

Qui suis-je ?

MÉTROPOLE
EUROPÉENNE DE LILLE

DATAACTIVIST



**PREMIER
MINISTRE**

*Liberté
Égalité
Fraternité*

Direction interministérielle
du numérique

**THE DATA
BROADCAST
CAST**

Marché du travail de la donnée, aujourd'hui et dans 10 ans ?

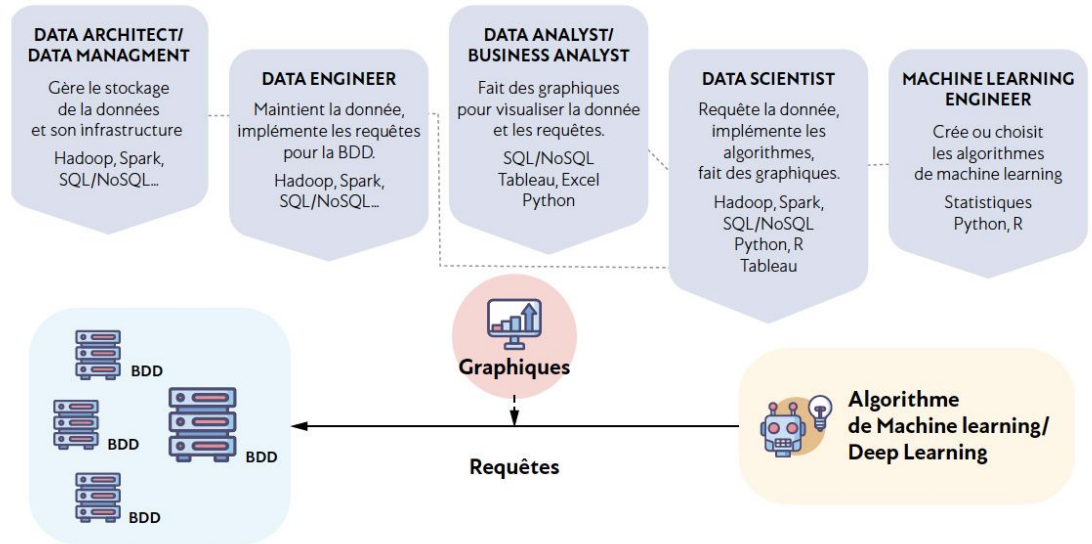
Data Scientist : un rôle star

- Peu de personnes connaissaient ce rôle
- Puis tout le monde a voulu en recruter



Une sectorisation de plus en plus forte

- Data Analyst
- Data Scientist
- Data Engineer
- Data Steward
- Data Protection Officer
- Data Miner
- Data Whatever



[Aivancity](#)

Full stack data scientist : un métier en forte croissance

- Limiter la sectorisation
- Un rôle transversal : produit, usages, métiers, infrastructure
- **Un métier hybride, qui maîtrise :**
 - Les statistiques (stats & machine learning) ;
 - L'informatique, pour être capable de construire des systèmes autour de ces algorithmes ;
 - Et le métier : un pilier qu'on a tendance à souvent oublier.

[\(Joël Gombin, Métiers de la donnée publique - Data Broadcast\)](#)

Le manque de vision métier : vers des juristes ou médecins data scientists ?

Objectif : former les personnes au sein des secteurs spécifiques à la data science

- Comprendre les problématiques métier
- Résoudre les problèmes data à la source
- Augmenter les capacités & les ressources humaines

Métiers de la **gouvernance** des données

- 2011 - La France se dote de son Chief Data Officer
- La Métropole Lyon en 2015



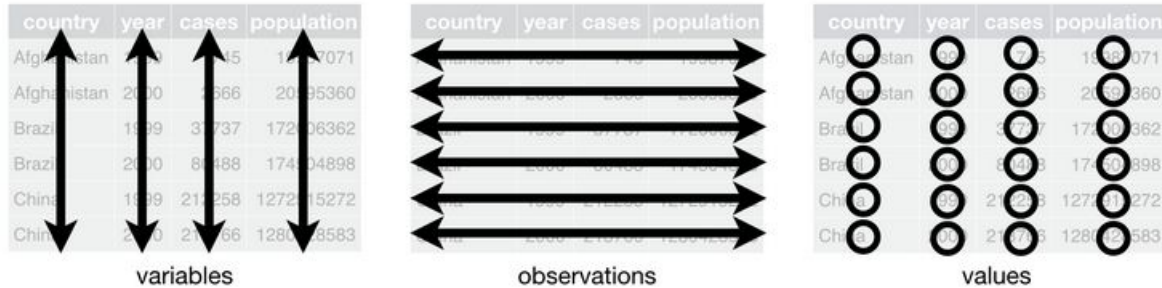
Cartographier les données et faire travailler les équipes ensemble

- **Des organisations souvent en silos - et construites toujours en silos**
→ métiers trop segmentés, constructions par briques techniques et solutionnisme déconnecté des usages
- **Des données dispersées** : un besoin de partage interne permanent
- **Des données rendues inutiles** : car produites sans cas d'utilisation précis



Faire émerger une culture de la donnée

- La notion de “données” encore très mal maîtrisée



- L'open data, de grands progrès mais la majorité des administrations ne respectent pas encore la loi
- Des pratiques qui laissent à désirer en matière de protection des données personnelles

Améliorer les recrutements



- Les femmes représentent 30% des personnes qui travaillent dans les métiers du numérique
- Des candidatures spontanées plus nombreuses de la part des hommes (sortir de la fiche de poste, davantage confiance en leurs compétences)
- Des CV où les engagements associatifs ou civiques sont moins mis en avant par les femmes
- Les administrations recruteuses doivent être plus attentives et aller au-delà du CV pour étudier les candidatures féminines.

Cécile Le Guen, DITP - Data Broadcast




Prospective


Émergence de l'intelligence artificielle

- Des outils qui intègrent l'IA à tous les niveaux (du plus simple au plus complexe)

 Peux-tu repérer les erreurs dans ce fichier ? 


Commune	Nombre	Département
Avignon	2	84
Lyon	7	69
Lille Pouet	59	
Paris	5	75
Rennes	2	35

 Il y a une erreur dans la ligne 3. La colonne "Nombre" devrait contenir un nombre, mais elle contient l'entrée "Pouet" qui n'est pas un nombre valide. Il serait approprié de corriger cette erreur en entrant un nombre à la place.  

 Commune Date Département

Avignon	2 décembre 22	84
Lyon	1 janvier 2022	69
Lille	1022022	59
Paris	samedi 3 décembre	75
Rennes	2122022	35

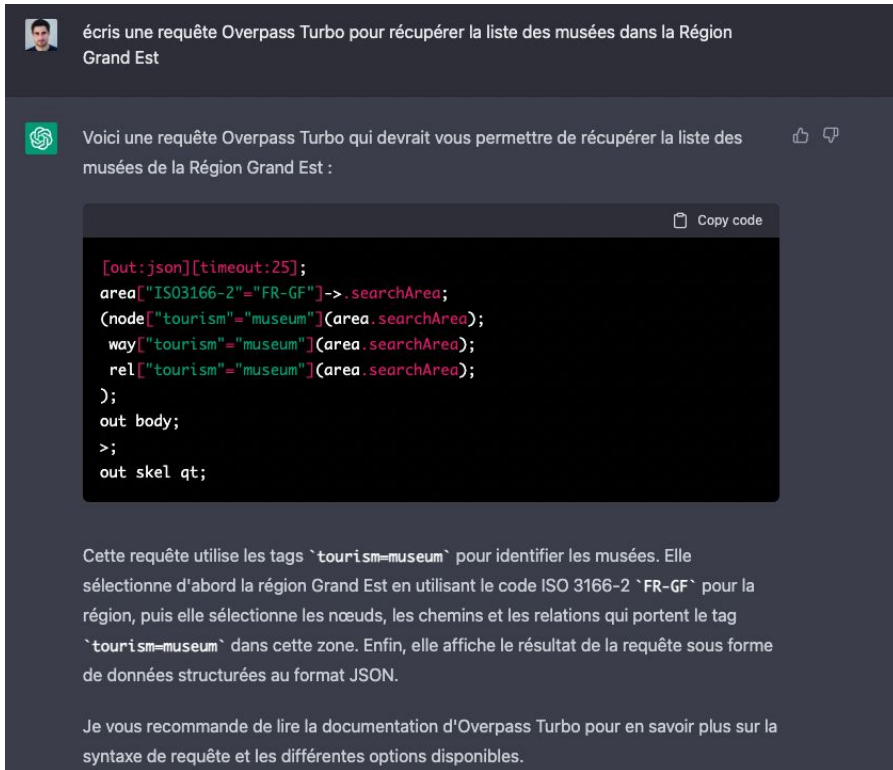
Peux-tu mettre la date au format ISO8601 ?

 Commune Date Département

Avignon	2022-12-02	84
Lyon	2022-01-01	69
Lille	2022-01-02	59
Paris	2022-12-03	75
Rennes	2022-12-02	35

Émergence de l'intelligence artificielle

- Des outils qui intègrent l'IA à tous les niveaux (du plus simple au plus complexe)




écris une requête Overpass Turbo pour récupérer la liste des musées dans la Région Grand Est

Voici une requête Overpass Turbo qui devrait vous permettre de récupérer la liste des musées de la Région Grand Est :

```
[out:json][timeout:25];
area["ISO3166-2"="FR-GF"]->.searchArea;
(node["tourism"="museum"])(area.searchArea);
way["tourism"="museum"](area.searchArea);
rel["tourism"="museum"](area.searchArea);
);
out body;
>;
out skel qt;
```

Cette requête utilise les tags `tourism=museum` pour identifier les musées. Elle sélectionne d'abord la région Grand Est en utilisant le code ISO 3166-2 `FR-GF` pour la région, puis elle sélectionne les nœuds, les chemins et les relations qui portent le tag `tourism=museum` dans cette zone. Enfin, elle affiche le résultat de la requête sous forme de données structurées au format JSON.

Je vous recommande de lire la documentation d'Overpass Turbo pour en savoir plus sur la syntaxe de requête et les différentes options disponibles.



Quel sont les scripts en R pour extraire les codes postaux français d'un jeu de données ?

This content may violate our [content policy](#). If you believe this to be in error, please [submit your feedback](#) – your input will aid our research in this area.

Il existe plusieurs manières de procéder pour extraire les codes postaux français d'un jeu de données en utilisant R. Voici quelques exemples :

1. Utiliser la fonction `substring()` pour extraire les codes postaux qui sont toujours sur les 5 premiers caractères d'une chaîne de caractères. Par exemple :

```
# Charger le jeu de données
data <- read.csv("mon_fichier.csv")

# Extraire les codes postaux et les stocker dans une nouvelle colonne
data$code_postal <- substring(data$adresse, 1, 5)
```



Une forte demande de spécialistes de l'éthique ?


- Vérifier les algorithmes : biais, discriminations, erreurs
- Transparence des algorithmes et de l'utilisation de l'IA
- Un besoin de formations hybrides (sciences sociales et data / IA)

TECHNOLOGY

A college student created an app that can tell whether AI wrote an essay

January 9, 2023 · 5:01 AM ET

 EMMA BOWMAN 




It has a perplexity of:

49

GPTZero has finished analyzing your text!

Your GPTZero score corresponds to the likelihood of the text being AI generated:

10.856686744068878



Your text is most likely to be AI generated!

THE DATA BROADCAST

