisation	35	4.12	302.591
21	exemple	78	102.405	296.700
22	publier	52	25.905	295.790
23	jeu	96	186.655	294.839
24	utilisateur	26	1.125	283.719
25	application	39	9.625	277.016

**Tableau 2 : analyse lexicale (les lemmes spécifiques)<sup>33</sup>**

D'une part, de nos résultats obtenus à savoir ceux des Lemmes spécifiques on se rend compte que dans le corpus les éléments qui reviennent le plus souvent et qui font écho à nos catégories de départ sont : « entrepris », « économique », « utilisation », « usage », « valeur », « qualité », « réutiliser », « impact », « application », « exemple ».

D'autre part, de notre analyse faite à la main, il ressort que la qualité est une catégorie qui revient beaucoup. De plus, au niveau des entretiens menés auprès des services internes de la métropole de Lille, c'est également, une préoccupation présente. Il existe un nombre important d'exemples d'innovation et ou de création d'entreprises par le biais de réutilisation de données ouvertes. Les domaines tels que la mobilité ou encore l'environnement semblent se démarquer des autres en termes de réutilisations des données. De tout ce qui précède, le mot "entreprise" apparaît 112 fois dans le corpus, « Usage » 79 fois, « impact » 53 fois, « utilisateur » 26 fois, « économique » 72 fois, « application » 39 fois.

<sup>33</sup> Voir Annexe

### 3.3 - Analyse Qualitative

D'après nos entretiens lorsqu'on demande aux professionnels du domaine en particulier les agents de la métropole lilloise, de citer des domaines où la réutilisation des données ouvertes pourrait avoir un impact significatif, les transports et la mobilité émergent comme des secteurs clés. L'intervenant numéro 5 s'exprime en ces termes : « *Je pense qu'il y a quand même de la créativité, on a des talents en France qui sont en capacité à partir de la donnée pour autant qu'elle soit de qualité, ils sont en mesure de l'exploiter et de développer différents services qui vont répondre là également aux besoins de leurs clients, leurs prospects comme on dit pour eux, et donc là voilà ça va générer de la création d'emploi. Une réussite dont je suis particulièrement fier est l'expérimentation liée à l'ouverture des données captées par des compteurs d'usagers dans le cadre des déplacements, que ce soit à pied, à vélo, etc. À partir de ces données, nous avons pu réguler les feux tricolores, en donnant la priorité aux véhicules en fonction de certaines règles algorithmiques. Les données collectées, traitées et rendues publiques ont suscité un grand intérêt de la part des gens. Des étudiants en école d'informatique ont développé une expérimentation basée sur ces données et ont compris pourquoi la durée des feux tricolores variait. Suite à cette expérience réussie, ces étudiants ont créé une start-up et développé le produit, qu'ils ont ensuite proposé à d'autres collectivités. Aujourd'hui, ils en vivent et ont réussi grâce à cette initiative.* » C'est dire que les données ouvertes ont un fort potentiel et permettent aisément aux acteurs de s'en servir pour innover en matière de technologie à condition que les données précédemment récoltées soient de qualité. Dans le même ordre d'idée, l'article *Open Data définition : qu'est-ce que c'est ? À quoi ça sert ?* identifie par exemple Google Traduction comme utilisatrice d'innombrables documents de l'union européenne traduits dans tous les langages européens pour développer des API et ses algorithmes de traduction et améliorer son service. A titre d'exemple l'intervenant numéro 1 cite Google, city mapper, des bureaux d'études, Waze, les promoteurs immobiliers.

L'Open Data a un impact qui va au-delà de la valeur pécuniaire. Elle améliore la qualité de vie sur les plans financier, social, culturel, et bien d'autres encore. Elle facilite l'accès aux services de santé et favorise la découverte de nouveaux traitements pour les maladies. Elle offre également une meilleure compréhension des gouvernements et facilite les déplacements. De nouvelles opportunités et possibilités sont constamment découvertes grâce à l'Open Data. Pour illustrer ces propos l'intervenant numéro 3 s'exprime en ces termes : « *Pour les bureaux d'études en urbanisme, les données ouvertes permettent de se poser certaines questions : il y a quoi dans ce quartier ? là il y a un arrêt de bus et là il y a le train, la fréquence du bus qui passe à tel arrêt et puis il y a la fréquence du tour, il Ya du stationnement payant là-bas, gratuit ici en zone bleue là-bas. Sans avoir*



*forcément besoin de se déplacer, c'est un gain de temps pour énormément de secteur économique. Avant tout se faisait à la main, aujourd'hui forcément quelqu'un l'a fait. En tout cas nous on produit de la donnée publique pour qu'on s'en serve et qu'elle soit utile. Et après ils sont saisis comme ils veulent. Dans quel domaine particulier ? je dirai tous, le commerce, je veux installer mon commerce je regarde et je sais que finalement là aujourd'hui il n'y a rien mais demain un tram donc je peux avoir une grande terrasse, ce qui peut arriver sur plein de considération. Ou alors il y a un tram je n'ai pas envie de me mettre là parce qu'il y aura des travaux pendant 5 ans et mon modèle économique, il ne survivra pas. »* Ainsi expliqué les données ouvertes permettent donc de faire des estimations et prendre de manière éclairée des décisions positives qui impacteront positivement le devenir d'une entreprise. On peut donc aisément se servir des données ouvertes dans une procédure d'étude de marché.

Selon Alexandre GRASSIGNY, Chargé de mission - La Fabrique de la Cité dans l'article intitulé : « *La data au cœur d'un nouvel équilibre dans les échanges public-privé* ». Le secteur public a encore un travail de mise en qualité de ses données comme des canaux de diffusion afin de susciter plus d'adhésion. Certains échanges sont plutôt dictés par le privé, qui s'appuie sur sa légitimité et la valeur ajoutée du service fourni, via la vente ou l'échange de données. D'autres sont plutôt aux mains du public au travers des régies de données ou de la mise en place de règles et conditions d'activité pour les data-services. Aucun modèle ne s'impose pour l'instant et la plupart garde un caractère expérimental. C'est dire donc qu'en fonction des intérêts poursuivis, il y aurait un acteur qui prime sur l'autre. Cependant il semblerait que le secteur privé ait plus ce privilège. C'est d'ailleurs ce qu'explique Pierre-Alexandre Conte dans l'article « *Coopération public – privé et données : comment et pourquoi les collectivités doivent reprendre la main* ». Pour lui, les acteurs privés ont généralement la mainmise sur les échanges avec le secteur public. Ils dictent la nature de la coopération et ne donnent finalement que peu de garanties en retour. Certaines sociétés se contentent ainsi de vendre les données dont elles disposent. L'intervenant numéro 2 dit également à ce propos : *l'autre frein c'était sur la relation avec ceux avec qui on était en contrat. Obligation d'ouverture des données encore une fois sur Ilévia. Au moment de la disposition d'ouverture des données, ce n'était pas prévu au contrat initial donc ça a provoqué dans les relations avec eux des réticences, d'une part ça générait des coûts chez eux, mise en forme, des coûts sur lesquels il n'existait pas forcément de retour en termes de recette et culturel il y a quand même une perte de pouvoir, et de souveraineté sur la donnée, sur l'exploitation. Il y a quand même, une exploitation, se dire qu'on a moins la main sur tout le cycle d'exploitation de son réseau y compris la donnée* ». Simon Chignard, data editor chez Etalab et auteur de plusieurs livres sur les données, souligne la différence entre les acteurs privés et les collectivités en termes d'échelles et de temps d'action. Les

collectivités ont une perspective territoriale, tandis que des entreprises comme Waze, Strava ou Uber ont des stratégies globales, qui peuvent évoluer pour des raisons qui ne sont pas liées spécifiquement aux territoires. Isabelle Pellerin, vice-présidente de Rennes Métropole, confirme ces propos en mentionnant qu'il est difficile pour les collectivités d'être associées à la gouvernance au-delà de l'expérimentation. Une fois que les entreprises privées ont accès aux données et les utilisent, elles valorisent ces données et mènent leur activité indépendamment des collectivités.

### Partie III : Exploration des résultats

L'exploration des résultats sera guidée par une interprétation contextuelle et une

réflexion critique. Tout d'abord, nous entamerons une discussion approfondie des résultats obtenus. Nous analyserons les tendances, les corrélations et les différences significatives qui ont émergé de nos données. Nous examinerons également les éventuelles contradictions ou les résultats surprenants qui pourraient nécessiter une exploration plus approfondie. Ensuite, nous procéderons à l'interprétation des résultats en lien avec les connaissances existantes et les théories pertinentes. Cette interprétation nous permettra de formuler des hypothèses explicatives et d'identifier les facteurs clés qui influencent les résultats observés.

Enfin, nous adopterons une approche critique pour évaluer la validité et la portée de nos résultats. Nous examinerons attentivement les limites de notre méthodologie, les biais potentiels et les variables non mesurées qui pourraient avoir influencé nos conclusions. Cette réflexion critique nous aidera à évaluer la fiabilité de nos résultats et à prendre en compte les nuances nécessaires lors de l'interprétation.

## 1- Interprétation des résultats

Les résultats obtenus confirment partiellement nos hypothèses. D'une part, notre analyse de corpus a révélé que l'ouverture des données stimule effectivement l'innovation et la création d'entreprises. Les données ouvertes sont considérées comme une ressource essentielle permettant aux utilisateurs et aux opérateurs économiques de générer de nouvelles idées et de lancer des start-ups. Cette utilisation des données ouvertes favorise ainsi l'innovation dans différents domaines.

D'autre part, en ce qui concerne l'évaluation précise des risques et la prise de décisions éclairées, bien que nous n'ayons pas pu obtenir des éléments concrets dans notre analyse, nous avons pu déduire que l'accès aux données ouvertes peut jouer un rôle crucial. Par exemple, dans le domaine de l'immobilier, l'utilisation des données ouvertes peut permettre d'identifier les zones à risques et d'aider à la prise de décisions éclairées. Cela peut conduire également à la création d'espaces urbains durables et adaptés aux besoins de la population, ainsi qu'à des choix d'investissement plus informés pour les promoteurs.

Cependant, notre troisième hypothèse concernant l'émergence de partenariats public-privé et de la société civile comme moteurs de l'économie semble ne pas se confirmer. L'analyse suggère qu'il existe encore des obstacles et des différences d'intérêts qui empêchent une réelle collaboration entre les acteurs publics et privés. Bien que l'idéal soit d'établir des partenariats solides pour favoriser le développement économique, la réalité actuelle semble éloignée de cette vision.

Pris ensemble, nos résultats suggèrent que les données ouvertes peuvent stimuler l'innovation et faciliter la prise de décision éclairée dans certains domaines, tels que l'immobilier. Cependant, des obstacles à des partenariats public-privé efficaces subsistent. Il reste donc des aspects à explorer et à approfondir dans le domaine de l'impact économique des données ouvertes, notamment pour rechercher des solutions pour surmonter ces barrières et faciliter une coopération fructueuse entre les différents acteurs.

## 2- Réflexion critique sur les résultats

Dans cette section, nous examinerons attentivement les résultats à la lumière de certains éléments essentiels, notamment la discussion des biais potentiels et des limites d'interprétation, ainsi que des suggestions pour de futures recherches ou des approfondissements. L'une des considérations clés de notre réflexion critique porte sur l'identification et la discussion des biais potentiels qui pourraient avoir influencé nos résultats. Nous analyserons de manière critique les éventuels biais méthodologiques ou de sélection qui pourraient avoir affecté la validité ou la représentativité de notre échantillon.

Nous reconnaissons que nos conclusions sont basées sur des données spécifiques, une méthodologie particulière et une approche analytique donnée. Il est essentiel d'explorer les éventuelles lacunes, les aspects non mesurés ou les variables omises qui pourraient limiter la validité de nos conclusions. Cette discussion critique des limites de notre interprétation nous permettra d'adopter une perspective plus équilibrée et de reconnaître les zones d'incertitude.

## 2.1- Discussion des biais potentiels et des limites de l'interprétation

Dans notre méthodologie de recherche, il est important de reconnaître que le choix d'un corpus composite peut potentiellement introduire des biais dans les résultats. En combinant l'utilisation de l'outil en ligne Ana text 2.3 pour une analyse automatique et une analyse manuelle, certains biais peuvent se manifester. Tout d'abord, le choix d'un corpus composite peut engendrer un biais de sélection. Malgré nos efforts pour construire un corpus représentatif et équilibré en incluant différentes sources et perspectives, il est possible que certaines données ou certains domaines soient privilégiés ou négligés, ce qui pourrait influencer les résultats obtenus. Les effets du biais de sélection de corpus sur l'analyse linguistique ont été examinés dans une étude descriptive, en particulier en relation avec l'utilisation du programme *Linguistic Inquiry and Word Count* (LIWC) dans la recherche politique. LIWC a été utilisé pour explorer plusieurs phénomènes, notamment la détection de la tromperie, la radicalisation, la diplomatie, le populisme et la popularité présidentielle (Windsor et al., 2019)<sup>34</sup>.

L'utilisation de l'outil en ligne Ana text 2.3 pour l'analyse automatique peut également introduire des biais. Les résultats de l'outil dépendent de son algorithme et de sa capacité à traiter et interpréter les données. Il est essentiel de prendre en compte les limites et les biais potentiels associés à cet outil. Par exemple, il peut y avoir des difficultés à saisir certains contextes ou subtilités linguistiques, ce qui peut conduire à des résultats partiels ou inexacts. Aussi, les algorithmes utilisés par cet outil peuvent ne pas être parfaitement adaptés à tous les types de texte ou de langage, ce qui peut conduire à une analyse automatisée imprécise. Il existe également un risque de biais lorsque l'analyse est effectuée manuellement. Les analystes peuvent avoir leurs propres préférences, interprétations subjectives et limitations personnelles, qui peuvent affecter les résultats. Par exemple, différentes personnes peuvent catégoriser les mots légèrement différemment en fonction de leur compréhension et de leur interprétation des parties du discours. De plus, l'analyse manuelle est sujette à des erreurs humaines telles que des biais de lecture, des omissions ou des interprétations erronées.

---

<sup>34</sup> Windsor, L. C., Cupit, J., & Windsor, A. (2019). Automated content analysis across six languages. *PLOS ONE*, 14(11), e0224425. <https://doi.org/10.1371/journal.pone.0224425>

En tenant compte des facteurs mentionnés précédemment, nous reconnaissons que nos travaux de recherche peuvent être potentiellement biaisés. Les choix que nous avons fait, dans l'utilisation d'outils d'analyse lexicale en ligne en combinaison avec l'analyse manuelle, et la composition du corpus composite, influent sur les résultats obtenus. Il est important de souligner que la présence de biais dans notre travail de recherche ne nie pas sa valeur ou son importance. Les biais sont inévitables dans toute méthode d'analyse et de recherche, cependant pour atténuer ces biais, nous avons pris des mesures telles qu'une sélection rigoureuse des sources de données, la validation croisée des résultats obtenus à partir de différentes sources de données.

## 2.2- Suggestions pour de futures recherches ou approfondissements

Bien que nous ayons exploré les avantages potentiels des données ouvertes, en particulier leur intérêt économique et effectué une analyse de corpus, il reste encore beaucoup à découvrir et à explorer. Les suggestions suivantes offrent des pistes intéressantes pour de futures recherches. Combiner des méthodes quantitatives et qualitatives dans l'étude de l'impact économique des données ouvertes permet d'obtenir une perspective plus complète et nuancée des résultats. Les méthodes quantitatives fournissent des données chiffrées et mesurables sur les indicateurs économiques clés tels que la croissance économique, l'emploi, les investissements, etc. Ces données quantitatives offrent une vue d'ensemble et permettent des comparaisons entre différentes régions, secteurs ou initiatives.

Parallèlement, les méthodes qualitatives, telles que les études de cas, les entretiens ou les observations, permettent de comprendre les mécanismes sous-jacents et les expériences des acteurs impliqués dans l'utilisation des données ouvertes. Elles fournissent une compréhension approfondie des contextes spécifiques, des défis rencontrés, des bénéfices perçus et des répercussions socio-économiques réelles. Les études de cas qualitatives peuvent également mettre en évidence des facteurs clés de réussite ou d'échec dans la mise en œuvre des initiatives de données ouvertes.

En combinant ces approches, les chercheurs peuvent établir des liens entre les données quantitatives et les aspects qualitatifs, en fournissant des explications et des interprétations plus

approfondies des résultats économiques observés. Par exemple, une analyse quantitative peut montrer une corrélation entre l'ouverture des données et la croissance économique, tandis qu'une étude de cas qualitative peut expliquer les mécanismes par lesquels les entreprises utilisent ces données pour innover, créer des emplois ou optimiser leurs opérations. Elles offrent la possibilité de recueillir des informations sur les défis et les opportunités spécifiques rencontrés dans différents contextes géographiques, économiques et institutionnels. Cette approche enrichit la compréhension de l'impact économique des données ouvertes en fournissant des informations détaillées sur les mécanismes, les expériences et les défis rencontrés par les acteurs impliqués. Dans le cadre des études qualitatives, il serait bénéfique de concentrer notre attention sur les acteurs clés du domaine.

En ciblant spécifiquement les parties prenantes directement impliquées dans l'utilisation des données ouvertes, nous pourrions recueillir des informations concrètes et des perspectives approfondies sur l'impact économique de ces données. En interrogeant les acteurs du secteur, tels que les entrepreneurs, les chercheurs, les décideurs politiques et les représentants d'organisations pertinentes, nous pourrions obtenir des connaissances de première main sur les mécanismes, les opportunités et les défis liés à l'utilisation des données ouvertes à des fins économiques. Leurs expériences et leurs observations pourraient fournir des exemples concrets, des études de cas et des points de vue essentiels pour compléter les données quantitatives et enrichir notre compréhension globale.

## Conclusion

Cette étude nous a permis de mieux comprendre l'impact économique des données ouvertes et leur potentiel d'innovation, de prise de décision éclairée et de développement économique. Les résultats de notre analyse de corpus confirment dans le même ordre d'idée qu'en pratique les données ouvertes fournissent une matière première précieuse pour la génération de nouvelles idées et l'amélioration des services existants, contribuant ainsi à la stimulation de l'innovation et de la création d'entreprises. De plus, l'accès aux données ouvertes permet des évaluations des risques plus précises, permettant aux différents acteurs de prendre des décisions plus éclairées en matière d'urbanisme, d'immobilier et d'autres domaines où l'accès aux données joue un rôle clé dans la planification et la prise de décision stratégique.

Il est important de reconnaître les limites et les biais potentiels associés à notre méthodologie et plus particulièrement à l'utilisation de corpus composés et d'analyses lexicosémantiques. Nous soulignons la nécessité de méthodes mixtes, combinant des approches quantitatives et qualitatives, et croisant les résultats avec d'autres sources d'information pour une compréhension plus complète et nuancée de l'impact des données ouvertes. Pour les recherches futures, il serait intéressant d'approfondir la recherche qualitative en ciblant davantage d'acteurs clés dans le domaine, tels que les utilisateurs de données ouvertes, les entrepreneurs et les décideurs politiques. Une enquête approfondie sur leurs expériences, leurs défis et leurs besoins fournira un aperçu des mécanismes sous-jacents à l'impact économique des données ouvertes. L'étude des données ouvertes est un domaine en pleine expansion qui offre de nombreuses opportunités de recherche et d'approfondissement.

L'étude des données ouvertes est un domaine en constante évolution, offrant de nombreuses opportunités de recherche et d'approfondissement. L'exploitation efficace des données ouvertes demande une collaboration entre les acteurs du secteur public et du secteur privé, ainsi qu'une attention particulière aux enjeux de transparence, de gouvernance et de protection des données.



# Bibliographie

- **Articles scientifiques**

Benyayer, L., & Chignard, S. (2015). Focus - Les enjeux économiques de l'ouverture des données : pas de marché, pas de valeur. *Informations sociales*, n° 191(5), 36. <https://doi.org/10.3917/inso.191.0036>

Dalbera, J. (2002). Le corpus entre données, analyse et théorie. *Corpus*, 1. <https://doi.org/10.4000/corpus.10>

Droit à l'information. (2022). *Articles*. <https://www.unesco.org/fr/right-information#:~:text=L'UNESCO%20d%C3%A9fend%20l'acc%C3%A8s,des%20droits%20de%20l'homme>.

Garric, N., & Longhi, J. (2012). L'analyse de corpus face à l'hétérogénéité des données : d'une difficulté méthodologique à une nécessité épistémologique. *Langages*, n° 187(3), 3-11. <https://doi.org/10.3917/lang.187.0003>

Kariel, H. S. (1956). Political Science in the United States : Reflections on One of its Trends. *Political Studies*, 4(2), 113-127. <https://doi.org/10.1111/j.1467-9248.1956.tb00947.x>

Negura, L. (2006). L'analyse de contenu dans l'étude des représentations sociales. *Sociologies*. <https://doi.org/10.4000/sociologies.993>

Warusfel, B. (2018a). Enjeux et limites de l'ouverture des données en matière de sécurité et de défense. *Revue française d'administration publique*, 167(3), 551. <https://doi.org/10.3917/rfap.167.0551>

Warusfel, B. (2018b). Enjeux et limites de l'ouverture des données en matière de sécurité et de défense. *Revue française d'administration publique*, 167(3), 551. <https://doi.org/10.3917/rfap.167.0551>

Windsor, L. C., Cupit, J., & Windsor, A. (2019). Automated content analysis across six languages. *PLOS ONE*, 14(11), e0224425. <https://doi.org/10.1371/journal.pone.0224425>

- **Ressources documentaires**

Digitale, U. (2018, 19 décembre). [Avis d'expert] L'open data privé, le nouvel eldorado des entreprises ? *www.usine-digitale.fr*. <https://www.usine-digitale.fr/article/avis-d-expert-l-open-data-prive-le-nouvel-eldorado-des-entreprises.N781694>

Dor, F. (2020, 10 janvier). Les 5 chiffres à savoir sur Citymapper, l'appli qui cartonne. *Les Echos Start*. <https://start.lesechos.fr/innovations-startups/tech-futur/les-5-chiffres-a-savoir-sur-citymapper-lappli-qui-cartonne-1175380>

Jacqué, P. (2016, 9 avril). Citymapper, cette start-up qui agace la RATP. *Le Monde.fr*.  
[https://www.lemonde.fr/economie/article/2016/04/09/citymapper-cette-start-up-qui-agace-la-ratp\\_4899122\\_3234.html](https://www.lemonde.fr/economie/article/2016/04/09/citymapper-cette-start-up-qui-agace-la-ratp_4899122_3234.html)

Marzolf, E. (2023, 9 mai). Météo France tarde à faire sa révolution de l’“open data”.  
*Météo France tarde à faire sa révolution de l’“open data” | À la une | Acteurs Publics*.  
<https://acteurspublics.fr/articles/meteo-france-tarde-a-faire-sa-revolution-de-l-open-data>

Publications Office of the EU. (2022, 14 décembre). The Open Data Maturity Report 2022 is out ! *Publications Office of the EU*. <https://data.europa.eu/en/news-events/news/open-data-maturity-report-2022-out>

<https://finq.org/wp-content/uploads/2020/02/Open-Data-impact-bilan-2010-2018-de-l-open-data.pdf>

## Sitographie

Agueke. (2023a, mai 5). *Données ouvertes institutionnelles*. ODEP.  
<https://odep.hypotheses.org/789>

Akbaraly, M. (s. d.-a). *Fracture numérique : Quelles Conséquences sur l’Accessibilité ?* <https://blog.ipedis.com/fracture-numerique-queelles-consequences-sur-laccessibilite>

*Comprendre ce qu’est une API*. (s. d.-b). <https://www.redhat.com/fr/topics/api>

De l’Ardèche, S. G. P. L. C. D. (2019a, décembre 15). *Chapitre 1 Les grands principes d’une démarche d’open data | Rapport sur l’open data*.  
<https://dataactivist.coop/ardeche/rapport/partie1.html>

De Vasconcelos Aguiar, J. P., & Muller, C. (2022a, juillet 11). *Qu'est-ce que la fracture numérique ?* - *Internet Society*. Internet Society.

<https://www.internetsociety.org/fr/blog/2022/03/quest-ce-que-la-fracture-numerique/>

Dymytrova, V. (2020a, août 5). *Du traitement des données à la création de valeur : comprendre les pratiques professionnelles des réutilisateurs des données ouvertes*.

<https://hal.science/hal-02913346>

Harold Dwight Lasswell (1902-1978). (s. d.-a). [https://www.reseau-](https://www.reseau-canope.fr/savoirscdi/societe-de-linformation/le-monde-du-livre-et-de-la-)

[canope.fr/savoirscdi/societe-de-linformation/le-monde-du-livre-et-de-la-  
presse/histoire-du-livre-et-de-la-documentation/biographies/harold-dwight-lasswell-  
1902-1978.html](https://www.reseau-canope.fr/savoirscdi/societe-de-linformation/le-monde-du-livre-et-de-la-presse/histoire-du-livre-et-de-la-documentation/biographies/harold-dwight-lasswell-1902-1978.html)

*Le règlement général sur la protection des données - RGPD*. (s. d.-a). CNIL.

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

*L'ouverture des données publiques*. (s. d.-a). Gouvernement.fr.

<https://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>

Meszaros, B. (2015a, juin 10). *Livre blanc sur les données ouvertes*.

<https://inria.hal.science/AO-DROIT/halshs-01162692v1>

*Nos réflexions sur la qualité des données* - *data.gouv.fr*. (s. d.-a).

<https://www.data.gouv.fr/fr/posts/nos-reflexions-sur-la-qualite-des-donnees/>

*Open Data à Paris : et maintenant ?* (s. d.-a). Citymapper.

<https://citymapper.com/news/1307/open-data-a-paris-et-maintenant>

*Site du laboratoire LIDILEM - Université Grenoble Alpes - Plateformes*. (s. d.-a).

<https://lidilem.univ-grenoble-alpes.fr/ressources/plateformes>

Van Cuyck, A. (2013a, mars 14). *L'open data comme nouvelle forme de gouvernance numérique : enjeux, marchés, modèles, idéologies*. <https://hal.science/hal-00963337>

World Development Report 2021 team. (s. d.-a). *Rapport sur le développement dans le monde 2021 : Des données au service d'une vie meilleure | Le potentiel des données ouvertes au service des entreprises*. <https://wdr2021.worldbank.org/fr/spotlights/the-potential-of-open-data-to-help-businesses/>

[Wikipedia contributors. \(2023a\). Aspects of the Theory of Syntax. Wikipedia. https://en.wikipedia.org/wiki/Aspects\\_of\\_the\\_Theory\\_of\\_Syntax](https://en.wikipedia.org/wiki/Aspects_of_the_Theory_of_Syntax)

[Wikipedia contributors. \(2023b\). Syntactic Structures. Wikipedia. https://en.wikipedia.org/wiki/Syntactic\\_Structures](https://en.wikipedia.org/wiki/Syntactic_Structures)

# Annexes

## Analyse lexicométrique

Rang	Lemme	Fréquence
1	donnée	919
2	service	154
3	ouverture	133
4	acteur	114
5	entreprise	112
6	jeu	96
7	collectivité	90
8	information	82
9	usage	79
10	exemple	78
11	plateforme	67
12	valeur	59
13	citoyen	54
14	qualité	54
15	impact	53
16	territoire	48
17	question	47
18	temps	45
19	fait	44
20	portail	42
21	accès	41
22	administration	40
23	loi	40
24	application	39
25	cas	38
26	manière	38
27	besoin	38
28	<u>gouv</u>	37
29	terme	36
--	--	--

28	<u>gouv</u>	37
29	terme	36
30	travail	36
31	base	36
32	fr	36
33	utilisation	35
34	culture	35
35	mise	34
36	transparence	33
37	sujet	32
38	usager	32
39	transport	31
40	société	31
41	outil	31
42	site	31
43	réutilisation	30
44	demande	30
45	disposition	30

Rang	Lemme	Fréquence
46	organisation	29
47	démarche	29
48	personne	28
49	mission	28
50	année	28
51	ville	28
52	bibliothèque	28

## Guide d'entretien 1

### *Introduction*

Présentation de l'objectif de l'enquête : Nous réalisons une enquête auprès des services de la Métropole Européenne de Lille afin de pouvoir évaluer les différentes utilisations de la plateforme Open Data MEL.

Rappel de la confidentialité et de l'anonymat des réponses : Veuillez noter que vos réponses sont confidentielles et anonymes. Toutes les informations échangées lors de cette conversation sont traitées de manière anonyme et sont utilisées uniquement dans le cadre de cette étude.

Demande de consentement pour l'enregistrement de l'entretien : Avant de poursuivre, j'aimerais vous demander votre consentement pour l'enregistrement de cet entretien. Soyez assuré(e) que toute information enregistrée sera traitée de manière confidentielle et utilisée uniquement dans le cadre de notre étude.

Profil :

- Quel(s) service(s) de la MEL représentez-vous ?
- Quel poste occupez-vous ?
- Depuis combien de temps êtes-vous à ce poste ?

**Connaissez-vous la loi du numérique de 2016 sur l'ouverture des données par les institutions ? \*OUI\***

**Quels sont les principaux objectifs de cette loi selon votre compréhension ?**

**Avez-vous remarqué des changements ou des impacts dans votre entreprise ou dans votre secteur suite à l'adoption de cette loi ?**

Utilisation de la plateforme :

**Connaissez-vous la plateforme open data mel : si oui**

Comment avez-vous connu cette plateforme ?

Quels sont les avantages que vous retirez de l'utilisation de cette plateforme ?

Quelles sont les fonctionnalités de la plateforme que vous utilisez le plus souvent ? Pourquoi ?

**Si vous déposez des données, à quelle fréquence le faites-vous ?**

**Si vous récupérez des données, à quelle fréquence le faites-vous ?**

Et sur quelle(s) thématique(s) ?

**Que pensez-vous globalement de la plateforme (données, services, accès...) ?**

**Données :**

Quelle est votre opinion sur la qualité et la pertinence des données disponibles sur la plateforme ?

Dans l'ensemble, pensez-vous que les données fournies sur la plateforme répondent aux besoins de votre service ?

**Services :**

Comment évaluez-vous la gamme de services proposés par la plateforme ? Sont-ils adaptés aux besoins de votre service ?

Avez-vous trouvé les fonctionnalités de la plateforme utiles et faciles à utiliser ?

Ces données sont-elles utiles à votre service ? Si oui, dans quel(s) usage(s) ?

**Accès :**

Trouvez-vous qu'il est facile de naviguer et de trouver les informations dont vous avez besoin ?

Avez-vous rencontré des difficultés techniques lors de l'accès à la plateforme ou de la récupération des données ?



Quels sont vos commentaires sur la convivialité et l'ergonomie de la plateforme en termes de facilité d'utilisation et de navigation ?

**Connaissez-vous la plateforme open data mel : si non**

Pourquoi n'utilisez-vous pas cette plateforme ?

Seriez-vous intéressé par le fait d'apprendre à l'utiliser ?

Savez-vous si quelqu'un s'occupe de la donnée ouverte dans votre service ?

**Attentes et besoins de l'outil et des gens :**

Avez-vous des besoins en termes de jeux de données de l'institution, dont vous avez besoin et qui ne sont pas disponibles ? Si oui, lesquels ?

Avez-vous des jeux de données que votre service produit, et que vous pourriez d'après vous rendre disponible sur la plateforme ? Si oui, lesquels ?

Pensez-vous qu'il serait judicieux de sensibiliser, documenter, ou de former davantage les différents acteurs des collectivités territoriales sur l'ouverture des données ?

Former

Documenter

Accompagner au dépôt

Autre

**En fonction de la réponse :**

**Sensibiliser**

**Dans votre perception, quels sont les avantages de sensibiliser les acteurs des collectivités territoriales à l'ouverture des données ?**

**En quoi cette sensibilisation pourrait-elle favoriser une meilleure compréhension de l'enjeu et susciter un engagement plus fort dans ce domaine ?**

**Quelle est votre opinion sur la nécessité de former les acteurs des collectivités territoriales à l'ouverture des données ?**

**En quoi cette formation pourrait-elle être bénéfique pour ces acteurs et pour les collectivités territoriales dans leur ensemble ?**

**Quels types de compétences ou connaissances spécifiques devraient être abordés lors de ces formations ?**

### **Documenter**

**Selon vous, quel rôle joue la documentation dans la promotion de l'ouverture des données au sein des collectivités territoriales ?**

**Accompagnement au dépôt :**

**Quels types d'accompagnement pourraient être mis en place pour soutenir les acteurs territoriaux dans le dépôt de leurs données ouvertes ? Cela pourrait inclure des formations, des guides pratiques, des outils techniques, des conseils personnalisés, etc.**

**Existe-t-il des partenaires externes ou des ressources spécifiques (organismes publics, associations, experts, etc.) qui pourraient être mobilisés pour fournir un accompagnement efficace aux acteurs territoriaux dans le processus de dépôt des données ouvertes ?**

Sur la base de vos connaissances et de votre expérience, quelles seraient vos recommandations pour renforcer la sensibilisation, la documentation et la formation des acteurs des collectivités territoriales en ce qui concerne l'ouverture des données ? Quelles mesures concrètes pourraient être prises pour promouvoir une culture d'ouverture des données au sein des collectivités territoriales ?

### **Pour Tous**

En dehors de la plateforme Open Data Mel connaissez-vous d'autres plateformes de type data.gouv , la plateforme hauts-de -France ?

Les utilisez-vous ? **si oui Quelles sont ces plateformes ?**

**Réutilisation et impacts économiques :**

Avez-vous déjà réutilisé des données ouvertes dans le cadre de votre travail ou de vos projets ?

Pouvez-vous identifier des domaines où la réutilisation des données ouvertes pourrait avoir un impact significatif ?

Avez-vous connaissances d'entreprises ou d'organisations qui ont pu tirer parti des données ouvertes pour créer de la valeur économique ?

Pensez-vous que la donnée ouverte puisse être génératrice d'emplois ? **Oui ? pouvez-vous aller au bout de vos idées svp ?**

### **Limites**

Quels sont les principaux obstacles ou défis auxquels vous avez pu être confronté en voulant vous conformer à la politique d'ouverture des données et ou réutilisation ?

### Conclusion

Merci pour votre précieuse contribution à notre étude. Nous nous engageons à utiliser les informations recueillies de manière confidentielle et à les utiliser uniquement à des fins de recherche.

### **Guide d'entretien 2**

Rappel de la confidentialité et de l'anonymat des réponses : Veuillez noter que vos réponses sont confidentielles et anonymes. Toutes les informations échangées lors de ces conversations sont traitées de manière anonyme et sont utilisées uniquement dans le cadre de cette étude.

Demande de consentement pour l'enregistrement de l'entretien : Avant de poursuivre, j'aimerais vous demander votre consentement pour l'enregistrement de cet entretien.

Soyez assuré(e) que toute information enregistrée sera traitée de manière confidentielle et utilisée uniquement dans le cadre de notre étude.

Profil :

Quel(s) service(s) de la MEL représentez-vous ?

Quel poste occupez-vous ?

Depuis combien de temps êtes- vous à ce poste ?

Connaissez-vous la loi du numérique de 2016 sur l'ouverture des données par les institutions ?  
\*OUI\*

Quels sont les principaux objectifs de cette loi selon votre compréhension ?

Avez-vous déjà réutilisé des données ouvertes dans le cadre de votre travail ou de vos projets ?

Pouvez-vous identifier des domaines où la réutilisation des données ouvertes pourrait avoir un impact significatif ?

Quelle est votre expérience ou votre connaissance de l'open data et de son utilisation dans un contexte économique ?

Impact économique : Quels sont, selon vous, les principaux avantages économiques de l'open data ?

Avez-vous des exemples concrets d'entreprises ou de secteurs qui ont bénéficié de l'utilisation de données ouvertes ?

Comment l'open data peut-il stimuler l'innovation et la création d'entreprises ?

Quels sont les effets potentiels de l'open data sur la productivité et l'efficacité des entreprises ?

Dans quelle mesure l'open data peut-il favoriser la transparence et la responsabilité des entreprises et des institutions publiques ?

Y a-t-il des défis ou des obstacles spécifiques à l'utilisation de l'open data dans un contexte économique ? Comment peuvent-ils être surmontés ?

Quels sont les éléments essentiels à prendre en compte lors de l'évaluation de l'impact économique de l'open data ?

Exemples et cas concrets :

Connaissez-vous des études de cas ou des exemples où l'utilisation de l'open data a eu un impact économique significatif ?

Pouvez-vous fournir des exemples de modèles économiques basés sur l'open data ?

Quels sont les défis auxquels les entreprises sont confrontées lorsqu'elles cherchent à monétiser les données ouvertes ?

Comment les organisations peuvent-elles évaluer la valeur économique des données ouvertes qu'elles possèdent ?

Avez-vous connaissances d'entreprises ou d'organisations qui ont pu tirer parti des données ouvertes pour créer de la valeur économique ?

Pensez-vous que la donnée ouverte puisse être génératrice d'emplois ? Oui ? pouvez-vous aller au bout de vos idées svp ?

Quels sont les facteurs clés qui peuvent encourager les entreprises à embaucher des talents spécialisés dans l'exploitation des données ouvertes ?

Conclusion

Merci pour votre précieuse contribution à notre étude. Nous nous engageons à utiliser les informations recueillies de manière confidentielle et à les utiliser uniquement à des fins de recherche.

## **Corpus**

### **Entretiens qualitatifs**

#### **Articles :**

- Avis d'expert : L'open data privé, le nouvel eldorado des entreprises ?
- Open Data : comment créer de la valeur à partir d'une donnée « gratuite »
- Open Data : un deal parfait ?
  
- Open Data définition : qu'est-ce que c'est ? À quoi ça sert ?
  
- Open data : de nouveaux jeux de données adaptés aux attentes des utilisateurs
- L'impact économique des données ouvertes : que savons-nous déjà ?
- Open Data définition : qu'est-ce que c'est ? À quoi ça sert ?
- La data au cœur d'un nouvel équilibre dans les échanges public-privé
- Coopération public – privé et données : comment et pourquoi les collectivités doivent reprendre la main
- Dix ans d'ouverture des données publiques : un bilan critique

#### **Dossiers des étudiants du Master 2 information-documentation de la promotion 2022-2023**

## Résumé

Les données ouvertes, sont considérées comme un mouvement visant à rendre accessibles au public des ensembles de données de manière libre et gratuite. En comprenant les principes et les valeurs fondamentales du mouvement des données ouvertes, il devient aisé de saisir pleinement son potentiel et les implications qu'il peut avoir dans divers domaines. Cette pratique revêt une grande importance pour les organisations, qu'elles soient publiques ou privées, car la gestion efficace des données est devenue une préoccupation majeure à mesure que leur valeur économique est reconnue. A travers une approche d'analyse de corpus de la question de l'impact économique des données ouvertes, trois hypothèses sont examinées : l'ouverture des données favorise l'innovation et la création d'entreprises, l'accès aux données ouvertes améliore l'évaluation des risques et la prise de décision, et la disponibilité des données encourage la collaboration entre les acteurs économiques. Les résultats de cette recherche permettent de mieux comprendre comment l'open data peut stimuler l'économie et comment elle se manifeste.

**Mots-clés :** données ouvertes, économie, réutilisation, innovation