

Big Data

Open Data

MMI 2 – TP#3 S4

The background features a dark gray field with two large, abstract geometric shapes. A bright yellow triangle is positioned on the left side, pointing towards the bottom-left corner. A light red triangle is located on the right side, pointing towards the top-right corner.

Danielo **JEAN-LOUIS**
Développeur front-end

Data-science (Science des données)

- Domaine en aval du big data (stockage) et en amont du produit (marketing)
 - Ex : Le pôle data-science va trouver une tendance pour sortir un nouveau produit
- Modernisé grâce aux algorithmes

Sources :

- <https://www.lebigdata.fr/data-science-definition>

Data-science (Science des données)

- Deux grandes branches (version moderne) :
 - data-architects : définit la plateforme technique
 - data-analystes : applique des algorithmes sur la donnée – C'est ce qu'on va voir en cours

Sources :

- <https://www.lebigdata.fr/data-science-definition>

Data-science traditionnelle

- Utilisation de logiciels payants, limités et coûteux en terme de mise à l'échelle
- Facile à utiliser, apprentissage facile
- Gestion du big data très limité
- Appelé aussi BI pour Business Intelligence

Sources :

- <https://www.lebigdata.fr/data-science-definition>

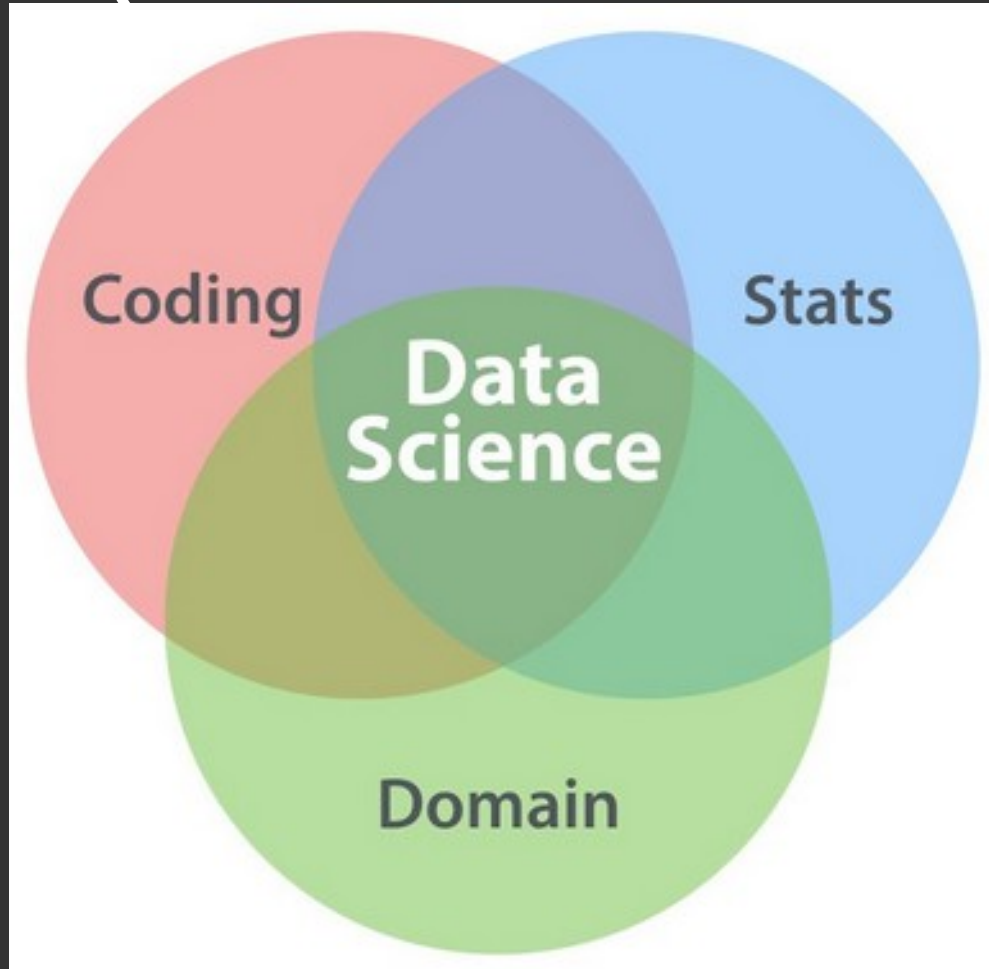
Data-science moderne (Science des données)

- Mélange des compétences suivantes :
 - Maths et statistiques
 - Métier / produit
 - Programmation et Algorithmique
- Pas besoin d'être excellent dans les trois domaines, mais de s'y connaître

Sources :

- <https://www.lebigdata.fr/data-science-definition>

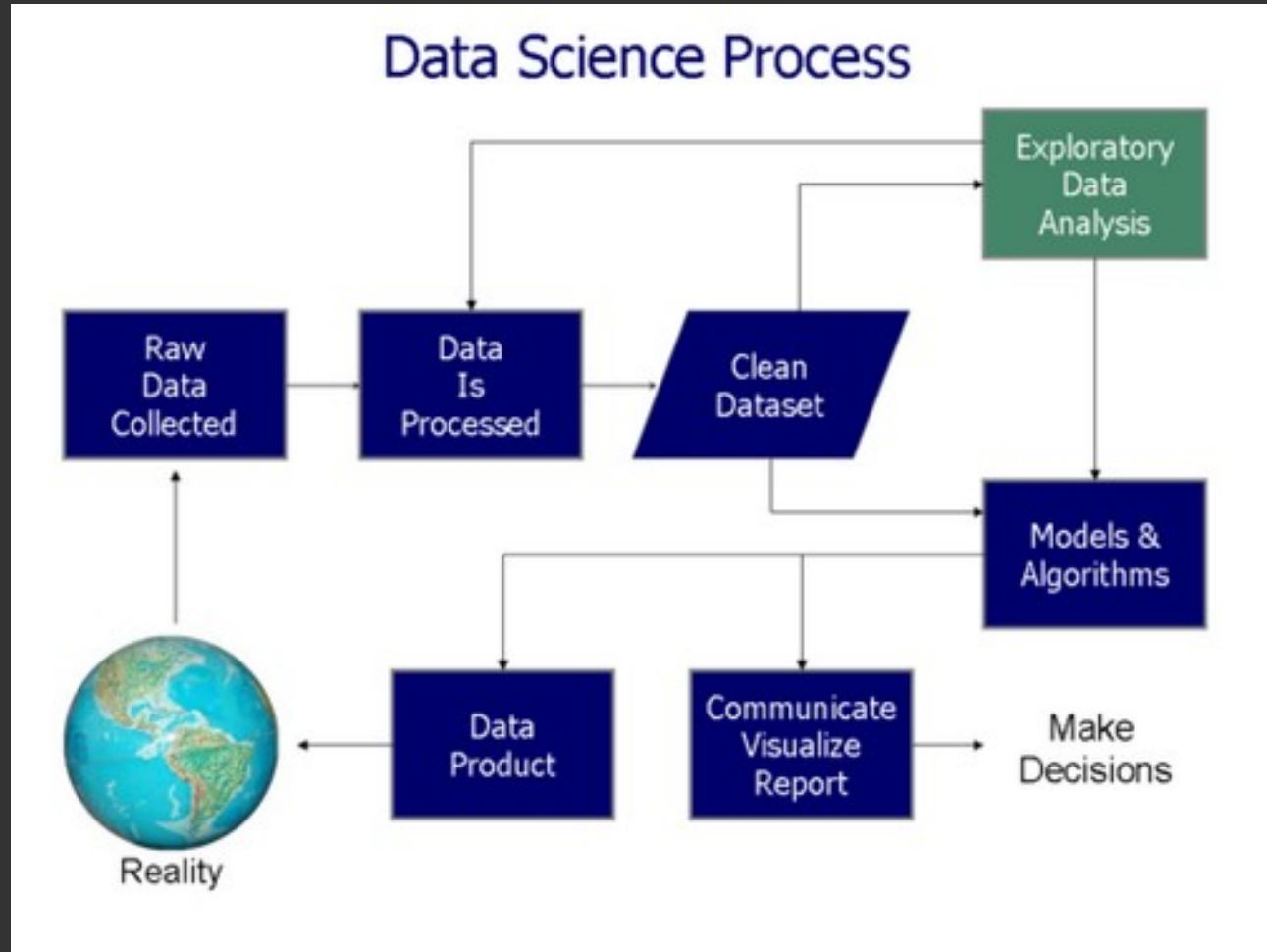
Data-science (Science des données)



Sources :

- <https://www.youtube.com/watch?v=r2l3IDKwyMw> - anglais

Processus



Sources :

- <https://www.kdnuggets.com/2016/03/data-science-process.html>

Langages de programmation du scientifique de la donnée

- Julia : Dédié et pensé pour la data-science
- R : Langage "matheux", syntaxe pouvant être déroutante de prime à bord
- Python : langage haut niveau → Syntaxe simple à assimiler

Langages de programmation du scientifique de la donnée - Popularité

1. Python – Celui qu'on utilisera
2. R
3. Julia (Langage récent - 2012)

Questions ?

