

# **Exploitation de l'Open Data avec Python par l'architecte**

Démonstration par la pratique des apports potentiels au sein de la conception architecturale

## **ABSTRACT**

Ce mémoire a pour but d'aborder les enjeux de l'appropriation du langage de programmation Python par les architectes afin d'exploiter les données issues de plateformes en Open Data. Plus précisément, le potentiel de ce langage sera démontré à travers une approche pratique de plusieurs scripts répondant aux principaux enjeux autour de l'appropriation des données ouvertes par les architectes. Comment manipuler de manière universelle les formats de données ouverts et leur structure ? De quelle manière peut-on les intégrer dans l'environnement de travail de la conception architecturale ? Peut-on finalement se servir de Python pour mener ses propres analyses ?

# SOMMAIRE

<b>ABSTRACT</b>	<b>2</b>
<b>SOMMAIRE</b>	<b>3</b>
<b>INTRODUCTION</b>	<b>5</b>
<b>1 Etude de cas des données des « volumes bâtis » de l'Open Data Paris : le script Python comme outil unique pour gérer la complexité.</b>	<b>9</b>
1.1 L'Open Data : entre nomenclature et variables . . . . .	10
1.1.1 Variables et typologies des valeurs . . . . .	10
1.1.2 Les métadonnées : clé de compréhension des données . . . . .	10
1.2 La donnée ouverte : hiérarchisée et hétérogène . . . . .	10
1.2.1 L'hétérogénéité : caractéristique intrinsèque aux données ouvertes . . . . .	10
1.2.2 Le format JSON : une structure plus souple que le tableur . . . . .	10
1.3 Un script Python pour extraire, comprendre et épurer . . . . .	10
1.3.1 Un outil adapté à la manipulation de formats ouverts . . . . .	10
1.3.2 Transformer des notations complexes en valeurs compréhensibles . . . . .	10
<b>2 Synthétiser et intégrer les données ouvertes au sein du « workflow » de l'architecte avec Python : du graphique au modèle 3D</b>	<b>11</b>
2.1 Le graphique statistique : visualisation basique des données . . . . .	13
2.2 Cartographie personnalisée . . . . .	13
2.3 Production de dessins vectorisés organisé . . . . .	13
2.4 Génération d'un modèle 3D . . . . .	13

<b>3</b>	<b>Analyse approfondie d'une masse de données en Open Data : révéler et prédire des liens pour aiguïser sa conception</b>	<b>14</b>
	<b>CONCLUSION</b>	<b>15</b>
	<b>BIBLIOGRAPHIE</b>	<b>18</b>

## INTRODUCTION

Au cours des dernières années, un déploiement prolifique de jeux de données est en train d'avoir lieu sous l'égide de «l'Open Data». Sur le territoire Français, des dispositifs mis en place par le gouvernement tel qu'«Etalab», chargé de la coordination et la mise en place de l'ouverture de jeux de données (par décret du 30 Octobre 2019) incarnent cette volonté de faciliter la diffusion de données ouvertes, tout en promouvant leur réutilisation.<sup>1</sup>

Le gouvernement définit l'Open Data comme «l'effort que font les institutions, notamment gouvernementales, qui partagent les données dont elles disposent».<sup>2</sup> En effet, c'est avant tout une stratégie prônant l'ouverture du plus grand nombre de bases de données au public, les rendant ainsi totalement accessibles. A l'instar des autres mouvements du même type, tel que «l'Open Source», le traitement et la rediffusion des données sont autorisées, voir même encouragées comme c'est le cas par le gouvernement français : «les données partagées trouvent des réutilisateurs qui les intègrent dans de nouveaux services à forte valeur ajoutée économique ou sociale.». Les règles relatives à leur réutilisation font l'objet d'une licence publique et universelle, ne réclament pas ou peu de démarches pour se l'approprier.

Ainsi, de nombreuses plateformes mises en place par diverses instances opérant dans des domaines très variés ont vu le jour au cours des dernières années, allant d'organismes spécialisés dans les données géographiques comme l'Institut national de l'information géographique et forestière (IGN)<sup>3</sup> jusque dans le domaine des transports comme Ile de France Mobilités,<sup>4</sup> en passant par l'environnement et l'écologie tel que l'ADEME.<sup>5</sup>

---

<sup>1</sup> *Etalab - Qui sommes-nous. Le blog d'Etalab* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.etalab.gouv.fr/qui-sommes-nous>.

<sup>2</sup> *L'ouverture des données publiques. Gouvernement.fr* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>.

<sup>3</sup> *Géoservices | Accéder au téléchargement des données libres IGN* [en ligne]. [s. d.]. [Consulté le 15 septembre 2020]. Disponible à l'adresse : <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html>.

<sup>4</sup> *Portail Open data Île-de-France Mobilités* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://data.iledefrance-mobilites.fr/pages/home/>.

<sup>5</sup> *Portail open data de l'ADEME* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://data.ademe.fr/>.

Bien que cette nécessité étatique de partager l'information publique ne date pas de l'apparition du Web (comme l'explique la loi Cada de 1978), Ce dernier a permis, au-delà de la dispense de tout intermédiaire (notamment humain) entre le fournisseur et l'utilisateur, d'exploiter de nouvelles formes d'accès et surtout de consommation, en particulier via des scripts ou des algorithmes écrits dans un langage de programmation afin d'automatiser la récupération de données depuis les formats de fichiers ouverts.

Ainsi, quiconque cherche à mettre en place un travail d'analyse le plus exhaustif possible d'un contexte donné peut, grâce aux plateformes et moyens cités ci-dessus, disposer très rapidement de données riches et abondantes.

De ce point de vue-là, il paraît extrêmement pertinent pour les métiers issus de l'architecture et de l'urbanisme, et en particulier le métier d'architecte, de se saisir des données issues de l'Open Data afin de renforcer leur compréhension du territoire sur lequel ils construisent, que cela soit par la simple analyse statistique ou bien la récupération d'informations géométriques d'un site.

Or, les architectes ont tendance à préférer, de par leur expertise orientée sur la conception, réclamant un esprit de synthèse affûté, les résultats explicites d'analyse de données plutôt que les données en elles-mêmes. De plus, les outils numériques sur lesquels les architectes se forment relèvent très majoritairement des domaines du dessin, de la modélisation ou de la communication plutôt que de l'analyse en elle-même, qui accentue leur besoin de résultats synthétiques «préfabriqués».

Cependant, il existe depuis les années 2010 un certain essor des travaux de recherche basés sur des données issues en partie ou totalement de l'Open Data, et ce, grâce à un langage de programmation en particulier, dont la simplicité de la syntaxe couplée à une profusion de bibliothèques (comparables à des «plug-in») spécialisées dans le traitement de données informatiques en ont fait un outil populaire pour la recherche d'aujourd'hui,

le Python. En effet, comme l'illustre l'exemple du travail de recherche «CityEngine - Twitter» mené au «Centre for Advanced Spatial Analysis» de Londres<sup>6</sup>, proposant une cartographie urbaine de densité basée sur des Tweets géolocalisés dans cette même ville, un seul et unique script en Python permet à la fois de récupérer les messages sur une plage de 24 heures (via une bibliothèque, nommée «Tweepy», permettant au code d'interagir avec l'API de Twitter), de les trier et d'en extraire leurs coordonnées et leur horaire de publication, et enfin de fournir ces données directement à l'outil de génération de modèles 3D urbains «CityEngine» (publié par l'ESRI) afin que ce dernier puisse constituer une carte procédurale (animée selon le nombre de tweets sur une plage de 24 heures).

Une telle étude étant désormais possible sur des données massives privées, ce type d'exploitation peut encore plus aisément être mis en place lorsque les données utilisées sont totalement ouvertes et avec accès illimité.

Ainsi, grâce à des données massives accessibles (tant en termes de tarifs qu'en terme de facilité d'extraction) couplées à un langage de programmation comme Python, développer ses propres analyses par exploitation de données brutes est désormais à la portée des chercheurs, sans avoir besoin d'un bagage informatique conséquent.

Dès lors, face à la complexité des enjeux auxquels la conception architecturale fait appel (climatique, socio-économique, écologique ou structurel par exemple), il semble pertinent d'envisager que des architectes se saisissent de ce type d'outil, dans le but de construire, au prisme de leurs propres volontés d'intervention (même complexes), leurs propres modèles de compréhension du territoire. Ce nouveau regard, personnalisé par l'architecte, pourrait alors apporter à ce dernier des éléments susceptibles de le guider de manière bien plus significative, en particulier dans les premières phases d'esquisse, afin d'améliorer la qualité de sa production.

---

<sup>6</sup>HÜGEL, Stephan et ROUMPANI, Flora. *CityEngine-Twitter* [logiciel]. [S. l.] : Zenodo, 14 mai 2014. [Consulté le 28 décembre 2020]. DOI 10.5281/ZENODO.9795.

**Ainsi, dans quelle mesure l'exploitation de données issues de l'Open Data grâce au langage Python représente-t-elle un avantage certain pour l'architecte ?**

Après avoir initialement démontré l'intérêt du langage Python dans l'extraction et la manipulation des données issues des plateformes accessibles en Open Data à travers l'élaboration complète d'un script de récolte de données, ce dernier sera complété à travers un aperçu constitué d'exemples clés de la capacité de Python à produire des documents de travail utiles à l'architecte (cartographie, dessin et modélisation). Enfin, ce travail d'exploitation sera abouti en montrant la prodigieuse capacité du langage Python à permettre de manière accessible l'analyse complexe de ces données ainsi que la mise en place d'algorithmes de prédiction.



## 1 Etude de cas des données des « volumes bâtis » de l'Open Data Paris : le script Python comme outil unique pour gérer la complexité.

Tel que le stipule le portail européen de données, au-delà de l'accessibilité en elle-même des données, la question de la lisibilité des structures de données et des formats de fichiers disponibles sur les plateformes relevant de l'Open Data est d'importance cruciale : « On peut utiliser les données car elles sont disponibles sous une forme commune et lisibles par des machines. ».<sup>7</sup> Cet organisme relève également un autre aspect primordial, celui de la facilité du traitement des données par les outils informatiques. En effet, elles ont davantage vocation à faire l'objet de manipulations automatiques (synthèse, tri, etc...) plutôt que d'être simplement lues par un utilisateur humain.

Pour partager des données tabulaires (sous forme de tableur) par exemple, là où un utilisateur humain préférera un format Excel (.XLSX) (en y incluant notamment couleurs et styles de polices pour améliorer sa lisibilité), le portail européen des données recommande plutôt d'autres formats comme le .CSV (Comma Separated Values), format ouvert constitué de texte brut séparé par des caractères spéciaux, compatible avec un large panel d'outils logiciels capable d'opérations de traitement.

Face à ce besoin de compréhension et de manipulation de données brutes, les langages de programmation de haut niveau d'abstraction (possédant une syntaxe plus lisible et concise pour l'humain, rendant leur utilisation accessible) et en particulier le Python apparaissent alors comme des outils offrant la souplesse et la puissance nécessaire pour répondre à cette problématique.

**Au sein de cette section, le jeu de données « Volumes bâtis » de la plateforme Open Data Paris sera étudié de près en tant qu'exemple type, à travers une approche concrète de sa complexité. Afin de permettre son exploitation, un script Python sera élaboré en fin de section. Ce travail servira également de base pour**

---

<sup>7</sup>What is open data? [en ligne]. [s. d.]. [Consulté le 28 décembre 2020]. Disponible à l'adresse : <https://www.europeandataportal.eu/earning/en/module1/#/id/co-01>.

**aborder les concepts plus approfondis des chapitres suivants.**

## 1.1 L'Open Data : entre nomenclature et variables

### 1.1.1 Variables et typologies des valeurs

### 1.1.2 Les métadonnées : clé de compréhension des données

## 1.2 La donnée ouverte : hiérarchisée et hétérogène

### 1.2.1 L'hétérogénéité : caractéristique intrinsèque aux données ouvertes

### 1.2.2 Le format JSON : une structure plus souple que le tableur

## 1.3 Un script Python pour extraire, comprendre et épurer

### 1.3.1 Un outil adapté à la manipulation de formats ouverts

### 1.3.2 Transformer des notations complexes en valeurs compréhensibles

## 2 Synthétiser et intégrer les données ouvertes au sein du «workflow» de l'architecte avec Python : du graphique au modèle 3D

La production de documents graphiques faisant partie intégrante du «workflow» de l'architecte, il est primordial de s'y intéresser au sein de ce mémoire.

En effet, c'est via ce type de document que l'architecte est capable de non seulement communiquer sa production ou encore sa démarche de conception, mais également de se documenter en vue de ses travaux de conception. Cependant, ce sont aussi des moyens de communication plutôt chronophages à produire, en particulier lorsque l'architecte est amené à dessiner des éléments existants.

Or, comme démontré dans le chapitre précédent, les jeux de données massifs disponibles en Open Data représentent une ressource parfaite afin de réduire cette charge de travail, pourvu que l'on dispose d'outils capables de dessiner ou de modéliser grâce à des données brutes.

Encore une fois, le langage Python est tout autant capable de produire des documents que d'en lire, et les domaines de la modélisation et du graphisme ne font pas exception.

**Ce chapitre présentera plusieurs variantes d'exploitation du script obtenu à la fin du chapitre précédent dans le but d'obtenir des documents synthétiques à partir des données des bâtiments obtenues. Elles seront respectivement consacrées à l'élaboration de graphiques statistiques, puis d'une carte interactive, d'un dessin vectorisé avec gestion des calques, et enfin d'un modèle 3D du contexte bâti.**

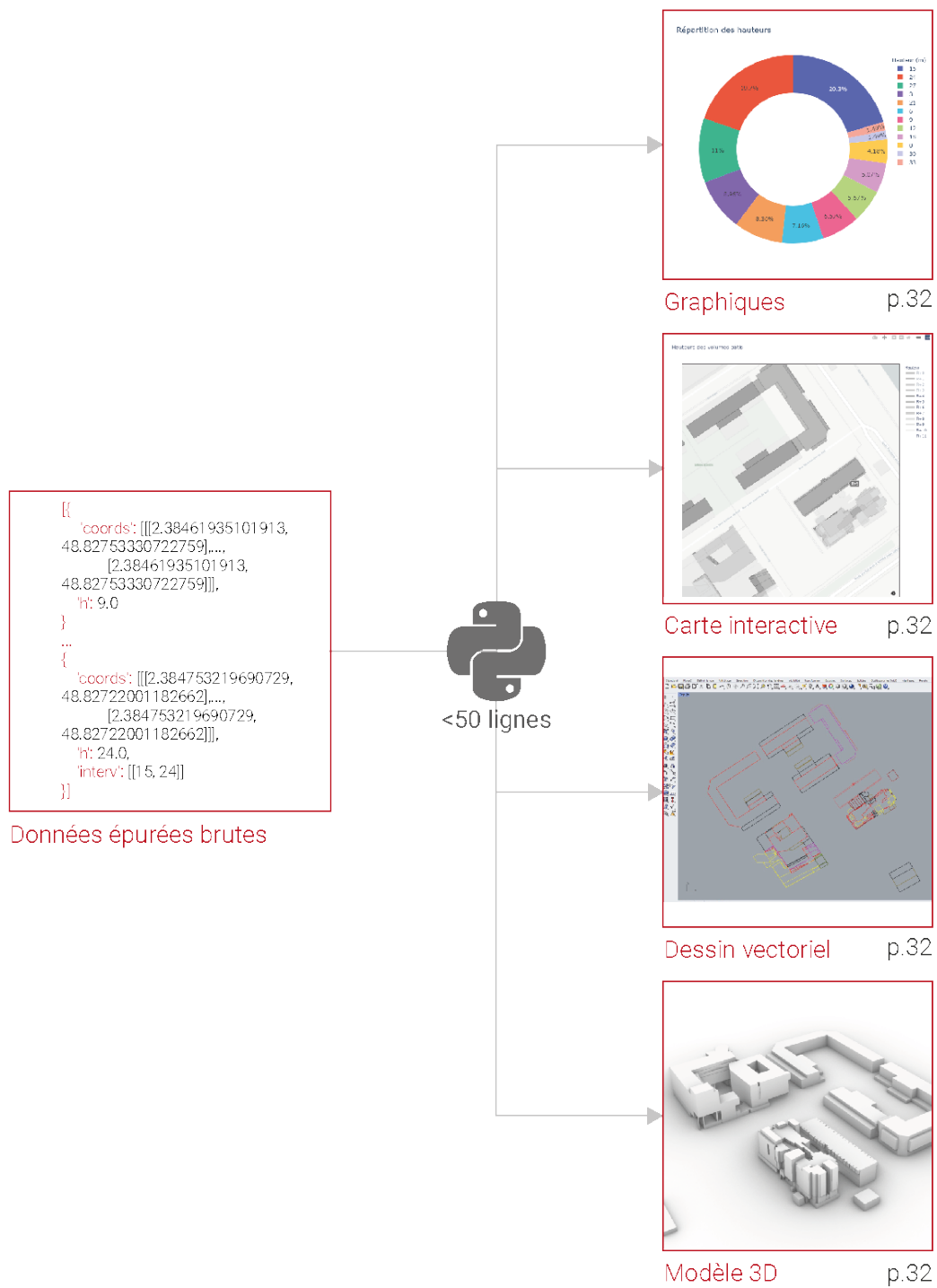


FIG. 1 :

2.1 Le graphique statistique : visualisation basique des données

2.2 Cartographie personnalisée

2.3 Production de dessins vectorisés organisé

2.4 Génération d'un modèle 3D

### **3 Analyse approfondie d'une masse de données en Open Data : révéler et prédire des liens pour aiguïser sa conception**

Montrer l'exemple de l'outil de prédiction de gisement de matériaux développé dans le cadre du PFE, en tant qu'« aboutissement » de ce que l'on est actuellement capable de tirer des données en Open Data.

## CONCLUSION

Grâce à l'approche pratique proposée au sein de ce mémoire, nous avons pu démontrer que le langage de programmation Python constitue un outil souple et performant permettant de d'accompagner l'architecte tout au long de son exploitation des données issues de l'Open Data.

En effet, permettant en premier lieu de rendre compréhensible et manipulable aux plus novices les structures de données couramment rencontrées, en particulier les données hiérarchisées hétérogènes, le langage est doté d'outils de production de documents synthétiques tout à fait capables de convertir les données extraites en formats de fichiers courants pour l'architecte, élément primordial au sein du « workflow » de l'architecte. Ainsi, au-delà des graphiques statistiques, Python permet la génération de cartographies interactives, mais surtout des dessins techniques hiérarchisés et même des modèles 3D de manière automatique et personnalisée, représentant un gain de temps considérable.

Enfin, l'architecte devient capable grâce à Python de mener ses propres analyses de données afin de les mettre à profit dans son activité de conception, grâce à une capacité d'identification de corrélations entre différentes données agrégées, ou encore celle de pouvoir prédire l'impact de son intervention grâce au Machine Learning.

De plus, épaulé par divers modules permettant tous ces usages de manière simplifiée tout au long du processus d'exploitation des données ouvertes, le Python acquiert un véritable caractère universel, autant dans le sens où il est capable de se suffire à lui-même que pour qualifier sa compatibilité extraordinaire avec d'autres outils et services (comme Rhinoceros).

Au-delà du cadre de l'exploitation des données issues de l'Open Data, ce langage représente un véritable pivot afin d'accompagner les agences vers les nouveaux outils plus intelligents.

Premièrement, l'expérience acquise durant la manipulation des jeux de données ouverts (en particulier sur les notions des structures de données) sera extrêmement précieuse lorsqu'il s'agira d'exploiter des jeux de données plus directement liés à la pratique architecturale en elle-même, comme depuis un parc de « Smart Buildings » qu'il faudra surveiller et analyser par exemple.

Au-delà de cet aspect, des solutions logicielles émergentes tel que « Générative Design in Revit » publiée par Autodesk,<sup>8</sup> désormais intégrée à Revit 2021 proposent déjà à l'architecte de travailler aux côtés d'algorithmes évolutifs et prédictifs, que ce soit de manière simplifiée par interface graphique ou bien personnalisable grâce au langage de programmation visuelle Dynamo. La collaboration entre des agences d'architecture et des solutions intégrant de l'Intelligence Artificielle commence également à se développer, tel que l'agence Viguiet et son partenariat avec « SpaceMaker AI »,<sup>9</sup> société norvégienne spécialisée dans la conception urbaine générative.

Toutes ces opportunités nouvelles requièrent une certaine familiarité avec le fonctionnement algorithmique, et surtout une assimilation de sa manière de fonctionner afin de l'intégrer efficacement à son « workflow »

Dès lors, de par son statut de langage de programmation lui conférant un caractère universel quant aux notions de bases informatiques et surtout algorithmiques (même avec une syntaxe simplifiée), le Python se présente comme un atout « futur-proof » face à ces nouvelles interactions architecte-machine. De plus, l'adoption massive du Python par les architectes permettrait d'amorcer une solution efficace afin de lutter contre la crainte de la montée en puissance de l'Intelligence Artificielle au sein des métiers du bâtiment, par une atténuation considérable de l'effet « Boite noire » qu'elle suscite. Ceci faciliterait donc grandement la transition du corps architectural dans la numérisation du domaine du bâtiment.

---

<sup>8</sup>*Generative Design in Revit now available. Revit* [en ligne]. 8 avril 2020. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://blogs.autodesk.com/revit/2020/04/08/generative-design-in-revit/>. Section : What's New.

<sup>9</sup>*VIGUIER noue un partenariat avec SPACEMAKER AI* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.viguiet.com/fr/actualites/journal/viguiet-noue-un-partenariat-avec-spacemaker-ai>.



Au-delà de cette appropriation, une certaine motivation de créer des outils logiciels par les architectes pour les architectes intégrant ces nouvelles compétences pourrait émerger, permettant alors de réaffirmer la place du métier d'architecte au sein d'un écosystème de plus en plus techno-centré.

## BIBLIOGRAPHIE

*Etalab - Qui sommes-nous. Le blog d'Etalab* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.etalab.gouv.fr/qui-sommes-nous>

*Generative Design in Revit now available. Revit* [en ligne]. 8 avril 2020. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://blogs.autodesk.com/revit/2020/04/08/generative-design-in-revit/>. Section : What's New

*Géoservices | Accéder au téléchargement des données libres IGN* [en ligne]. [s. d.]. [Consulté le 15 septembre 2020]. Disponible à l'adresse : <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html>

*L'ouverture des données publiques. Gouvernement.fr* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>

*Portail open data de l'ADEME* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://data.ademe.fr/>

*Portail Open data Île-de-France Mobilités* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://data.iledefrance-mobilites.fr/pages/home/>

*VIGUIER noue un partenariat avec SPACEMAKER AI* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.viguier.com/fr/actualites/journal/viguier-noue-un-partenariat-avec-spacemaker-ai>

*What is open data?* [en ligne]. [s. d.]. [Consulté le 28 décembre 2020]. Disponible à l'adresse : <https://www.europeandataportal.eu/elearning/en/module1/#/id/co-01>

HÜGEL, Stephan et ROUMPANI, Flora. *CityEngine-Twitter* [logiciel]. [S. l.] : Zenodo, 14 mai 2014. [Consulté le 28 décembre 2020]. DOI 10.5281/ZENODO.9795

VANNIEUWENHUYZE, Aurélien. *Intelligence artificielle vulgarisée : le Machine Learning et le Deep Learning par la pratique*. [S. l.] : [s. n.], 2019. ISBN 978-2-409-02073-5. OCLC : 1127535504