



## Chapitre 1

# Introduction

# Plan

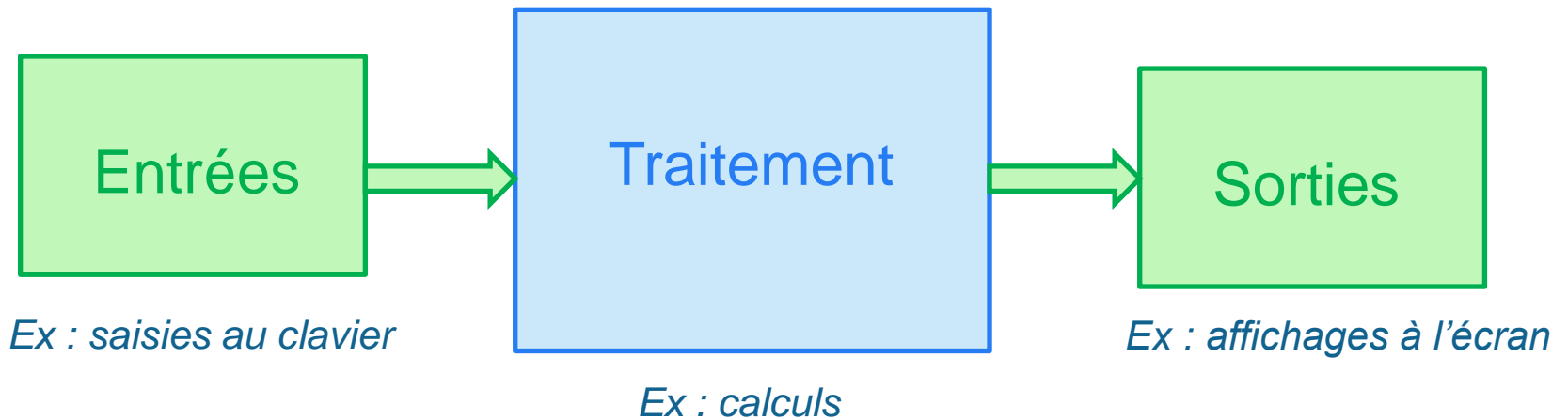
- Ordinateur
- Programme
- Donnée >< information
- Volatilité des données
- Persistance des données
- But de l'organisation des données
- Inconvénients des données mal structurées
- Fiche de cours

# Ordinateur

Machine, système de traitement de **l'information**  
+ programmable



# Programme



# Donnée >< information

## Donnée

**Enregistrement** dans un code donné d'un objet, texte, concept, fait ... en vue de transmettre ou stocker de l'information, interpréter ou effectuer un traitement par l'homme ou la machine, ou encore en déduire de nouvelles informations



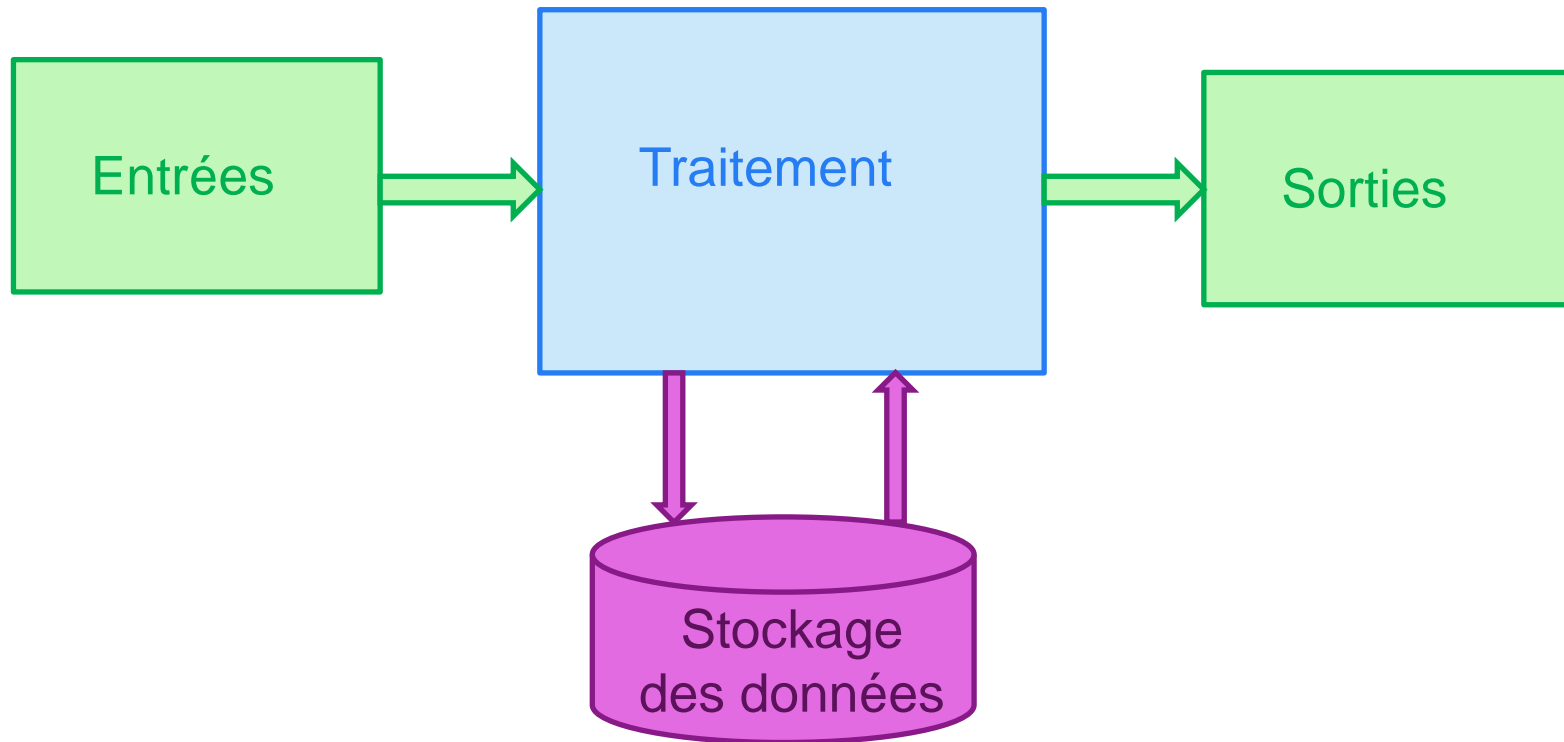
## Information

**Sens** qu'on donne à une donnée  $\Rightarrow$  subjectif

*Ex. de donnée : 181023*

*$\hookrightarrow$  Information possible : un prix  
un nombre d'habitants  
un code secret  
un matricule  
une date ...*

# Persistance des données



Problème : données perdues après traitement !

⇒ Les **enregistrer** (les persister) pour pouvoir les réutiliser dans le futur

# Persistance des données

Exemples de stockage des données :

Fichiers



Base de données



# But de l'organisation des données

Enregistrer des données réutilisables

⇒ correctes et cohérentes

⇒ Structurer et organiser les données





# Inconvénients des données mal structurées

Contre-exemple : mauvaise structure et organisation des données

Domaine d'application : gestion d'une librairie

Formulaire d'encodage d'un nouveau livre mis en vente

Informations demandées pour chaque nouveau livre :

Titre, nombre de pages,

Si c'est un livre destiné aux enfants,

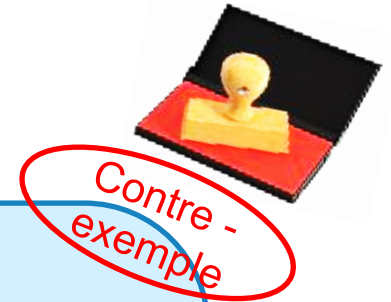
Nom, date de naissance et date de décès de l'auteur,

Prix de vente

# Inconvénients des données mal structurées

## Formulaire d'encodage d'un nouveau livre

|                               |                      |
|-------------------------------|----------------------|
| Titre                         | <input type="text"/> |
| Nombre de pages               | <input type="text"/> |
| Livre destiné aux enfants ?   | <input type="text"/> |
| Nom de l'auteur               | <input type="text"/> |
| Date de naissance de l'auteur | <input type="text"/> |
| Date de décès de l'auteur     | <input type="text"/> |
| Prix de vente                 | <input type="text"/> |



Avec tous les **champs** de type chaîne de caractères sans aucune restriction  
⇒ l'utilisateur peut introduire **n'importe quelle chaîne de caractères !**

# Inconvénients des données mal structurées

Hypothèse :

Chaque livre encodé via le formulaire est enregistré dans un fichier Excel :

Une ligne du fichier = enregistrement d'un livre

cf. *CatalogueLivres.xlsx*

À partir de ce fichier,  
tentez de répondre aux différentes demandes

# Inconvénients des données mal structurées

*Fichier CatalogueLivres.xlsx*

| Titre                        | Nb pages       | Pour enfant ? | Auteur           | Date naissance   | Date décès       | Prix         |
|------------------------------|----------------|---------------|------------------|------------------|------------------|--------------|
| Cubitus : l'intégrale 6      | 182 p.         | oui           | Dupa             | 12-february-1945 | 08-november-1940 | 17,95 euros  |
| Les 7 boules de cristal      | soixante-deux  | V             | HERGÉ            | 22 may 1907      | 3 march 1983     | 16,9 dollars |
| L'île mystérieuse            | 439 pages      | pour tous     | Jules Verne      | 8 FEV 1828       | 24 mars 1905     | 11,75        |
| Barbe bleue                  | 170            | Non           | Amélie Nothomb   | 9 juillet 66     |                  | 15,60 €      |
| Tintin au Tibet              | 62 p.          | OK            | Hergé            | 22 may 1907      | 3 mars 1983      | 15,95 euros  |
| Le petit prince              | nb. pages : 93 | pour enfants  | Saint-Exupéry    | 29-06-00         | 31 juillet 44    | 24,32        |
| Le lotus bleu                | 64             | OUI           | hergé            | 22 may 1907      | 3/3/1983         | 13,50 €      |
| Journal d'hirondelle         | 144 PAGES      | X             | Amélie Nothom    | 09-07-1966       |                  | 17,5         |
| Objectif lune                | 68             | 1             | Ergé             | 22/06/07         | 03/03/83         | 14,75 €      |
| La jeunesse de Mickey        | 80 PAGES       | < 12 ans      | Walt Disney      | 12/05/1901       | 12/15/1966       | 17 euro      |
| Oliver Twist                 | 608            | Non           | Charles Dickens  | 7 fevrier 1812   | 9 juin 1870      | 10,99        |
| La gloire de mon père        | 309 p          | NO            | Marcel Pagnol    | 29 février 1895  | 31 avr 74        | 7,50 €       |
| Les misérables               | 2.598          | Pas OK        | Victor Hugo      | 26 fév. 02       | 22 mai 85        | 11,99        |
| Le roi Babar                 | 48 p           | VRAI          | Jean de Brunhoff | 9/12/1899        | 16/10/1937       | 6,9          |
| Voyage au centre de la terre | 225            | tout public   | Verne Jules      | 18/02/1829       | 25/03/1905       | 14,5 eur     |
| Acide sulfurique             | 224 p          | Pas OK        | Notomb Amélie    | 9/7/ 66          |                  | 16           |
| L'île noire                  | 65 pages.      | VRAI          | HERGE            | 22/06/1907       | 3/3/83           | 13,6 euros   |
| La Nostalgie heureuse        | 162            | Pour adultes  | Amelie Nothomb   | 9 july 66        |                  | 15.95 eur    |
| De la terre à la lune        | 195 pg         | ok            | JULES VERNE      | 8 FÉVRIER 1818   | 26 mars 1905     | 9,8 euros    |

# Inconvénients des données mal structurées

## Demandes

*Quel est le prix de vente moyen des livres ?*

*Quel est le plus grand nombre de pages que contient un livre ?*

*Quels sont les livres qui coûtent entre 20 et 85 euros ?*

*Quel est le plus petit prix de vente ?*

⇒ Impossible de répondre car ces **calculs ne peuvent se faire** que sur des **nombre**s et pas sur des caractères



⇒ Il aurait fallu restreindre les valeurs qu'on peut introduire :

- Nombre de pages : uniquement des **entiers**  $>0$
- Prix de vente : uniquement des **réels**  $\geq 0$  (0 admis si gratuit)



# Inconvénients des données mal structurées

*Demande*

*Quels sont les livres destinés aux enfants ?*

Encodage libre  $\Rightarrow$  trop de valeurs possibles pour les livres pour enfants :  
*Oui, OUI, vrai, VRAI, V, 1, OK, ok, pour enfants, < 12 ans, tout public...*



Pfff... Trop de recherches à faire !

$\Rightarrow$  Il aurait fallu restreindre les valeurs qu'on peut introduire  
Ex : seulement deux valeurs **vrai** ou **faux**



# Inconvénients des données mal structurées

## Demande

Les dates "31 avr 74", "29 février 1895" et "12/15/1966" sont-elles valides ?

Max 30 jours en avril  $\Rightarrow$  ~~31 avr 74~~

1895 pas année bissextile  $\Rightarrow$  ~~29 février 1895~~

N.B. 29 février 1988 ✓ car année divisible / 4  
~~29 février 1900~~ sauf multiple de 100 (siècle)  
29 février 2000 ✓  $\Rightarrow$  exception : multiple de 400



Année bissextile ?

Si format français  $\Rightarrow$  ~~12/15/1966~~  $><$  si format anglais  $\Rightarrow$  12/15/1966 ✓

# Inconvénients des données mal structurées

*Demande*

*Quelle est l'année de naissance des auteurs du Petit Prince et des Misérables ?*

Année de naissance encodée pour l'auteur du Petit Prince : 29-06-00

⇒ 1800 ? 1900 ? 2000 ?

Année de naissance encodée pour l'auteur des Misérables : 26 fév. 02

⇒ 1702 ? 1802 ? 1902 ?



# Inconvénients des données mal structurées

*Demande*

*Quels sont les auteurs nés en février ?*

Beaucoup de valeurs encodées pour ce mois

⇒ Trop de recherches à faire !



# Inconvénients des données mal structurées

## Conclusion

Si champs de type chaîne de caractères pour encoder des dates  
⇒ les dates peuvent être non valides et non exploitables

⇒ Il aurait fallu imposer des champs de type date :  
seulement des dates valides peuvent y être encodées



# Inconvénients des données mal structurées

On a un problème : d'après les données encodées, Dupa est décédé avant sa naissance !



⇒ Il aurait fallu ajouter la contrainte :  
la date de décès doit être postérieure à la date de naissance



# Inconvénients des données mal structurées

*Demande*

*Quels sont les titres des ouvrages de Hergé ?*

Les informations concernant Hergé ont été encodées plusieurs fois avec des orthographes différentes

*Hergé, HERGE, HERGÉ, hergé, Ergé...*

⇒ Plusieurs recherches à faire !



# Inconvénients des données mal structurées

*Demande*

*Quand est né Jules Verne ?*



Les informations concernant Jules Verne ont été encodées plusieurs fois

Des dates de naissance différentes ont été encodées :

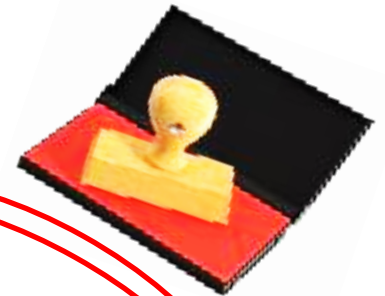
*8 FEV 1828, 18/02/1829 et 8 FÉVRIER 1818*

⇒ Erreur d'encodage

# Inconvénients des données mal structurées

*Demande*

*Quand Amélie Nothomb décèdera,  
combien de fois faudra-t-il encoder sa date de décès ?*



Les informations concernant Amélie Nothomb ont été encodées plusieurs fois

À son décès, il faudra encoder 4 fois sa date de décès

⇒ **Modification à exécuter à plusieurs endroits**

**Redondance**

# Inconvénients des données mal structurées

## Conclusion

Si redondance de données :

- Risque d'erreur d'encodage  $\Rightarrow$  incohérence des données
- En cas de modification de donnée  $\Rightarrow$  mises à jour à effectuer à plusieurs endroits

$\Rightarrow$  Il aurait fallu  
n'encoder qu'une seule fois les données de chaque auteur



# Inconvénients des données mal structurées



OK ! Je comprends bien l'intérêt d'éviter la redondance et de ne stocker qu'une seule fois les données de chaque auteur.  
***Mais alors, comment fait-on ?***

On sépare les livres et les auteurs dans deux entités séparées, puis on établit une association entre ces deux entités.

***C'est ce que nous allons voir dans le chapitre suivant !***





## Acquis d'apprentissage

Au terme de l'activité d'apprentissage, l'étudiant sera capable de :

- Interpréter un schéma entités-associations qui modélise des concepts et des liens entre ces concepts
- Traduire un schéma entités-associations en un schéma adapté aux bases de données relationnelles (schéma relationnel)
- Interpréter un schéma relationnel donné
- Créer à partir d'un schéma relationnel donné la structure d'une base de données relationnelle ou modifier la structure d'une base de données relationnelle existante, via des instructions SQL
- Insérer, modifier et supprimer des données dans une base de données relationnelle via des instructions SQL
- Interroger une base de données relationnelle via des requêtes SQL

# Fiche de cours

Modalités d'évaluation

Interro formative

Examen : 100 %