

Sciences Numériques et Technologie

Ressources dans <https://philippereneviergonin.github.io/snt/>

Philippe.Renevier-Gonin@ac-grenoble.fr

Les données

- Repères historiques
- Les structures et les formats de données
- Les opérations sur les données
- Les métadonnées

Les repères historiques

- Page 14-15 du manuel « SNT, édition Delagrave »
- 1725 – Rubans perforés puis cartes perforées
- 1956 – 1^{er} disque dur
- 1970 – modèle relationnel (base de données)
- 1979 – 1^{er} tableur (ancêtre de excel)
- 1984 – mémoire flash (usb, ssd)
- 1997 – Big Data (traitement d'énormes quantités d'informations)
- 2009 – Open Data : données libres d'accès : OpenStreetMap vs GoogleMaps

Qu'est-ce qu'une donnée ?

- une valeur ou un ensemble de valeurs
 - collectées,
 - enregistrées,
 - traitées ou analysées.(exemple la circulation dans waze)
(les applications waze envoient les infos à Waze)
(stockage pour estimer des temps de trajet)
(sur votre application, vous avez un trajet selon les infos)
- Les données peuvent prendre différentes formes et proviennent de diverses sources. Elles sont essentielles pour la prise de décision, la recherche, l'analyse et la communication.
- Quelques types de données :
 - Données numériques (ex : températures, ventes, scores, ...)
 - Données textuelles : (des commentaires, des posts, des documents, des emails...)
 - Données catégorielles : (couleurs, types de produits, ...)
- Les informations peuvent être reliées :
 - au temps (Données temporelles)
 - Et/ou à l'espace (Données géographiques, géolocalisation)

Les structures de données

- Les données peuvent être brutes (non traitées) ou structurées (organisées dans un format spécifique). Elles jouent un rôle crucial
- Structurer les données facilite leur exploitation (utilisation, compréhension, analyse, etc.)
- La structuration dépend de l'usage
 - Autrement dit : une donnée est adaptée à son utilisation

Le format csv

- Un fichier texte avec un séparateur entre les données
 - En France : « ; »
 - Cela peut être aussi « , » ou une tabulation
- Lecture/Écriture avec
 - un éditeur de texte (notepad++ par exemple)
 - Avec un tableau (excel par exemple)
- Généralement « en colonne » : les informations au même niveau sur des lignes différentes contiennent des informations de même nature
- Exemple :

```
ANAI S;ACCOUCHR;AGEMERE;AGEPERE
```

```
2021;ES;31;34
```

```
2021;ES;32;32
```

```
2021;ES;25;35
```

```
2021;ES;28;39
```

Représentation plus lisible du csv

ANNAIS	; ACCOUC	HR	; AGE	MERE	; AGE	PERE
2021	; ES		; 31		; 34	
2021	; ES		; 32		; 32	
2021	; ES		; 25		; 35	
2021	; ES		; 28		; 39	

La structuration est en « table »

Le format xml

- Un format texte avec des « balises » qui encadrent les valeurs
 - Un fichier xml suit une grammaire qui définit les balises, leur signification et ce qu'elles peuvent contenir
 - Une balise peut contenir du texte ou des nombres, mais aussi d'autres balises
- Lisible par une personne, utilisable par les programmes.
- HTML est une forme particulière de XML
- Exemple :

```
<naissance>  
    <année>2021</année>  
    <condition_d_accouchement>ES</condition_d_accouchement>  
    <âgeDeLaMère>31</âgeDeLaMère>  
    <âgeDuPère>34</âgeDuPère>  
</naissance>
```


Le format JSON

- C'est un format plus « léger que le xml »
- Utilisé dans les échanges entre programme (pris la place du xml)

```
{ "naissance": { "année": 2021,  
"condition d accouchement": "ES", "âgeDeLaMère": 31,  
"âgeDuPère": 34 } }
```

Ou en version plus lisible :

```
{ "naissance":  
  { "année": 2021,  
    "condition d accouchement": "ES",  
    "âgeDeLaMère": 31,  
    "âgeDuPère": 34 }  
}
```

Descripteur des données

- Généralement la première ligne d'un csv (mais ce n'est pas obligatoire)
- Les balises utilisées dans le XML
- Les propriétés au format JSON
- Un objet sera un ensemble de données
 - Une ligne d'un csv
 - Le contenu d'une balise en XML
 - Ce qui est entre { } en JSON
 - Les caractéristiques de « l'objet » sont les descripteurs


Opérations sur les données au format « texte »

- Un programme peut lire les fichiers de données
- Le traitement peut être « générique »
 - Affichage d'une page web dans un navigateur
 - Ouverture d'un fichier csv dans excel
- Le traitement peut être spécifique
 - Il y a un programme qui lit, analyse et produit un résultat
 - Ex : un programme python
 - Ex : un code javascript dans une page web qui reçoit des données

Avantages / inconvénients des formats textes

- Avantages :
 - Facilement utilisables / réutilisables : ce sont des fichiers que vous pouvez « lire »
 - Peuvent servir d'import/export entre logiciels
 - Pour csv : adapter aux « tables », manipulables par un éditeur de texte et par les tableurs, simple
 - Pour XML : lisible, évolutif, peut contenir sa grammaire
 - Pour JSON : populaire en programmation, lisible, évolutif
- Désavantages :
 - Exploitation généralement à « programmer » ou « à refaire »
 - Csv : peu souple (forcément une table, pas de retour à la ligne dans les valeurs)
 - Xml : très verbeux, nécessite un programme pour son
 - JSON : nécessite un programme et des notions de programmation orientée objet

Le format « sql » dans une base de données

- Les bases de données (BD) sont des systèmes qui
 - Archivent
 - Classent
 - Indexent
 - Lisent/Choisissent
 - (ajoutent, modifient, suppriment)

les données
- Pour les BD relationnelles, les données sont structurées en table
 - Comme en csv
 - Mais des liaisons peuvent être faites en tables
 - Les fichiers sont internes à la BD et ne sont pas exploitables sans la BD
 - Nécessitent des requêtes pour les exploiter
- On parle de transaction : si l'opération s'arrête en cours d'exécution, la BD est censée revenir à un état cohérent

Opérations sur les données au format « sql »

- Il faut faire des requêtes
- Voici quelques mots clefs pour la « lecture » extraits du manuel « SNT édition Delagrave »
- SQL (Structured Query Language) est un langage informatique pour exploiter les données.
- Exemple de requête :
 - pour tout sélectionner de la table naissance_2021 :
`SELECT * FROM `naissance_2021`;`
 - Pour compter les filles nées en 2021 :
`SELECT COUNT(*) FROM `naissance_2021` where sexe = "Fille";`
où “sexe” est un des descripteurs (colonne) des données

Commandes SQL	Usage
<code>select</code>	Sélectionner une donnée
<code>from</code>	Dans une table
<code>where</code>	Avec une condition
<code>order by</code>	Trier le résultat dans l'ordre croissant
<code>count</code>	Compter
<code>avg</code>	Faire la moyenne

Les métadonnées

- donne des informations sur la donnée principale :
 - C'est une donnée descriptive
 - Qui accompagne une autre donnée
- Des exemples :
 - Les descripteurs des données donnent des informations sur les données
 - L'endroit où une photo a été prise
 - L'auteur d'un message sur les réseaux sociaux
 - Le nombre de « like » d'un message sur les réseaux sociaux