

**Introduction :** Depuis plusieurs années maintenant, il n'est pas un journal, pas une émission de radio, pas un podcast qui ne parlent pas de données. Les fameuses data. Osint le parallèle avec ce que fut le pétrole au siècle dernier et assimilant ces dernières à un moteur de changement et de croissance.

De nouvelles infrastructures, de nouvelles entreprises, de nouveaux monopoles, de nouvelles économies ont été créées par ces flux de données. Et pour aller au bout de l'analogie, on emprunte au vocabulaire minier les termes d'extraction, de raffinage, de valorisation, d'achat et de vente comme pour mieux signifier l'importance de cette ressource. Ce qui n'est pas sans incidence sur l'appréhension de la problématique.



La question de l'accès à ces données se pose tant sous le prisme de la **qualité** que de la **quantité**. Et lorsqu'il s'agit d'appréhender ces données, la question de leur **accessibilité** se pose et là la difficulté commence.

L'Osint (open source intelligence), est une parmi d'autres techniques, elle facilite cette quête et s'appréhende d'abord comme une démarche. Bloc de compétences pour manier et valoriser l'information au quotidien, ses principales sources sont l'open data, les sites officiels, la presse, les blogs, les médias sociaux...etc .



### **L'accès aux données ouvertes.**

Utilisée entre autres par les journalistes d'investigation mais aussi dans le cadre d'enquêtes en criminalité informatique, cette technique s'assimile à un nouveau régime de vérité qui repose sur l'étude des traces pour établir des preuves qui viennent étayer une logique démonstrative et explicative, notamment pour répondre à de la désinformation. Utilisée également en IE, l'Osint peut être présente aux différentes étapes du cycle de l'information : dans la Collecte, l'analyse, la visualisation, ou la diffusion.

### **Vouloir savoir, vouloir comprendre, vouloir vérifier.**

Donc avoir recours à l'**Osint**, c'est avoir recours à des techniques d'investigation, c'est maîtriser certains outils et posséder des compétences techniques. D'un autre côté, pour capter ces dites data, qu'elles soient structurées ou non, l'on peut avoir recours au langage

## Fiche 05 :Osint et Python par Philippe Dion, Consultant RH

informatique comme Python. Capturer les commentaires d'un produit par exemple via le **webscraping** ou les tweets autour d'un **hashtag** via une api sont une autre façon de capturer l'information. Et ainsi, par le biais d'un script récupérer davantage de tweets que sur une application qui peut devenir payante assez rapidement après sa création. Il ne s'agit bien évidemment de ne pas opposer ces deux démarches. **Elles sont complémentaires.**

Si la manière d'appréhender la data est différente, travailler en Osint comme en python est une démarche qui demande de la créativité. Beaucoup moins si on imbrique des outils. ou avant l'élaboration d'un script.

### Mais alors quel avantage a coder ?

A mon sens, l'avantage est de faire du sur mesure. Si l'on prend l'exemple de **Twitter**, les requêtes à l'Api ne permettent pas plus d'un certain nombre de données. L'avantage de python, par les boucles, est de reproduire l'exécution du script et donc la récupération de données.

Certes, tout ceci nécessite des compétences techniques mais avec l'avantage d'un traitement qui ira plus loin que la simple collecte. La data science recèle d'outils statistiques, mathématiques qui permettent d'aller plus loin dans l'aide à la décision. Et faire appel à des algorithmes de **machine learning** permet de franchir un autre cap. Dans une même base de données, on peut stocker des données provenant de sources différentes et lorsque le jeu de données comporte beaucoup de variables, on peut résumer l'information en composantes principales et en avoir une représentation plus simple (Analyse en Composantes Principales).

Pouvoir la stocker dans une base de données, dans un dataframe pandas et utiliser un **algorithme** de NLP (Natural language processing) afin de mesurer la tonalité positive ou négative d'avis client est une autre possibilité. **Connecter** ses propres outils pour diffuser ou collecter facilement des informations, de manière manuelle ou automatisée.

Et lorsque la réponse à sa problématique devient difficile, faire appel à une communauté de développeurs, de data analysts présents sur des forums ou via des plateformes comme [github](https://github.com).

### Quelques conseils pour débiter ?

L'exercice de la manipulation des données commence déjà par une manière de percevoir python. Certes l'apprentissage de la syntaxe, des librairies peut être un passage obligé. Mais ce qui me semble plus intéressant comme approche c'est de commencer par vouloir résoudre des petits problèmes, des questions que l'on rencontre dans sa collecte d'informations. Se confronter à des cas réels.

Si l'on souhaite se diriger vers la data-analyse, il me semble impératif d'avoir (si ce n'est pas déjà le cas) de bonnes bases en statistiques. Pouvoir ainsi réaliser et expliquer une étude exploratoire de son jeu de données.

Dans le domaine des ressources humaines qui est le mien, l'on dispose de données très diverses (du temps, des coûts, des effectifs, des sessions de formation...), ayant donc des

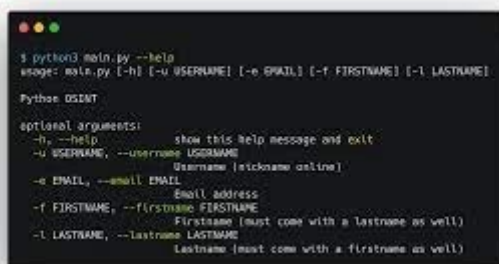
## Fiche 05 :Osint et Python par Philippe Dion, Consultant RH

caractéristiques très différentes. Les axes d'analyse sont divers comme l'axe temporel, l'axe des salariés, l'axe financier. Les chiffres sont très présents au sein de la fonction RH et en alimentent les constats et actions en complément des autres compétences traditionnelles des RH (dialogue social, droit du travail, développement des compétences...).

L'analytique RH permettent de mieux connaître le capital humain, un meilleur pilotage des politiques RH.

La data visualisation permettra de rendre ses reportings plus efficaces en facilitant la restitution de l'information et donc la communication.

**Les contraintes** : le respect des règles des plateformes est essentiel, dans l'usage de la collecte des données, des quantités de flux collectées. En cas de non-respect, le bannissement est le moindre des maux. La sévérité des sanctions allant croissant, un mésusage des données pourrait être l'occasion d'arrêt de référence.



```
$ python3 main.py --help
usage: main.py [-h] [-u USERNAME] [-e EMAIL] [-f FIRSTNAME] [-l LASTNAME]

Python OSINT

optional arguments:
  -h, --help            show this help message and exit
  -u USERNAME, --username USERNAME
                        Username (nickname online)
  -e EMAIL, --email EMAIL
                        Email address
  -f FIRSTNAME, --firstname FIRSTNAME
                        Firstname (must come with a lastname as well)
  -l LASTNAME, --lastname LASTNAME
                        Lastname (must come with a firstname as well)
```

### Quelques ressources :

<https://openfacto.fr/category/trucs-et-astuces/page/2/>

[https://www.ndi.org/sites/default/files/280055%20NDI%20Data%20Analytics%20Guide\\_9\\_9\\_2021\\_FINAL.pdf](https://www.ndi.org/sites/default/files/280055%20NDI%20Data%20Analytics%20Guide_9_9_2021_FINAL.pdf)

<https://www.psbedu.paris/fr/actus/open-source-intelligence-technique-renseignement-service-intelligence-economique>

<https://fr.acervolima.com/twosint-outil-osint-automatise-pour-enqueter-sur-un-compte-twitter/>

<https://lampyre.io/>

<https://www.octoparse.fr/blog/top-30-des-logiciels-de-web-scraping-gratuits-en-2021#>

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>