

Les données structurées et leurs traitements.

Organisation.

- 4 séances.
- Exemple filé du carnet d'adresse à construire et à exploiter.
- Une séance sur la nature des données. Données et métadonnées. Descripteurs. Collection. Base de données. Data science.
- Créer des tables de données, créer son propre carnet d'adresse. Partir de la version papier et voir comment structurer ? les inconvénients ?
- Une séance de traitement de données sous forme de csv. Utilisation du tableur. Représenter graphiquement les données.
- Une séance sur l'exploitation des données dans un csv par un logiciel non tableur. Recherche dans des données structurées. Trier. Filtrer.
- Une séance sur le stockage matériel de données et la consommation d'énergie correspondante. Big data.

Contenus et capacités (programme officiel).

Données	Définir une donnée personnelle. Identifier les principaux formats et représentations de données.
Données structurées	Identifier les différents descripteurs d'un objet. Distinguer la valeur d'une donnée de son descripteur. Utiliser un site de données ouvertes, pour sélectionner et récupérer des données.
Traitement de données structurées	Réaliser des opérations de recherche, filtre, tri ou calcul sur une ou plusieurs tables.
Métadonnées	Retrouver les métadonnées d'un fichier personnel.
Données dans le nuage (cloud)	Utiliser un support de stockage dans le nuage. Partager des fichiers, paramétrer des modes de synchronisation. Identifier les principales causes de la consommation énergétique des centres de données ainsi que leur ordre de grandeur.

Activités suggérées (programme officiel).

- Consulter les métadonnées de fichiers correspondant à des informations différentes et repérer celles collectées par un dispositif et celles renseignées par l'utilisateur.
- Télécharger des données ouvertes (sous forme d'un fichier au format CSV avec les métadonnées associées), observer les différences de traitements possibles selon le logiciel choisi pour lire le fichier : programme Python, tableur, éditeur de textes ou encore outils spécialisés en ligne.
- Explorer les données d'un fichier CSV à l'aide d'opérations de tri et de filtre, effectuer des calculs sur ces données, réaliser une visualisation graphique des données.
- À partir de deux tables de données ayant en commun un descripteur, montrer l'intérêt des deux tables pour éviter les redondances et les anomalies d'insertion et de suppression, réaliser un croisement des données permettant d'obtenir une nouvelle information.
- Illustrer, par des exemples simples, la consommation énergétique induite par le traitement et le stockage des données.

Activités.

En vrac :

- Données et données personnelles. Maison, quelles sont les données que j'utilise quelles sont mes données personnelles ?
- Créer un répertoire.
- Ouvrir un fichier csv avec un bloc note pour repéré les descripteurs déjà vus lorsque ouvert avec excel. Image d'un nouveau csv (impossible à ouvrir avec excel donc) et repérer à nouveau les descripteurs.
- Séance sur Python utilisation des modules pour le traitement de données. Gros fichier pour montrer la pertinence des langages de programmations plutôt que excel pour les big datas.
- Stocker les données : les divers support. Un logiciel de synchronisation et les différents paramétrages.
- Pour la leçon un texte ou une vidéo puis un schéma de tableau avec les noms à compléter, les différents types de tables usuels en informatique.

Séance 1 : données et tableurs.

I Préséance : repères historiques.

Voici quelques dates importantes sur les données numériques.

Les liens ci-dessous pointent vers des pages de Wikipedia, projet d'encyclopédie collective en ligne, et vous permettent de savoir de quoi on parle.

1. 1930 : utilisation des **cartes perforées**, premier support de stockage de données ;
2. 1956 : invention du **disque dur** permettant de stocker de plus grandes quantité de données, avec un accès de plus en plus rapide ;
3. 1970 : invention du **modèle relationnel** (E. L. Codd) pour la structuration et l'indexation des bases de données ;
4. 1979 : création du premier tableur, **VisiCalc** ;
5. 2009 : **Open Government Initiative** du président Obama ;
6. 2013 : **charte du G8** pour l'ouverture des données publiques.

Approfondissez vos connaissances sur l'histoire des données numériques.

Sachez utiliser les liens donnés dans la liste des repères historiques pour répondre aux vingt questions suivantes.

Exercice 1

L'utilisation des cartes perforées comme premier support de stockage de données date :

- o du XVIIe siècle,
- o des années 1930,
- o des années 1960.

Exercice 2

Par quelle technologie bon marché les cartes perforées ont-elles été progressivement remplacées dans les années 1970 ?

- o Les clefs USB.
- o Les disques durs.
- o Des bandes magnétiques.

Exercice 3

Les cartes perforées ont été utilisées avec :

- o Le métier à tisser Jacquart.
- o La machine analytique de Charles Babbage.
- o Ipad.
- o Les premières machines de la marque IBM.

Exercice 4

En quelle année a été inventé le disque dur ?

- 1912.
- 1956.
- 2002.

Exercice 5

Quel était le prix de 1 Mo du premier disque dur IBM 350 Ramac ?

- 1 \$.
- 100 \$.
- 10000 \$.

Exercice 6

En 1980, le prix de 1 Mo du modèle ST-506 est de :

- 300 \$.
- 3000 \$.
- 30000 \$.

Exercice 7

En 2008, le mégaoctet d'un disque dur ne coûte plus qu'environ ...

- 2 \$.
- 0,02 \$.
- 0,0002 \$.

Exercice 8

Quelle est la capacité du premier disque dur IBM 350 Ramac ?

- 1 Mo
- 5 Mo
- 5 Go

Exercice 9

Qu'est ce qu'un modèle relationnel ?

- Le modèle relationnel est une manière de se faire connaître pour avoir beaucoup de relations.
- Le modèle relationnel est une manière de modéliser les relations existantes entre plusieurs informations, et de les ordonner entre elles.

Exercice 10

En quelle année a été inventé le modèle relationnel ?

- 1930
- 1970
- 2001

Exercice 11

Les modèles relationnels sont utilisés dans la pratique dans :

- Les bases de données.
- Les relations intergouvernementales.

Exercice 12

Qu'est-ce que Visicalc ?

- Le premier tableur.
- La première calculatrice de la marque Visi.

Exercice 13

Pour quel ordinateur le logiciel VisiCalc a été conçu en 1979 ?

- IBM PC.
- Apple II.
- Iphone 4.

Exercice 14

Aujourd'hui, des tableurs utilisés se nomment :

- LibreOffice Calc.
- Microsoft Excel.
- Microsoft Word.
- Numbers.
- Google Sheets.
- OpenOffice Table.

Exercice 15

L'Open Government Initiative du président Obama date de :

- 2000
- 2009
- 2019

Exercice 16

Quel est le but de l'Open Government Initiative du président Obama ?

- Ouvrir les portes de la maison blanche à tout citoyen américain.
- Créer un niveau sans précédent de transparence et d'ouverture du gouvernement américain.
- Ouvrir les portes du congrès américain.

Exercice 17

L'Open Government Initiative du président Obama s'inscrit une tendance visant à appliquer en politique et dans la gouvernance des démocraties

- des philosophies telles que l'open source et les « contenus ouverts » (Open data).
- d'une philosophie consistant à ouvrir un dialogue entre les citoyens et les dirigeants.

Exercice 18

La charte du G8 pour l'ouverture des données publiques date de :

- 2002
- 2013
- 2019

Exercice 19

Charte du G8 : un communiqué commun des chefs d'État a expliqué que cette charte contribuera à :

- accroître la quantité de données publiques ouvertes dans les secteurs essentiels de l'action publique.
- accroître la quantité de pages web.
- à soutenir le débat démocratique.
- à soutenir les initiatives commerciales de marchandisation des données.
- et à faire en sorte de faciliter la vente des données publiques.
- et à faire en sorte de faciliter la réutilisation des données publiées.

Exercice 20

En France, la charte doit être appliquée par :

- Etalab, service placé aux côtés du premier ministre et chargé de la question de l'ouverture des données publiques.
- aucun organisme, les données publiques n'appartenant à personne en particulier.
- les fournisseurs d'accès à l'internet.

II Les données, les métadonnées et les données personnelles. (Sans PC)

Données personnelles : <https://donneespersonnelles.fr/donnees-personnelles>

Vidéo c'est pas sorcier (à télécharger) : <https://www.youtube.com/watch?v=7NKEBW0cgZE>

Vidéo à regarder sur le buisness des données personnelles : <https://www.youtube.com/watch?v=dztbDIPux9A>.

C'est pas sorcier des vidéos à tester

[facebook](#)

Représentant de la CNIL un peu lourd pas très pratique : [donnees personnelles telephone nous espionne](#)

Bien mais plante en permanence : [protéger sa vie privée](#).

Débat chiant : [big data](#)

Hors sujet : [wifi et téléphone la fin de la vie privée](#)

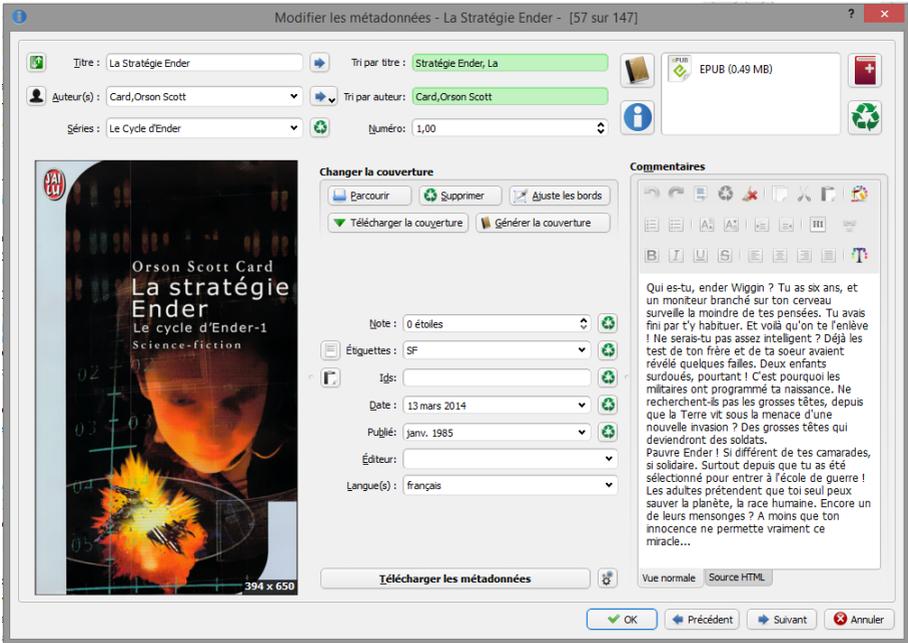
Les *données* sont les informations brutes (pas encore interprétées) que nous souhaitons conserver. Depuis très longtemps nous enregistrons et stockons différentes données (comptabilité d'un commerce, registres de baptêmes, etc) de différentes façons. Actuellement elles sont le plus souvent enregistrées via l'informatique.

Le monde économique enregistre de grandes quantités de données (souvent des nombres) qui sont enregistrées et stockées sous forme de tableaux.

Nous conservons d'autres formes de données : des textes, des images (photographies), des sons (musique, chansons) ou des films.

Les sites web conservent les adresses IP et les pages consultées par chaque internaute venant sur son site.

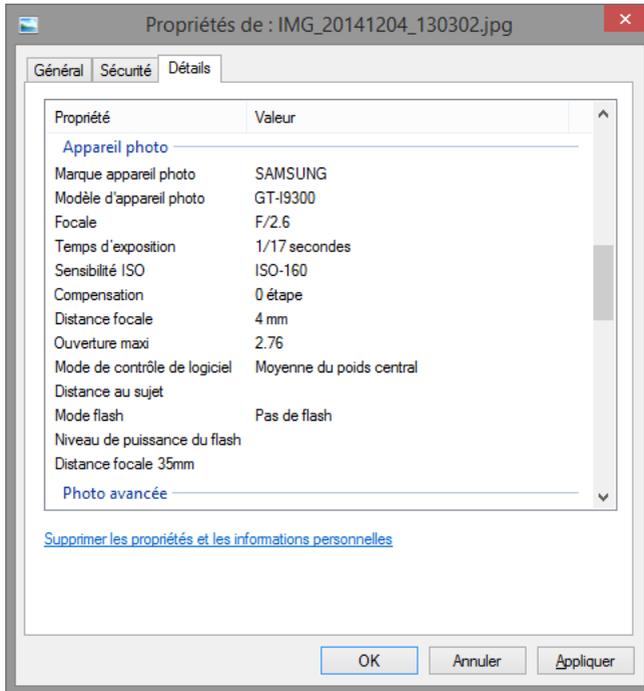
Les *métadonnées* sont les informations ajoutées à d'autres données pour les classer, les préciser. Ainsi lors de la création d'une photographie l'appareil photographique ajoute à l'image (qui est elle-même un ensemble de données) des informations sur la résolution de l'image, la date et l'heure, la localisation GPS, etc.



Métadonnées pour un livre

Sous le système d'exploitation Windows, on peut accéder aux métadonnées d'un fichier en effectuant successivement les actions suivantes :

1. Clic droit sur son icône.
2. Choisir Propriétés.
3. Cliquer sur l'onglet "Détails".



Métadonnées pour une photographie sous Windows.

Voici d'autres exemples de métadonnées :

```

«id»=>14465482354176,
«text»=>
«Tu crois vraiment que c'est vrai ? perso j'ai vu des articles qui indiquent que non !
http://www.monsite.com/indication que non»,
«created at»=>«Wed Apr 10 23:56:41 +0000 2019»,
«in_reply_to_usuer_id»=>ACastanet,
«in_reply_to_screen_name.»=>A. Castanet,
«in_reply_to_status_id»=>IsalineMA,
«favorited.»=>true
«user.»=>
  «id»=>6253282,
  «screen_name»=>E. Chryssos»,
  «name»=>E. Chryssos»,
  «description»=>
  «Ceci est ma description.»,
  «url»=>«https://twitter.com/E_Chryssos»,
  «location»=>Nancy, FR»,
  «profile_background_color»=>cldfee»,
  «profile_background_image_url»=>
  «https://www.nathan.fr»,
  «profile_link_color»=>0000ff»,
  «profile_sidebar_border_color»=>87bc44»,
  «profile_sidebar_fill_color»=>e0f92»,
  «profile_text_color»=>000000»,
  «created at»=>«Thu June 12 14:13:51 +0000 2014»
  «contributors_enabled»=>true,
  «favourites_count»=>1628,
  «statuses_count»=>13417,
  «friends_count»=>448,
  «time_zone»=>Paris (France)»,
  «utc_offset»=>+3600

#Identifiant du Tweet
#Contenu du tweet
#Date de création du tweet
#Identifiant @ de l'utilisateur à qui ré
#Nom de l'utilisateur à qui répond l'au
#Identifiant @ de l'utilisatrice qui a é
#L'auteur a aimé le tweet

#Identification du compte Twitter dans l
#Nom de l'auteur du tweet
#Identifiant @ de l'auteur du tweet

#Description du profil de l'auteur du tweet
#Adresse URL du profil
#Ville depuis laquelle l'auteur écrit son tweet
#Thème de couleur du profil de l'auteur

#Image d'arrière-plan choisie par l'utilisateur pour sa page Twitter
#Pour cette catégorie et les 3 suivantes : couleurs des caractères et des
barres latérales choisies par l'utilisateurs ou par défaut

#Date de creation du profil
#L'utilisateur permet à quelqu'un d'autre de publier des messages en son nom
#Nombre de tweets favoris de l'utilisateur
#Nombre de tweets écrits par l'utilisateur
#Nombre de comptes que l'utilisateur suit
#Fuseau horaire dans lequel se trouve l'auteur
#Décalage en minute par rapport au fuseau horaire de Greenwich

```

Métadonnées pour un tweet.

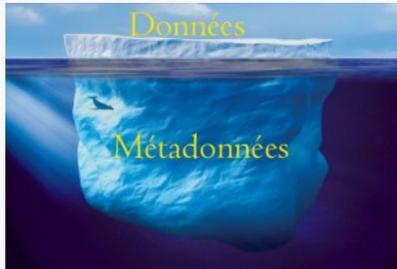
```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html><head>
<meta http-equiv="Content-Language" content="fr"> <meta name="author" content="Pierre Achethémel">
<meta name="category" content="software"> <meta name="DC.Creator" content="Pierre Achethémel - Mon guide">
<meta name="DC.Title" content="M&eacute;tadonn&eacute;es: une initiation">
<meta name="DC.Title.Subtitle" content="Dublin Core, IPTC, EXIF, RDF, XMP">
<meta name="description" content="Techniques relatives aux métadonnées (IPTC, EXIF)">
<meta name="GENERATOR" content="Microsoft FrontPage 4.0">
<meta name="keywords" content="ressources,XML,standards informatiques">
```

Métadonnées pour une page HTML.

Les métadonnées jouent un rôle de plus en plus important au vu de la surabondance d'information sur internet : <https://www.youtube.com/watch?v=Bn0asZOS3GE> (2 minutes).

Exercice 21

1. En vous inspirant de l'exemple des métadonnées pour le livre (voir plus haut) imaginez les catégories pertinentes de métadonnées pour décrire un film (ou un jeu vidéo si vous préférez).
2. Quelle est la nature des métadonnées nécessaires à l'envoi d'un tweet ?
3. En considérant l'exemple des métadonnées d'un tweet indiquez ce qu'illustre l'image de l'iceberg ci-dessous.



4. Quel est le nom de l'émetteur du tweet ? Est-ce obligatoirement une identité exacte ? Combien de fois ce message a-t-il été retweeté ? Pourquoi la photographie indiquée en bas n'apparaît-elle pas ici ?
5. Qui est l'auteur de la page HTML ? Avec quel logiciel a-t-elle été créée ? Qu signifie Dublin core ?

Je mets quatre images de métadonnées et je pose des questions.

Le libre accès au métadonnées présente néanmoins des risques : https://www.youtube.com/watch?v=kz3Zb_Y_wJw (5 minutes). C'est pourquoi il a fallu définir la notion de donnée personnelle.

Les *données personnelles* sont celles qui permettent de connaître la vie, les goûts, les habitudes, les idées d'une personne qu'il est possible d'identifier : les photographies personnelles, les sites internet consultés, les listes d'amis, les notes scolaires, etc.

Les données personnelles et la loi : <https://www.youtube.com/watch?v=7NKEBW0cgZE> (10 minutes).

Exercice 22

QCM sur les données, métadonnées et données personnelles : vocabulaire et distinguer ce qui relève de la donnée personnelle ou pas.

1. En quoi les métadonnées facilitent-elles l'accès aux documents ?
2. Les métadonnées peuvent-elles permettre de s'immiscer dans le vie privée de quelqu'un ? Expliquez par un exemple.
3. En quoi la connaissance des données personnelles d'autrui peuvent-elles être problématiques ? Illustrez par des exemples.
4. Quel sont les droits de l'utilisateur concernant ses données personnelles ?
5. Quel est, en France, l'organisme qui s'occupe de vérifier qu'il n'y a pas d'utilisation illégale des données notamment personnelles ?

III Distinguer les différents formats de fichiers. (Sans PC)

Petite leçon sur les fichiers (les exemples les plus classiques) puis QCM sur les différents formats avec recherche sur internet si besoin est.

Les données informatiques sont enregistrées de différentes façons (nous parlons de *formats*) dans des *fichiers*.

Un fichier informatique contient toujours au moins une métadonnée : son nom. Le plus souvent le nom du fichier est suivi de son format pour indiquer le type de données dont il s'agit. Ainsi le fichier « *last_splash.mp3* » s'appelle en fait *last_splash* et l'extension *.mp3* indique qu'il s'agit d'un fichier audio.

Le format d'un fichier est indiqué à la fin du nom du fichier après un point.

Exercice 23

Regroupez les fichiers d'après leurs extensions suivant les catégories : audio, vidéo, texte, table (tableau de données), image, programme (application), archive.



IV Explorer un fichier table avec un tableur et trier les données. (Avec PC)

Programme : *Identifier les différents descripteurs d'un objet. Distinguer la valeur d'une donnée de son descripteur.*

Petite leçon de vocabulaire sur les trucs utilisés comme entête et autres (descripteurs).

Programme : *Utiliser un site de données ouvertes, pour sélectionner et récupérer des données.*

Télécharger un fichier csv et l'ouvrir avec notepad. téléchargés.

Les données sont très souvent rangées sous forme de *tables* c'est-à-dire de tableaux dans lesquels la première ligne sert à décrire la forme des lignes suivantes. Dans une table le *descripteur* ou *nom du champ* est l'intitulé, inscrit dans la première ligne, qui indique la nature des informations contenues dans une colonne.

Nom	Prénom	Taille
Fontaine	Philomène	175 cm
Hoarau	Amédée	160 cm

Dans la table ci-contre les champs sont « Nom », « Prénom » et « Taille ». Nous dirons que « Taille » est le descripteur de la donné « 175 cm ».

Le site de données ouvertes data.gouv.fr regroupe plusieurs milliers de fichiers disponibles sur tous les sujets. On y trouve aussi bien des données sur la pêche que sur l'enseignement, la population ou la localisation des vélos en libre-service. Ces données sont librement utilisables et peuvent permettre à chacun de mieux comprendre un domaine d'activité en ayant accès aux données de ce domaine.

Exercice 24

1. Allez sur le site data.gouv.fr et téléchargez le fichier, au format *csv*, recensant les Poilus morts pendant la première guerre mondiale.
2. Ouvrez le fichier téléchargé (*tracesdesoldats-lespoilus.csv*) avec un tableur (LibreOffice Calc) : clic droit sur le fichier puis « Ouvrir avec ».
3. Quels sont les trois premiers descripteurs de cette table.
4. En utilisant les fonctions de tri ([vidéo d'explication de 4 minutes](#)) du menu « Données » du tableur, déterminez le nombre de personnes dont le département de décès est le Loir-Et-Cher.
5. En utilisant les fonctions de tri du tableur déterminez le nombre de personnes nées à Brest et mortes dans la Meuse.
6. En utilisant les fonctions de filtres du tableur ([vidéo d'explication 4 minutes](#)) déterminez le nombre de Poilus nés en 1875.

Exercice 25

Pour l'exercice suivant il faut utiliser le fichier de l'INSEE des [populations légales](#). Les réponses pourront être écrites sur une nouvelle feuille du fichier. L'exercice s'intéresse à la feuille *Communes*.

1. Observer le fichier fourni par l'INSEE, comment les communes sont-elles triées ?
2. Les communes de moins de 10 000 habitants sont recensées par roulement tous les 5 ans, alors que pour les communes dont la population est supérieure ou égale à 10 000 habitants, une enquête par sondage est réalisée chaque année. Quel est le pourcentage des communes concernées par les enquêtes par sondage ?
3. Choisir l'un des diagrammes proposés par le tableur pour représenter les données. Cette représentation vous semble-t-elle satisfaisante ?

Exercice 26

Téléchargez le fichier des **populations des communes**.

1. (a) Trouvez (**ctrl+F**) une commune dont la population est de 9 053.
 (b) Colorez en bleu le fond de la cellule contenant le nom de cette commune.
2. Calculez la population moyenne des communes françaises (vous pourrez utiliser la formule « `=somme()` »).
3. Déterminez la population médiane des communes françaises.
4. Déterminez les premier et troisième quartiles de cette série statistique.
5. Complétez chacune des affirmations suivantes :
 - (a) Environ 25% des communes françaises ont une population inférieure à ...
 - (b) Environ ... des communes françaises ont une population comprise entre 191 et 1041 habitants.
 - (c) Environ 50% des communes françaises ont une population inférieure à ...
 - (d) Environ 25% des communes françaises ont une population supérieure à ...
6. Illustrez les résultats obtenus sur une droite graduée.

Exercice 27

Téléchargez le fichier des **populations des communes des départements de l'Ardèche et de la Drôme**.

1. Comparez les populations des villes de deux départements voisins à l'aide des médianes, quartiles, moyennes.
2. Une étude est menée pour un éventuel regroupement de ces deux départements.
 - (a) Quelle serait la nouvelle population moyenne? La nouvelle population médiane?
 - (b) Comment calculer la population moyenne du regroupement à partir des moyennes?
 - (c) Peut-on trouver la nouvelle médiane à partir des populations médianes des deux départements?

Séance 2 : trier, filtrer des tables.

I Disséquer un fichier csv. (Avec PC)

Faire une activité de tri avec les fichiers csv déjà téléchargés trop gros pour être manipulé avec tableur.

Le format *csv* (*comma separated values*) est un format de table ouvert extrêmement simple. Un fichier csv est un fichier texte autrement dit il peut être ouvert par un simple éditeur de texte (*Notepad* ou *Blocnote* sur Windows, *TextEdit* ou *Bracket* sur Mac).

Exercice 28

1. Nous allons regarder comment est construit un fichier csv en utilisant le fichier de l'exercice précédent.
 - (a) Ouvrez le fichier *tracesdesoldats-lespoilus.csv* avec un éditeur de texte (*Blocnote* sur Windows, *TextEdit* sur Mac).
Indication pour les ordinateurs du lycée : ouvrez d'abord le logiciel puis choisissez d'ouvrir le fichier en question.
 - (b) Retrouvez les descripteurs de la première lignes. Comment sont séparés les différents champs ? Comment sont séparées les différentes lignes de la table ?
2. Essayons de procéder de même avec un autre fichier contenant une table.
Téléchargez un fichier au format *ods* (par exemple celui-ci) créé avec le tableur Calc de LibreOffice.
Ouvrez-le avec un éditeur de texte (*Notepad* sur Windows, *TextEdit* sur Mac).
Que remarquez-vous ?
3. Nous allons maintenant créer un répertoire téléphonique au format *csv*.
 - (a) Ouvrez un éditeur de texte puis créez les champs des descripteurs « nom », « prenom », « mail », « telephone ». N'oubliez pas de les séparer avec le symbole qui convient.
 - (b) Enregistrez votre fichier en le nommant *repertoire.csv*.
Attention avec Windows, choisissez : sélectionnez tout type de fichier.
 - (c) Ajoutez les noms, prénoms, mails et téléphones imaginaires de trois personnes dans le fichier.
 - (d) Après avoir enregistré et fermé le fichier ouvrez-le à nouveau mais avec Calc de LibreOffice.

II Rechercher les données d'une table avec python. (Avec PC)

Lorsque les fichiers de données sont gros il peut être difficile pour un logiciel tableur de l'ouvrir. Dans ce cas il est pertinent d'utiliser Python pour rechercher dans le fichier.

Module python pour csv <https://docs.python.org/fr/3/library/csv.html>

Exercice 29

1. Téléchargez le fichier de la base nationales des adresses pour la Réunion : [ban-974.csv.gz](#) (archive au format .gz). Extrayez l'archive.
2. Voici un programme écrit en Python qui recense le nombre d'adresses dans la commune des Avirons.

```

1 import csv #Appel les fonctions pour lire le fichier csv
2 f=open('ban-974.csv', newline='') #Ouvre le fichier csv qui est appelé f
3 entete=f.readline() #Lecture de la première ligne qui est appelée entete
4 ligne=f.readline() #Lecture de la deuxième ligne qui est appelée ligne
5 compteuroccurrence=0
6 while ligne != "": #Tant que la ligne n'est pas une ligne vide il faut faire ce qui suit
7     id_ban_position,id_ban_adresse,cle_interop,id_ban_group,id_fantoir,numero,suffixe,nom_voi
8     #de la lign est séparé suivant les différents descripteurs
9     if nom_commune=="Les Avirons": #Si le descripteur nom_commune contient Les Avirons alors
10        compteuroccurrence=compteuroccurrence+1
11        ligne=f.readline() #Lecture d'une nouvelle ligne
12 print(compteuroccurrence) #Affiche la valeur de compteuroccurrence
13 f.close() #Referme le fichier csv

```

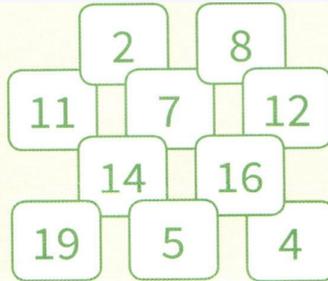
- (a) Téléchargez ce programme ([ici](#)). Copiez-le puis exécutez-le avec le logiciel Pyzo.
 - (b) Combien y a-t-il d'adresses aux Avirons ?
3. En modifiant le précédent programme python déterminez le nombre d'adresses qui sont des numéros BIS (le descripteur dans la table est *suffixe*).
 4. Nous souhaitons maintenant savoir combien d'adresses des Avirons sont des numéros BIS. Pour cela modifiez le précédent programme en combinant les deux conditions avec l'instruction Python `and`.

III Trier à la main. (Sans PC)

Exercice 30

Matériel

Préparer dix cartes de papier blanc de même taille (par exemple 5 cm × 5 cm). Inscrire sur chacune un entier choisi au hasard entre 1 et 20. Aligner ces cartes côte à côte sur une table, face inscrite cachée.



Objectif : Classer ces 10 cartes dans l'ordre croissant de gauche à droite.

Protocole

- Vous ne pouvez retourner que deux cartes simultanément.
- Vous pouvez inverser ou non les positions des deux cartes avant de les replacer face retournée.
- Vous ne devez pas mémoriser les valeurs inscrites sur les cartes au fur et à mesure de votre travail.
- Vous compterez le nombre de paires de cartes que vous aurez retournées et le nombre de paires de cartes que vous avez inversées.

Étapes

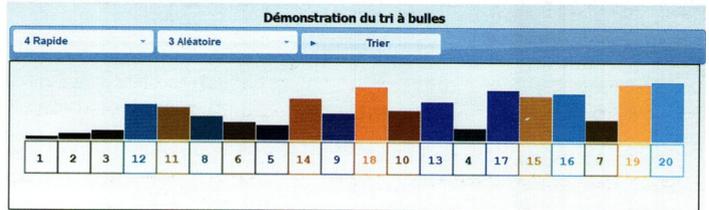
1. Quand vous pensez que votre tri est terminé, retourner les cartes pour vérification. Si elles ne sont pas correctement triées, recommencer après avoir replacé les cartes au hasard.
2. Donner un nom à votre méthode de tri et la rédiger sur feuille en utilisant obligatoirement les mots ou expressions suivants : d'abord – ensuite – enfin – jusqu'à ce que – tant que – alors – si.
Attention : toute personne doit être capable de tester votre méthode, à partir de votre texte.
3. Échanger la méthode d'un binôme avec celle d'un groupe voisin, et appliquer cette nouvelle méthode en comptant le nombre de comparaisons et d'échanges effectués.
4. Inscrire au tableau le nom de la méthode ainsi que le nombre de comparaisons et d'échanges réalisés pour trier les cartes. Noter quel groupe semble avoir la technique la plus efficace.

IV Différentes méthodes de tri. (Sans et avec PC)

Super vidéo sur les différents tris : <https://www.youtube.com/watch?v=ra79TDfotno>

Un exemple de tri : le tri à bulles

Le tri à bulle consiste à parcourir un tableau de valeurs tirées au hasard, de gauche à droite, en comparant les éléments côte à côte et en les permutant s'ils ne sont pas dans le bon ordre. Au cours d'un premier parcours, le plus grand élément remonte de proche en proche vers la droite, puis le tri recommence toujours à partir du premier élément, jusqu'à ce que l'ensemble soit trié.



D'autres exemples de tri sur [cette vidéo](#).

Exercice 31

1. Quels sont les différents algorithmes de tri évoqués dans la vidéo précédente ?
2. Que représente la complexité pour un algorithme (de tri) ?
3. Décrivez la tri par insertion.

Programmer un algorithme de tri à bulles sous Python

```

1 import random as rd # importation de la bibliothèque random
2 import time as ti # importation de la bibliothèque time
3
4 taille_tableau=20
5
6 #-----création d'une liste de valeurs aléatoires-----
7 T=[rd.randrange(10) for i in range(taille_tableau)]
8
9 #-----définition de la fonction de tri -----
10 def tri_bulle(liste):
11     """ cette fonction prend en entrée une liste quelconque
12     et retourne une liste triée """
13     for taille_liste in range(len(liste),2,-1):
14         for indice in range(taille_liste-1):
15             if liste[indice]>liste[indice+1]:
16                 liste[indice],liste[indice+1]=liste[indice+1],liste[indice]
17     return liste
18
19 #-----mesure du temps de traitement-----
20 date_debut=ti.time()
21 tri_bulle(T)
22 date_fin=ti.time()
23
24 print('durée du tri : ',date_fin-date_debut,'s')

```

Étape 1 : créer un tableau de données aléatoires sous Python.

Étape 2 : saisir le code d'exécution du tri sous Python.

Étape 3 : évaluer le temps nécessaire pour trier.

Lien vers le programme Python : [programme Python](#).

Exercice 32

1. Exécutez le programme après l'avoir télécharger avec Pyzo.
2. Indiquez le rôle de l'instruction « `taille_tableau=250` » et celui de l'instruction « `rd.randrange(10)` ».
3. À l'aide des instructions « `print(T)` » et « `print(tri_bulle(T))` », affichez le tableau avant puis après le tri. Le tri est-il correct ?
4. Modifiez le nombre de valeurs à trier. Cette méthode fonctionne-t-elle toujours ?
5. Chargez la bibliothèque « `times A` ». À l'aide des instruction de l'étape 3, calculez le temps nécessaire pour effectuer le tri. Calculez l'évolution de ce temps si la table du tableau double de taille.

Séance 3 : les cartes de visites.

I Le fichier vCard. (Avec PC)

https://www.youtube.com/watch?v=b_F1rERg30w genre de vidéo pub en anglais à propos des vcard.

vCard (Visit Card soit Carte de visite) est un format standard ouvert d'échange de données personnelles.

Il existe trois versions de vCard : 2.1, 3.0 et 4.0.

L'un ou l'autre de ces formats est utilisé par le logiciel Carnet d'adresses d'Apple ou encore les logiciels de courrier électronique comme Mozilla Thunderbird ou Outlook Express. vCard est également utilisé par des appareils mobiles (téléphones mobiles, assistants personnels, GPS) pour échanger des éléments de carnet d'adresses, mais aussi par des logiciels de messagerie instantanée comme Jabber ou Skype.

Le fichier vCard est un format de carnet d'adresses qui peut contenir une ou plusieurs adresses. Lorsqu'il n'en contient qu'une seule, c'est une carte de visite.

Exercice 33

Allez dans les contacts de votre téléphone, choisissez l'un d'entre eux et faites « Partager contact ». Choisissez de partager en envoyant par mail sur votre propre boîte mail. Exemple de procédure sur Iphone semblable à celle pour Android.

Ouvrez le fichier ainsi obtenu avec un éditeur de texte (Blocnotes, Notepad, Textedit, etc).

Pour exporter les contacts de son smartphone, il faut : -Rechercher, depuis l'application que vous utilisez pour vos contacts, dans les paramètres, la fonction exportation. La majorité des téléphones doit proposer cette possibilité : sous Android, il existe Exporter vers la carte SD, ou Exporter vers la carte SIM,... - Lancer cette fonctionnalité, le téléphone va générer un/des fichiers texte (vCard) de vos contacts - Naviguer dans les répertoires du téléphone pour trouver le fichier vCard, avec l'extension .vcf; par exemple : Contacts.vcf ou bien Contacts2.vcf si vous avez lancé deux fois la fonction! - Valider la date et l'heure de création du fichier - Connecter le mobile, votre téléphone, en USB, à votre ordinateur - Naviguer dans le répertoire du téléphone depuis l'ordinateur pour trouver le fichier Quand le téléphone est branché au PC : Mémoire de stockage interne/Système/PIM - Copier le fichier Contacts.vcf sur le disque dur de votre ordinateur - Démontez le périphérique USB, le mobile, et déconnectez le câble USB - Tester l'ouverture de votre fichier Contact.vcf , avec un éditeur de texte (Notepad++ ou Bloc Note). Si le fichier s'ouvre et affiche les contacts au format texte tout va bien (même si vous trouvez le tout un peu compliqué), si un message d'erreur s'affiche, il faut reprendre la procédure (le fichier est mal copié!). - Penser à sauvegarder toutes vos données!

Un fichier vCard porte une extension *.vcf* (acronyme de *Virtual Card File*). Son contenu est lisible par n'importe quel éditeur de texte (Bloc Note, Notepad++, ...) Il est utile pour conserver ou transmettre ses coordonnées à une relation.

Exercice 34

1. L'exemple suivant est un fichier avec une seule carte de visite

```
BEGIN:VCARD
VERSION:2.1
N:Levasasseur;Frederic;;Lycee;
NICKNAME:levav
FN:Lycee Levasasseur Frederic
TEL;CELL:069-221-0626
TEL;HOME:026-221-0626
EMAIL;HOME:secretariat@levasseur.re
ADR;HOME:;;4 rue Fenelon;Saint-Denis;Reunion;France;97400
URL:lycee-levasseur.re
BDAY:1911-02-25
END:VCARD
```

D'après vous quel est le rôle des différents descripteurs d'une vCard : N, NICKNAME FN, BDAY, ADR, TELL;CELL, TELL;HOME, EMAIL, BEGIN :VCARD et enfin END :VCARD ?

2. Rédigez votre propre vCard (fictive si vous le souhaitez).

Importer vcf sur android via gmail : https://www.frandroid.com/comment-faire/latelephonie/231193_importerlescontactsdepuisunfichiercsvouvcard
Ou créer/utiliser un site de partage pour mettre à jour ses vCard

II Un programme Python pour créer des vCard. (Avec PC)

Exercice 35

Programme final sans interface graphique.

Ajout d'une interface graphique : https://sebsauvage.net/python/gui/index_fr.html, un exemple dans le dossier à virer ensuite tktest.py

1. Téléchargez ([ici](#)) le programme Python puis exécutez-le avec Pyzzo. Quel est son rôle ?
2. Vous trouverez des explications sur les différents descripteurs de vCard en suivant ces liens : <https://fr.wikipedia.org/wiki/VCard> et <https://www.w3.org/TR/vcard-rdf/>.

Complétez le précédent programme Python de sorte que la vCard créée contienne aussi

- (a) la date d'anniversaire,
 - (b) le numéro de téléphone fixe,
 - (c) l'email,
 - (d) une adresse avec département, code postal et pays.
3. On se propose de créer une interface graphique au programme précédent.

Liste de commandes python3 https://www.tutorialspoint.com/python3/python_gui_programming.htm.

Tutorial complet à proposer aux élèves qui veulent personnaliser leur programme : <https://likegeeks.com/python-gui-examples-tkinter-tutorial/>

Attention en Python3 le module a changé de nom il est passé de Tkinter à tkinter.

Téléchargez ([ici](#)) le programme Python puis exécutez-le avec Pyzzo.

Complétez ce programme avec les mêmes descripteurs vCard qu'à la question précédente.

Logiciel de création de exe à partir du tkinter https://anthony-tuininga.github.io/cx_Freeze/

Correction exercice 35

- 1.
2. Programme complet.
3. Programme complet avec interface graphique.

III Un programme pour générer un carnet d'adresses. (Avec PC)

<https://codes-sources.commentcamarche.net/source/49068-concatener-des-f>
vielle version de Python.

Gestion des dossier et fichier avec Python3 : <https://python.doctor/page-gestion-f>

https://www.tutorialspoint.com/python3/tk_messagebox.htm

Un truc à tester :

```
import glob
import os
for f in glob.glob("file*.txt"):
os.system("cat "+f+" >> OutFile.txt")
```

Une qui marche bien et ne nécessite pas le module os. Le module glob semble fonctionner sur tous les os car préimplémenté dans python.

```
from glob import glob

with open('singleDataFile.txt', 'a') as singleFile:
for txtFile in glob('*.txt'):
for line in open(txtFile, 'r'):
singleFile.write(line)
singleFile.write("\n")
```

Une version sans le with :

```
from glob import glob

fichierouvert=open('singleDataFile.txt', 'a')
for txtFile in glob('*.txt'):
for line in open(txtFile, 'r'):
fichierouvert.write(line)
fichierouvert.write("\n")
fichierouvert.close()
```

Exercice 36

1. Créez plusieurs cartes de visites (avec un programme précédemment créé) que vous placerez dans un même répertoire.
2. Créez dans le même répertoire le programme Python suivant

```
from glob import glob

fichierouvert=open('carnet_d_adresses.vcf','a')
for txtFile in glob('*.vcf'):
    for line in open(txtFile, 'r'):
        fichierouvert.write(line)
        fichierouvert.write("\n")
fichierouvert.close()
```

puis exécutez-le.

3. Quel fichier a été créé dans votre répertoire? Que contient-il?
4. En quoi ce programme facilite-t-il l'export et le partage?

IV Utiliser Python pour convertir un csv en vCard (et réciproquement). (Avec PC)

Il s'agit d'utiliser python pour transformer un fichier csv avec la informations dessus en fichier vcard.

Le fichier csv aurait été rempli auparavant avec des données.

<https://codereview.stackexchange.com/questions/3517/convertng-a-csv-to-programme-Python>.

V Activité : métadonnées et synchronisation.

Trouver les métadonnées sur un fichier vcard.

VI Activité : manipuler un fichier .vcf avec python.

Programme python qui génère une carte de visite.

Programme comptant le nombre de contact.

Ranger dans l'ordre alphabétique les contacts.

Fusionner deux fichiers vcf (messagerie + téléphone par exemple).

VII Créer un exe à partir du Python.

Logiciel de création de exe à partir du tkinter https://anthony-tuininga.github.io/cx_Freeze/

Séance 4 : stocker les données.

I Stocker les données.

Lisez l'article <https://www.lebigdata.fr/stockage-donnees>, visionnez cet historique du stockage jusqu'en 1995 https://www.youtube.com/watch?v=50z_iSRw5XM (10minutes) et après <https://www.youtube.com/watch?v=oi4vM-WLtqY> (10 minutes) puis répondez aux questions suivantes.

Vidéos très longues, très complètes parfois un peu complexe.

Une autre vidéo plus élémentaire sur l'ordinateur de C'est pas sorcier : <https://www.youtube.com/watch?v=c96KP5jZVYk>.

1. Citez trois support de données d'avant 1995.
2. Quelle est l'unité de mesure de quantité de donnée la plus petite est à quoi correspond-elle ?
3. Donnez quatre autres unités de mesures de quantité de données informatiques d'usage courant.
4. Par combien est multipliée la quantité de données produite pour un individu chaque année ?
5. Qu'est-ce qu'un NAS ?
6. Quels supports de stockage récemment développés n'ont pas encore supplantés les HDD pour des raisons de coût ?

II Le cloud.

Il faudrait ajouter l'utilisation d'un stockage en ligne pour partager les vCard avec les smartphones.

- Intérêt du stockage dans le nuage : après avoir pris connaissance de la « Définition – Qu'est-ce que le stockage cloud ? » sur <https://www.lebigdata.fr/comparatif-cloud-gratuit-stockage-en-ligne>, donner deux avantages de ce type de stockage par rapport à un stockage sur son ordinateur personnel ?
- Selon le site <https://www.lebigdata.fr/comparatif-cloud-gratuit-stockage-en-ligne> sont les huit critères à prendre en compte pour choisir un cloud gratuit ? Pour chacun des critères, donner un exemple illustrant son importance en répondant dans le tableau suivant.

Critère	Intitulé du critère	Exemple pour illustrer son importance	Note
1			
2			
3			
4			
5			
6			
7			
8			

- Affecter une note à chaque critère :
 - 0 : aucune importance ;
 - 1 : peu important ;
 - 3 : assez important ;
 - 5 : très important ;
 - 8 : extrêmement important (indispensable)
- Choisir 5 fournisseurs de Cloud parmi ceux présentés sur <https://www.lebigdata.fr/comparatif-cloud-gratuit-stockage-en-ligne>, vérifier quels critères ils satisfont ou non, puis calculer la note globale en fonction de celle attribuée précédemment par critère et en déduire un classement en complétant le tableau ci-dessous.

Fournisseur	Critères								Note globale	Classement
	Sécurité									

III Les data centers et leurs impacts sur l'environnement. (Sans PC)

Les centres de données (en anglais *data centers*) regroupent dans un même lieu, un grand nombre de serveurs informatiques avec des capacités de stockage en mémoire qui se comptent en penta-octets (10^{15} octets ou 1 million de Go).

Ces centres consomment beaucoup d'énergie électrique pour alimenter les processeurs et les mémoires. Ils en consomment aussi beaucoup pour les systèmes de refroidissement nécessaires, pour dissiper la chaleur dégagée par le fonctionnement des processeurs.

Une analyse du gâchis énergétique des centres de données



Ordinateurs, *data centers*, réseaux... engloutissent près de 10 % de la consommation mondiale d'électricité. Et ce chiffre ne cesse d'augmenter. S'il n'est évidemment pas question de se passer des progrès apportés par le numérique, les scientifiques pointent un mode de fonctionnement peu optimisé et très énergivore.

Nous vivons dans un monde de plus en plus dématérialisé. Nous payons nos impôts en ligne, regardons nos séries préférées en streaming, stockons nos milliers de photos dans le cloud... Dématérialisé, vraiment ? « Si l'on considère la totalité de son cycle de vie, le simple envoi d'un mail d'1 mégaoctet (1 Mo) équivaut à l'utilisation d'une ampoule de 60 watts pendant 25 minutes, soit l'équivalent de 20 grammes de CO₂ émis », rappelle Françoise Berthoud, informaticienne au Gricad et fondatrice en 2006 du groupement de services EcoInfo pour une informatique plus respectueuse de l'environnement. Car les mots des nouvelles technologies sont trompeurs : ils évoquent l'immatériel comme le mot « virtuel », l'éthéré comme le mot « cloud », ou encore la pureté comme

l'expression de « salle blanche ». Et nous font oublier un peu vite les millions d'ordinateurs et de smartphones, les milliers de *data centers* et de kilomètres de réseaux utilisés pour traiter et acheminer ces données. Et la quantité considérable d'énergie qu'ils engloutissent. « Le secteur des nouvelles technologies représente à lui seul entre 6 et 10 % de la consommation mondiale d'électricité, selon les estimations soit près de 4 % de nos émissions de gaz à effet de serre, assène Françoise Berthoud. Et la tendance est franchement à la hausse, à raison de 5 à 7 % d'augmentation tous les ans [...].

Source : Laure Cailloce, CNRS Le journal, 16 mai 2018.

<https://lejournald.cnrs.fr/articles/numerique-le-grand-gachis-energetique>



Carte des data center et des clouds : <https://www.datacentermap.com/cloud.html>.

En procédant à des recherches sur le web apportez des éléments de réponse aux questions suivantes.

Pistes de réflexion

1. Les données stockées dans les centres de données ont-elles besoin d'être disponibles en permanence ?
2. Combien d'exemplaires identiques de la même donnée peuvent être stockés dans un centre de données ?
3. Quelles matières premières sont consommées par un centre de données ?

Sujets d'exposés ou de débats



1. L'évolution de la consommation électrique des centres de données dans le monde.
2. La localisation des centres de données dans le monde et le climat.
3. Les centres de données les plus énergivores.