



COURS PI

☆ *L'école sur-mesure* ☆

de la Maternelle au Bac, Établissement d'enseignement
privé à distance, déclaré auprès du Rectorat de Paris

Seconde - Module 2 - Traitement des données

Sciences Numériques et Technologie

v.5.1



- ✓ **Guide de méthodologie**
pour appréhender notre pédagogie
- ✓ **Leçons détaillées**
pour apprendre les notions en jeu
- ✓ **Exemples et illustrations**
pour comprendre par soi-même
- ✓ **Prolongement numérique**
pour être acteur et aller + loin
- ✓ **Exercices d'application**
pour s'entraîner encore et encore
- ✓ **Corrigés des exercices**
pour vérifier ses acquis

www.cours-pi.com

Paris & Montpellier



EN ROUTE VERS LE BACCALAURÉAT

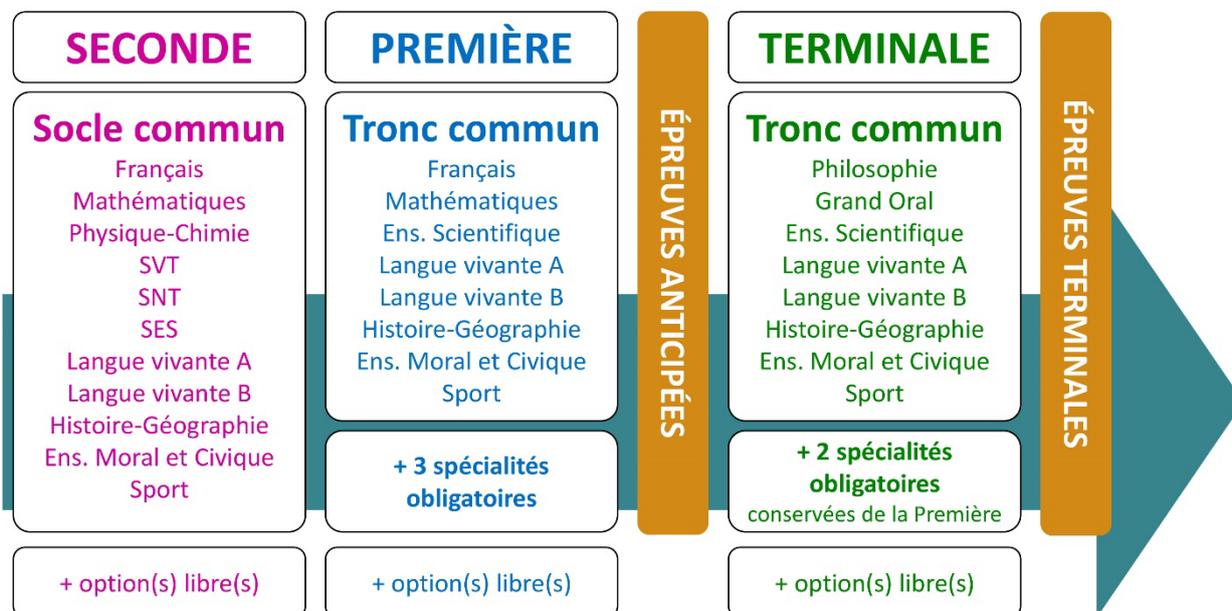
Comme vous le savez, la **réforme du Baccalauréat** est entrée en vigueur progressivement jusqu'à l'année 2021, date de délivrance des premiers diplômes de la nouvelle formule.

Dans le cadre de ce nouveau Baccalauréat, **notre Etablissement**, toujours attentif aux conséquences des réformes pour les élèves, s'est emparé de la question avec force **énergie** et **conviction** pendant plusieurs mois, animé par le souci constant de la réussite de nos lycéens dans leurs apprentissages d'une part, et par la **pérennité** de leur parcours d'autre part. Notre Etablissement a questionné la réforme, mobilisé l'ensemble de son atelier pédagogique, et déployé tout **son savoir-faire** afin de vous proposer un enseignement tourné continuellement vers l'**excellence**, ainsi qu'une scolarité tournée vers la **réussite**.

- Les **Cours Pi** s'engagent pour faire du parcours de chacun de ses élèves un **tremplin vers l'avenir**.
- Les **Cours Pi** s'engagent pour ne pas faire de ce nouveau Bac un diplôme au rabais.
- Les **Cours Pi** vous offrent **écoute** et **conseil** pour coconstruire une **scolarité sur-mesure**.

LE BAC DANS LES GRANDES LIGNES

Ce nouveau Lycée, c'est un enseignement à la carte organisé à partir d'un large tronc commun en classe de Seconde et évoluant vers un parcours des plus spécialisés année après année.



CE QUI A CHANGÉ

- Il n'y a plus de séries à proprement parler.
- Les élèves choisissent des spécialités : trois disciplines en classe de Première ; puis n'en conservent que deux en Terminale.
- Une nouvelle épreuve en fin de Terminale : le Grand Oral.
- Pour les lycéens en présentiel l'examen est un mix de contrôle continu et d'examen final laissant envisager un diplôme à plusieurs vitesses.
- Pour nos élèves, qui passeront les épreuves sur table, le Baccalauréat conserve sa valeur.

CE QUI N'A PAS CHANGÉ

- Le Bac reste un examen accessible aux candidats libres avec examen final.
- Le système actuel de mentions est maintenu.
- Les épreuves anticipées de français, écrit et oral, tout comme celle de spécialité abandonnée se dérouleront comme aujourd'hui en fin de Première.



A l'occasion de la réforme du Lycée, nos manuels ont été retravaillés dans notre atelier pédagogique pour un accompagnement optimal à la compréhension. Sur la base des programmes officiels, nous avons choisi de créer de nombreuses rubriques :

- **Suggestions de lecture** pour s'ouvrir à la découverte de livres de choix sur la matière ou le sujet.
- **À vous de jouer** pour mettre en pratique le raisonnement vu dans le cours et s'accaparer les ressorts de l'analyse, de la logique, de l'argumentation, et de la justification.
- **Réfléchissons ensemble** pour guider l'élève dans la réflexion.
- **L'essentiel** et **Le temps du bilan** pour souligner les points de cours à mémoriser au cours de l'année.
- **Pour aller plus loin** pour visionner des sites ou des documentaires ludiques de qualité.
- Et enfin... la rubrique **Les Clés du Bac by Cours Pi** qui vise à vous donner, et ce dès la seconde, toutes les cartes pour réussir votre examen : notions essentielles, méthodologie pas à pas, exercices types et fiches étape de résolution !

SCIENCES NUMÉRIQUES ET TECHNOLOGIE SECONDE

Module 2 – Traitement des données

L'AUTEUR



Adrien SAURAT

« L'enseignement, c'est favoriser l'autonomie et l'enrichissement des élèves, avec **autres objectifs, apprendre un métier.** » Professeur et formateur en informatique avec plus de douze ans d'expérience en développement web et dans l'animation du réseau Canopé, il se passionne aussi pour le théâtre et l'écriture de nouvelles. Des passions qui l'ont déjà conduit sur les planches du Festival d'Avignon.

PRÉSENTATION

Ce **cours** est divisé en chapitres, chacun comprenant :

- Le **cours**, conforme aux programmes de l'Education Nationale
- Des **applications** dont les **corrigés** se trouvent en **fin de chapitre**
- Des **exercices d'entraînement** et leurs **corrigés** en **fin de fascicule**
- Des **devoirs** soumis à correction (et **se trouvant hors manuel**). Votre professeur vous renverra le corrigé-type de chaque devoir après correction de ce dernier.

Pour une manipulation plus facile, les corrigés-types des exercices d'application et d'entraînement sont regroupés en fin de manuel.

CONSEILS A L'ÉLÈVE

Vous disposez d'un support decours complet : **prenez le temps** de bien le lire, de le comprendre mais surtout de **l'assimiler**. Vous disposez pour cela d'exemples donnés dans le cours et d'exercices types corrigés. Vous pouvez rester un peu plus longtemps sur une unité mais travaillez régulièrement.

LES DEVOIRS

Les devoirs constituent le moyen d'évaluer l'acquisition de **vos savoirs** (« Ai-je assimilé les notions correspondantes ? ») et de **vos savoir-faire** (« Est-ce que je sais expliquer, justifier, conclure ? »).

Placés à des endroits clés des apprentissages, ils permettent la vérification de la bonne assimilation des enseignements.

Aux *Cours Pi*, vous serez accompagnés par un **professeur selon chaque matière** tout au long de votre année d'étude. Référez-vous à votre « Carnet de Route » pour l'identifier et découvrir son parcours.

Avant de vous lancer dans un devoir, assurez-vous d'avoir **bien compris les consignes**.

Si vous repérez des difficultés lors de sa réalisation, n'hésitez pas à le mettre de côté et à revenir sur les leçons posant problème. **Le devoir n'est pas un examen**, il a pour objectif de s'assurer que, même quelques jours ou semaines après son étude, une notion est toujours comprise.

Aux Cours Pi, chaque élève travaille à son rythme, parce que chaque élève est différent et que ce mode d'enseignement permet le « sur-mesure ».

Nous vous engageons à respecter le moment indiqué pour faire les devoirs. Vous les identifierez par le bandeau suivant :



Vous pouvez maintenant
faire et envoyer le **devoir n°1**



Il est **important de tenir compte des remarques, appréciations et conseils du professeur-correcteur**. Pour cela, il est **très important d'envoyer les devoirs au fur et à mesure** et non groupés. **C'est ainsi que vous progresserez !**

Donc, dès qu'un devoir est rédigé, envoyez-le aux *Cours Pi* par le biais que vous avez choisi :

- 1) Par **soumission en ligne** via votre espace personnel sur **PoulPi**, pour un envoi **gratuit, sécurisé** et plus **rapide**.
- 2) Par **voie postale** à *Cours Pi*, 9 rue Rebuffy, 34 000 Montpellier
*Vous prendrez alors soin de joindre une **grande enveloppe libellée à vos nom et adresse**, et **affranchie au tarif en vigueur** pour qu'il vous soit retourné par votre professeur*

N.B. : *quel que soit le mode d'envoi choisi, vous veillerez à **toujours joindre l'énoncé du devoir** ; plusieurs énoncés étant disponibles pour le même devoir.*

N.B. : *si vous avez opté pour un envoi par voie postale et que vous avez à disposition un scanner, nous vous engageons à conserver une copie numérique du devoir envoyé. Les pertes de courrier par la Poste française sont très rares, mais sont toujours source de grand mécontentement pour l'élève voulant constater les fruits de son travail.*

VOTRE RESPONSABLE PÉDAGOGIQUE

Professeur des écoles, professeur de français, professeur de maths, professeur de langues : notre Direction Pédagogique est constituée de spécialistes capables de dissiper toute incompréhension.

Au-delà de cet accompagnement ponctuel, notre Etablissement a positionné ses Responsables pédagogiques comme des « super profs » capables de co-construire avec vous une scolarité sur-mesure.

En somme, le Responsable pédagogique est votre premier point de contact identifié, à même de vous guider et de répondre à vos différents questionnements.

Votre Responsable pédagogique est la personne en charge du suivi de la scolarité des élèves.

Il est tout naturellement votre premier référent : une question, un doute, une incompréhension ? Votre Responsable pédagogique est là pour vous écouter et vous orienter. Autant que nécessaire et sans aucun surcoût.

QUAND
PUIS-JE
LE
JOINDRE ?

Du **lundi** au **vendredi** : horaires disponibles sur votre carnet de route et sur PoulPi.

QUEL
EST
SON
RÔLE ?

Orienter les parents et les élèves.

Proposer la mise en place d'un accompagnement individualisé de l'élève.

Faire évoluer les outils pédagogiques.

Encadrer et **coordonner** les différents professeurs.

VOS PROFESSEURS CORRECTEURS

Notre Etablissement a choisi de s'entourer de professeurs diplômés et expérimentés, parce qu'eux seuls ont une parfaite connaissance de ce qu'est un élève et parce qu'eux seuls maîtrisent les attendus de leur discipline. En lien direct avec votre Responsable pédagogique, ils prendront en compte les spécificités de l'élève dans leur correction. Volontairement bienveillants, leur correction sera néanmoins juste, pour mieux progresser.

QUAND
PUIS-JE
LE
JOINDRE ?

Une question sur sa correction ?

- faites un mail ou téléphonez à votre correcteur et demandez-lui d'être recontacté en lui laissant **un message avec votre nom, celui de votre enfant et votre numéro.**
- autrement pour une réponse en temps réel, appelez votre Responsable pédagogique.

LE BUREAU DE LA SCOLARITÉ

Placé sous la direction d'Elena COZZANI, le Bureau de la Scolarité vous orientera et vous guidera dans vos démarches administratives. En connaissance parfaite du fonctionnement de l'Etablissement, ces référents administratifs sauront solutionner vos problématiques et, au besoin, vous rediriger vers le bon interlocuteur.

QUAND
PUIS-JE
LE
JOINDRE ?

Du **lundi** au **vendredi** : horaires disponibles sur votre carnet de route et sur PoulPi.
04.67.34.03.00
scolarite@cours-pi.com



CHAPITRE 1. Les données et leurs structures..... 1

Q OBJECTIFS

- Découvrir comment sont décrites et paramétrées les données afin de pouvoir les interpréter.
- Comprendre la différence entre données et métadonnées, et la lecture de ces deux types.
- Découvrir l'utilisation de supports de stockages en ligne (dans le nuage, ou « cloud » en anglais).

Q COMPÉTENCES VISÉES

- Savoir définir une donnée personnelle.
- Identifier les principaux formats et représentations de données.
- Identifier un descripteur et distinguer la valeur d'une donnée de son descripteur.
- Retrouver les métadonnées d'un fichier personnel.
- Expérimenter l'utilisation d'un support de stockage dans le nuage (cloud).
- Avoir connaissance des enjeux énergétiques liés au stockage de données.

Première approche : le stockage des données	2
1. L'évolution du stockage des données	4
2. Les données et leurs formats	6
3. Métadonnées	12
4. Données dans le nuage	14
Le temps du bilan	18
Les Clés du Bac	19

CHAPITRE 2. Traitement des données 21

Q OBJECTIFS

- Approfondir l'utilisation d'un site de données ouvertes.
- Réaliser des opérations sur des tables (ou collections de données) avec un tableur.
- Installer et expérimenter Python.
- Réaliser des traitements de données simples avec Python.

Q COMPÉTENCES VISÉES

- Connaître des sites de données ouvertes.
- Savoir utiliser un site de données ouvertes (pour y sélectionner et récupérer des données).
- Réaliser des opérations de recherche, filtre, tri ou calcul sur une table.
- Avoir installé Python et exécuté des scripts simples.

Première approche : les tableurs	22
1. Historique des tableurs	25
2. Importation de données ouvertes	27
3. Opérations de base dans un tableur	31
4. Première utilisation de Python	37
5. Traitement de données en Python	40
Le temps du bilan	44
Les Clés du Bac	45

CHAPITRE 3. La photographie numérique et le traitement d'image 47

Q OBJECTIFS

- La différence entre les photosites du capteur et les pixels de l'image.
- Retrouver les métadonnées d'une photographie.
- Traiter une image par programme pour la transformer.
- Identifier les étapes de construction d'une photo numérique.

Q COMPÉTENCES VISÉES

- Distinguer photosites du capteur et pixels de l'image.
- Retrouver les métadonnées d'une photographie.
- Traiter une image par programme pour la transformer.
- Identifier les étapes de construction d'une photo numérique.

Première approche : image numérique	48
1. Matériel et données associées.....	51
2. Métadonnées EXIF	56
3. Traitement d'images par programmation.....	57
4. Aperçu des algorithmes embarqués	62
5. Impact de ces nouvelles technologies	63
Le temps du bilan	65
Les Clés du Bac	66

CORRIGÉS à vous de jouer et exercices 69





Une donnée est une valeur ou une information qu'on peut vouloir sauvegarder pour être retrouvée plus tard ou pour être traitée par un programme. Ainsi, les données constituent des éléments de base servant de matière première pour tout traitement numérique. Même si certaines informations peuvent être stockées en « mémoire vive » et ainsi être vidées lorsqu'on éteint une machine, il est important de disposer aussi de données stockées sur des supports plus persistants (disques durs, cartes mémoires, etc.).

Ces données peuvent être organisées et structurées de différentes façons, que nous verrons plus en détail durant ce chapitre, ainsi que durant le programme de première.

Q OBJECTIFS

- Découvrir comment sont décrites et paramétrées les données afin de pouvoir les interpréter.
- Comprendre la différence entre données et métadonnées, et la lecture de ces deux types.
- Découvrir l'utilisation de supports de stockages en ligne (dans le nuage, ou « cloud » en anglais).

Q COMPÉTENCES VISÉES

- Savoir définir une donnée personnelle.
- Identifier les principaux formats et représentations de données.
- Identifier un descripteur et distinguer la valeur d'une donnée de son descripteur.
- Retrouver les métadonnées d'un fichier personnel.
- Expérimenter l'utilisation d'un support de stockage dans le nuage (cloud).
- Avoir connaissance des enjeux énergétiques liés au stockage de données.

Q MATÉRIEL NÉCESSAIRE

- Un ordinateur avec système d'exploitation Windows 7 ou supérieur.



Première approche

Le stockage des données



Document 1



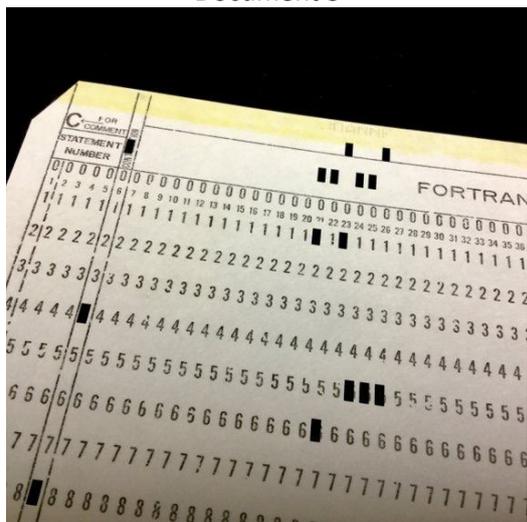
Document 2



Document 3



Document 4



Document 5



Document 6



LES DONNÉES ET LEURS STRUCTURES

L'évolution du stockage des données

Les premiers pas

Le plus ancien mode de stockage complexe lié à l'humanité (et au reste de la vie sur terre) n'est autre que l'ADN (ainsi que l'ARN) ! En effet, ce stockage biologique est très efficace. On commence même (au stade de recherche aujourd'hui) à utiliser des chaînes d'ADN synthétiques pour stocker des données informatiques.

Au niveau des créations humaines, on retrouve néanmoins dès la préhistoire l'usage de bâtons de comptage (ou bâtons de taille) qui ont beaucoup servi jusqu'au XIX^{ème} siècle. On peut grâce à ces bâtons faciliter le comptage instantané (une version plus évoluée du fait de compter sur ses doigts) ou bien « noter » des nombres à retenir.

Si l'on prend l'exemple d'un berger à l'ère du Néolithique : si enfin de journée il veut s'assurer que tous ses moutons rentrent dans l'enclos, il peut avoir recours à un tel bâton. La première fois qu'il fait sortir tout son troupeau, il fait passer les moutons un par un, et il grave une encoche pour chacun sur son bâton (qui peut être en bois ou en os). Son bâton de comptage contient alors une marque pour chaque mouton. En conséquence, lorsqu'il fait rentrer son troupeau dans l'enclos, il peut passer son doigt d'une encoche à l'autre à chaque fois qu'un mouton est passé. Il doit arriver à la dernière encoche en voyant passer le dernier mouton. S'il reste des encoches une fois le dernier mouton rentré, c'est que certains animaux se sont perdus !



Un des os d'Ishangon découverts au Congo et datés d'environ 20 000 ans.

Avec la sédentarisation et l'arrivée de l'agriculture, l'humanité a perfectionné ses outils de comptage et d'enregistrement. L'invention de l'écriture a ainsi surtout servi au début à noter des compte-rendus de transactions commerciales ou à rédiger des contrats. C'est en Mésopotamie, plus de 3000 ans avant notre ère, que la langue sumérienne a été la première à se voir transcrite sur un support physique, en l'occurrence des tablettes d'argile (l'écriture cunéiforme permettait d'écrire facilement sur l'argile à l'aide de calames, des roseaux taillés en pointe).



L'informatisation

L'écriture sous diverses formes s'est révélée bien pratique durant des millénaires, mais voici que depuis l'arrivée des ordinateurs, de nouveaux moyens de stockage devenus nécessaires. Ils sont de plus en plus performants.

Nous en avons vu certains lors de la première approche proposée en début de chapitre : cartes perforées, bandes ou disquettes magnétiques, disques optiques...

On peut en imaginer d'autres, comme les clés USB, ou les différentes cartes mémoire qui sont couramment utilisées dans les appareils photos, smartphones ou autres types de matériel transportable.

Tous ces supports de stockage ont pour but de stocker des informations numériques (sous format binaire, une suite de 0 et de 1), avec différents avantages en tête : gain de place, rapidité de lecture et/ou d'écriture, longue conservation, fiabilité, etc.

Nous avons ainsi surtout évoqué les supports amovibles, permettant d'effectuer des sauvegardes de sécurité, ou de transférer des informations d'une machine à une autre. Mais nos ordinateurs ont bien sûr besoin de stocker des données en leur sein. Pour cela, on utilise des disques appartenant généralement à deux grandes familles :

- les disques durs, ou Hard Drive Disks (HDD), inventés en 1956 et reposant sur l'utilisation de disques tournants magnétiques ;
- les disques SSD (solid-state drive), apparus en 1991, qui sont globalement plus robustes et rapides que leurs prédécesseurs (même si l'on peut trouver des avantages et inconvénients de chaque côté).



Intérieur d'un disque dur Seagate (taille : 3,5 pouces).



Disque SSD (taille : 2,5 pouces).



LES DONNÉES ET LEURS STRUCTURES

Les données et leurs formats

On parle de données pour tout ce qui est enregistré et restitué par une machine, il ne n'agit donc pas que de nombres ou de textes, mais aussi d'images, de vidéos, de sons, etc.

Il existe des données élémentaires, ne contenant d'une information : un nom, un prénom ou un numéro de téléphone par exemple. Et à un niveau supérieur, on trouve des données structurées. Ainsi, une ligne « client » dans une base de données commerciale pourrait contenir ces données élémentaires citées plus haut : nom, prénom et numéro de téléphone du client, mais aussi d'autres informations (adresse e-mail, adresse, etc.).

Ainsi, si une **donnée** est une information au niveau machine (stockée temporairement en mémoire vive ou durablement sur support disque), elle peut avoir différents niveaux de complexité.

On appelle **descripteur** ce qui décrit les informations décrites dans une donnée. Par exemple, si on travaille sur un fichier de clients, une donnée de base sera « un client ». Dans ce genre de cas, les libellés « nom », « prénom » ou « date de naissance » sont des descripteurs. Chaque descripteur de données est aussi généralement lié à un **type de données**. Pour reprendre le même exemple, le « nom » serait ici de type « chaîne de caractères » (avec peut-être une limite de 30 caractères), tandis que le descripteur « date de naissance » serait au format « date » (précisément de type DD/MM/YYYY pour une date affichant le jour, puis le mois, puis l'année, en chiffres).

L'exemple pris plus haut permet de mettre en évidence un cas particulier, celui des données personnelles. En effet, ce qu'on appelle **donnée personnelle** est ce qui permet d'identifier et décrire (dans les limites des besoins d'une application) une personne physique.

Enfin, pour conclure cette petite revue initiale de vocabulaire, évoquons les **collections de données** qui désignent des ensembles de données partageant les mêmes descripteurs. Par exemple, un ensemble de clients qui sont tous décrits par les mêmes champs (nom, prénom, adresse, etc.) peut être stocké dans une collection de données. Dans une base de données, on désigne généralement sous le nom de « table » ce genre de collection.



EN PASSANT, SUR LE WEB...

La protection des données

Avez-vous entendu parler du RGPD ? Cet acronyme signifie « Règlement général sur la protection des données » et désigne de nouvelles mesures prises depuis 2016 au niveau européen.

Question : un organisme français a pour but de veiller au bon respect des données, voyez-vous de quoi il s'agit ?

Voici le premier article du règlement RGPD, « Règlement général sur la protection des données ».

Article premier - Objet et objectifs

1. Le présent règlement établit des règles relatives à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et des règles relatives à la libre circulation de ces données.
2. Le présent règlement protège les libertés et droits fondamentaux des personnes physiques, et en particulier leur droit à la protection des données à caractère personnel.
3. La libre circulation des données à caractère personnel au sein de l'Union n'est ni limitée ni interdite pour des motifs liés à la protection des personnes physiques à l'égard du traitement des données à caractère personnel.

Comme on le voit, ce règlement a pour but d'aider les citoyens à rester maîtres des données les concernant. En vertu du RGPD, les sites internet disposant de données sur nous sont supposés nous les fournir sur simple demande. Ce genre d'opération est possible sur de gros sites comme Facebook ou auprès de Google. Avez-vous déjà essayé ?

En France c'est la **CNIL** qui a pour mission de faire respecter l'application du règlement RGPD ! Ces quatre lettres désignent la « Commission nationale de l'informatique et des libertés », qui a pour mission de veiller à ce que l'informatique soit au service du citoyen. Sur son site (cnil.fr), on peut trouver des informations sur le RGPD.

Examiner des données en libre accès

Certaines données sont mises à disposition du public. On parle alors d'**open data**. Ceci peut être fourni par des organismes privés (entreprises) ou publics (collectivités, services publics). Ces **données ouvertes** sont diffusées de manière structurée et sous une licence qui permet leur libre accès et leur réutilisation par toute personne, gratuitement.



UN PEU D'HISTOIRE

Politiques publiques

L'open data a plusieurs fondements, et parmi ceux-ci l'application de certaines politiques publiques.

Même si on peut en trouver certains fondements avant, c'est surtout au XXI^e siècle que ce mouvement a pu se doter de fondements juridiques. En effet, en 2003 une directive de l'Union Européenne, nommée « Informations du secteur public » a ouvert la voie de la réutilisation des données publiques. Un peu plus tard, toujours au sein de l'Europe, la directive INSPIRE encourage à partir de 2008 le recours à l'open data.

Aux États-Unis, sous le mandat de Barack Obama, l'ouverture des données publiques devient un axe important du gouvernement, avec notamment l'Open Government Initiative (adoptée en 2009).

Question : Sauriez-vous comment trouver les données publiques proposées par les pouvoirs publics français ?

Peut-être avez-vous eu l'occasion de remarquer que les sites gouvernementaux sont placés sous le nom de domaine « gov.fr », et les services précis occupent un sous-domaine, comme le site oups.gouv.fr qui se propose de réponse aux questions fréquentes à propos des démarches administratives.

Et bien si vous tapez dans votre moteur de recherche quelque chose comme « open data gov.fr » vous tomberez sur la « plateforme ouverte des données publiques françaises » située à l'adresse suivante : <https://www.data.gouv.fr>

Le site dispose d'une zone de recherche. Cherchez-y par exemple « Temps de parole des hommes et des femmes à la télévision et à la radio » et téléchargez un des fichiers proposés. Il s'agit de fichiers CSV, que vous pouvez ouvrir de différentes manières. Dans l'idéal, ouvrez cela d'abord avec un éditeur de texte, comme le bloc-notes (intégré à Windows), ou un logiciels gratuit comme Notepad++ ou Atom.

Par exemple, pour le fichier « Moyennes par années et par chaînes » (fichier csv de 46 kilo-octets) contient une première ligne qui ressemble à ceci :

media_type,channel_name,is_public_channel,year,women_expression_rate,speech_rate,nb_hours_analyzed

Il s'agit de la ligne d'en-tête, qui indique ce qui est contenu dans chaque colonne.

Les autres lignes ressemblent à ceci :

```
radio,Chérie FM,False,2002,47.10994424236209,15.738694355458852,718.0  
radio,Chérie FM,False,2003,46.03444471353921,16.250258185942855,1617.0  
radio,Chérie FM,False,2004,48.383607470997276,15.035449792785865,1644.0  
radio,Chérie FM,False,2005,45.45162674591405,16.063777716161454,1624.0  
radio,Chérie FM,False,2006,47.81930725989837,15.602842103982228,1604.0
```

Une ligne de ce type est constituée d'informations définies par les descripteurs qui sont dans l'en-tête. Chaque segment, comme « Chérie FM » ou « 47.10994424236209 », est une donnée (ou valeur). En outre, l'ensemble des lignes de données constitue une collection.

EXERCICE

01

Analyser les données récupérées.

A l'aide de la ligne d'en-tête, des lignes de données, mais aussi du texte explicatif accompagnant le jeu de données sur le site data.gouv.fr, analyser les informations contenues dans ce fichier. Pouvez-vous décrire chaque colonne ?

Les fichiers CSV



L'extension CSV signifie « comma-separated values » et désigne ces fichiers texte qui contiennent des données séparées par des virgules ou des points-virgules (parfois d'autres signes). La ligne d'en-tête contenant les descripteurs est fréquente mais pas obligatoire. Les lignes de données formant la collection ne sont pas limitées en nombre (dans la limite de l'espace disque disponible).

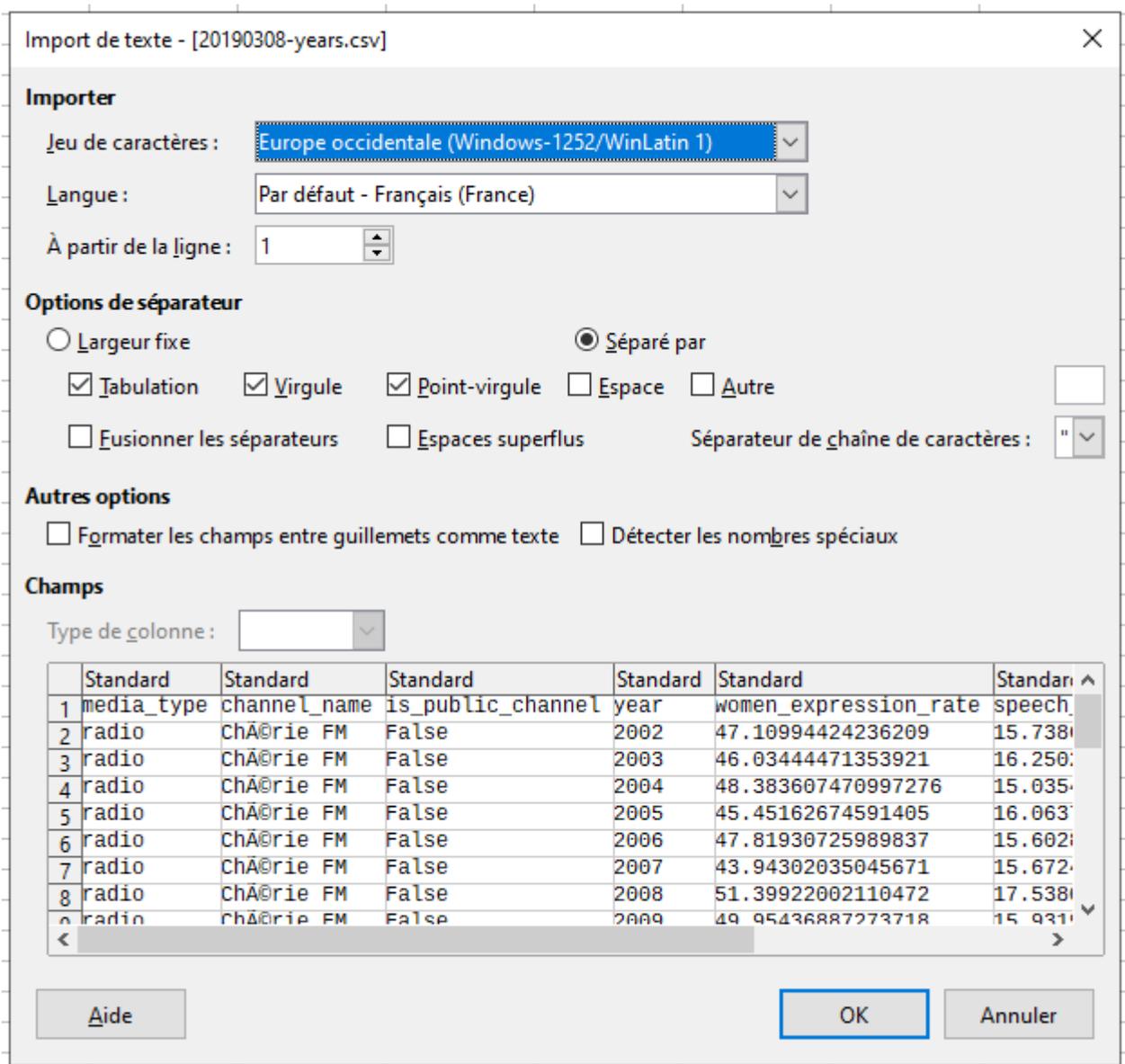
Nous avons vu qu'il était possible de les lire avec un éditeur de texte comme le bloc-notes, mais ils sont aussi particulièrement adaptés à une lecture via un tableur.

Question : connaissez-vous les logiciels désignés comme tableurs ? Quelle est leur utilité ?

Vous connaissez sans doute au moins Microsoft Excel, le tableur le plus connu aujourd'hui. Son équivalent chez Apple s'appelle Numbers. Ces logiciels permettent d'afficher, trier, traiter des données de manière pratique. Ils ont été créés à l'origine pour faciliter la gestion et la comptabilité des entreprises (et sont encore couramment utilisés pour ça), mais leur puissance a depuis multiplié leur usage.

Il existe une alternative libre et gratuite aux logiciels mentionnés plus haut, il s'agit du tableur Calc, contenu dans la suite Libre Office. Vous pouvez télécharger Libre Office, puis exécuter Libre Office Calc et ouvrir depuis ce logiciel le fichier CSV que vous avez téléchargé.

Vous devriez voir apparaître une fenêtre de ce type :



Dans ce cas précis, comme notre fichier est relativement standard (on constate juste des problèmes d’accents mais ce n’est pas bien grave), on voit dans la fenêtre de prévisualisation que les colonnes ont bien été repérées. Il suffit alors de cliquer sur OK pour que l’importation des données s’effectue. Vous verrez alors dans le tableur apparaître ces lignes :

	A	B	C	D	E	F	G
1	media_type	channel_name	is_public_channel	year	women_expression_rate	speech_rate	nb_hours_analyzed
2	radio	ChA@rie FM	False	2002	47.10994424236209	15.738694355458852	718.0
3	radio	ChA@rie FM	False	2003	46.03444471353921	16.250258185942855	1617.0
4	radio	ChA@rie FM	False	2004	48.383607470997276	15.035449792785865	1644.0
5	radio	ChA@rie FM	False	2005	45.45162674591405	16.063777716161454	1624.0
6	radio	ChA@rie FM	False	2006	47.81930725989837	15.602842103982228	1604.0
7	radio	ChA@rie FM	False	2007	43.94302035045671	15.672456051660472	1571.0
8	radio	ChA@rie FM	False	2008	51.39922002110472	17.538002781930835	1663.0
9	radio	ChA@rie FM	False	2009	49.95436887273718	15.931910989320668	1677.0
10	radio	ChA@rie FM	False	2010	48.06286688394447	15.948171474517942	1585.0
11	radio	ChA@rie FM	False	2011	47.83215331479781	17.981123963808614	1601.0

Vous pouvez dès lors utiliser les possibilités spécifiques d’un tableur : utiliser des formules, améliorer la forme (en rajoutant des couleurs, du gras, etc.), puis enregistrer tout ça dans un format de fichier qui stocke ces informations : pour Libre Office Calc il s’agit du format « Classeur ODF (*.ods) » mais ce logiciel est aussi capable d’enregistrer dans différents formats Excel, ou autres.

	A	B	C	D	E	F	G
1	<u>media_type</u>	<u>channel_name</u>	<u>is_public_channel</u>	<u>year</u>	<u>women_expression_rate</u>	<u>speech_rate</u>	<u>nb_hours_analyzed</u>
2	radio	Chérie FM	False	2002	47,11	15,74	718
3	radio	Chérie FM	False	2003	46,03	16,25	1617
4	radio	Chérie FM	False	2004	48,38	15,04	1644
5	radio	Chérie FM	False	2005	45,45	16,06	1624
6	radio	Chérie FM	False	2006	47,82	15,60	1604
7	radio	Chérie FM	False	2007	43,94	15,67	1571
8	radio	Chérie FM	False	2008	51,40	17,54	1663

Par exemple, dans l'image ci-dessus, les accents ont été corrigés (à l'aide de la fonction « Rechercher et remplacer »), l'en-tête a été colorée et centrée, les colonnes contenant les taux ont été converties en format français (décimale marquée par la virgule au lieu du point) et réduite à deux chiffres après la virgule. On peut enregistrer ce fichier modifié au format ODS ou XLS et le rouvrir plus tard : toutes ces modifications resteront visibles.

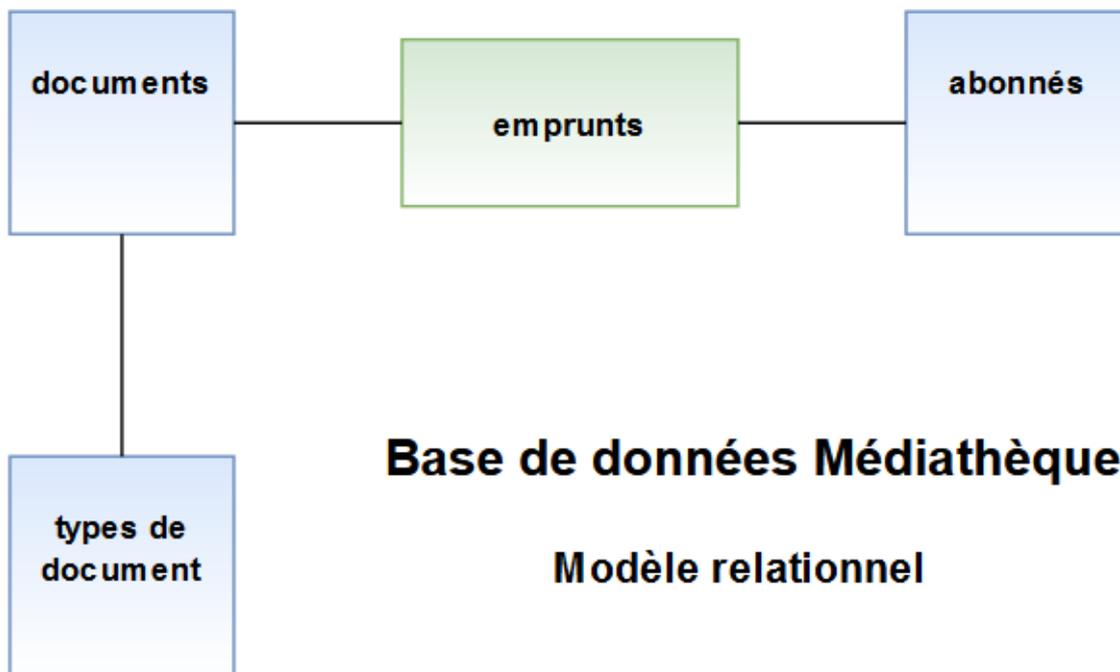
Si vous enregistrez cette version au format CSV, par contre, certains de ces changements seront perdus. Le CSV n'enregistre pas les couleurs, les mises en gras, etc.

Bases de données et modèle relationnel

Une **base de données** est un ensemble de collections (généralement appelées tables dans ces bases). Ce type d'organisation permet d'obtenir des structures de données complexes qui permettent de gérer toutes sortes de situations et de problématiques.

Par exemple, on peut imaginer que le fonctionnement d'une médiathèque passe par une base de données chargée d'enregistrer, entre autres, les informations relatives aux abonné·e·s et aux documents proposés par le lieu.

On représente les liens entre les collections d'une base de données à l'aide d'un modèle relationnel. Une version très simplifiée d'un modèle pourrait ressembler à ceci :



Les collections carrées, en bleu, contiennent les données sur lesquels reposent l'essentiel des traitements :

- un document est d'un certain type (livre, CD...);
- un document dispose dans sa collection de descripteurs comme son titre, mais aussi quelque chose comme « état de l'emprunt » qui permet de savoir si un document est disponible ou déjà prêté;
- un abonné est défini par des données personnelles (nom, prénom, date d'inscription, etc.).

Les traits représentent les relations entre les collections. Dans la collection « documents » un descripteur indique « type de document » et les valeurs correspondantes renvoient à un type de document décrit en détail dans la table qui porte ce nom.

La collection « emprunts » est un peu particulière car elle ne sert qu'à faire la relation entre les abonnés et les documents. Elle contient un identifiant pour chaque afin d'identifier qui a emprunté quoi. Pour chaque emprunt, elle contient en outre des informations complémentaires comme la date de retour prévue.

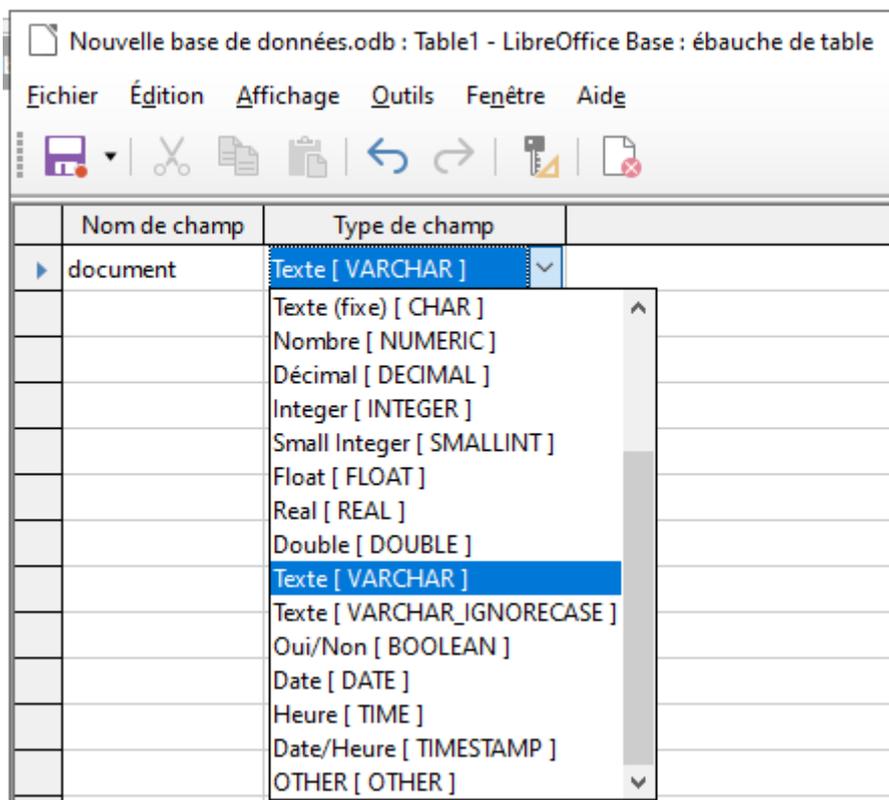
La base de données de Libre Office



Nous avons vu le tableur de Libre Office, Calc, qui est l'équivalent libre d'Excel. Et bien, si la suite Microsoft Office contient aussi une base de données nommée Access, cette dernière a aussi un équivalent chez Libre Office, nommé simplement Base. Si vous avez installé Libre Office, vous pouvez exécuter ce logiciel pour voir à quoi peut ressembler un outil de gestion de base de données.

Question : reconnaissez-vous dans Base certaines notions évoquées dans ce cours ?

Le fait d'apprendre à se servir d'une base de données dépasse le cadre de ce chapitre, mais vous pouvez néanmoins effectuer les premières opérations de la création d'une base. Après le choix du fichier de base de données (première étape obligatoire où il vous est demandé de choisir un espace de travail sur votre disque dur), vous aurez accès à un bouton nommé « Créer une table en mode Ebauche ». Cliquez dessus et saisissez un nom de champ, puis voyez quels types sont proposés :



Chaque ligne de ce tableau permet de définir un descripteur ! Si vous terminez la table, il vous sera ensuite possible de la remplir avec des valeurs. L'ensemble de la table (toutes les lignes de données) représentent une collection.

03

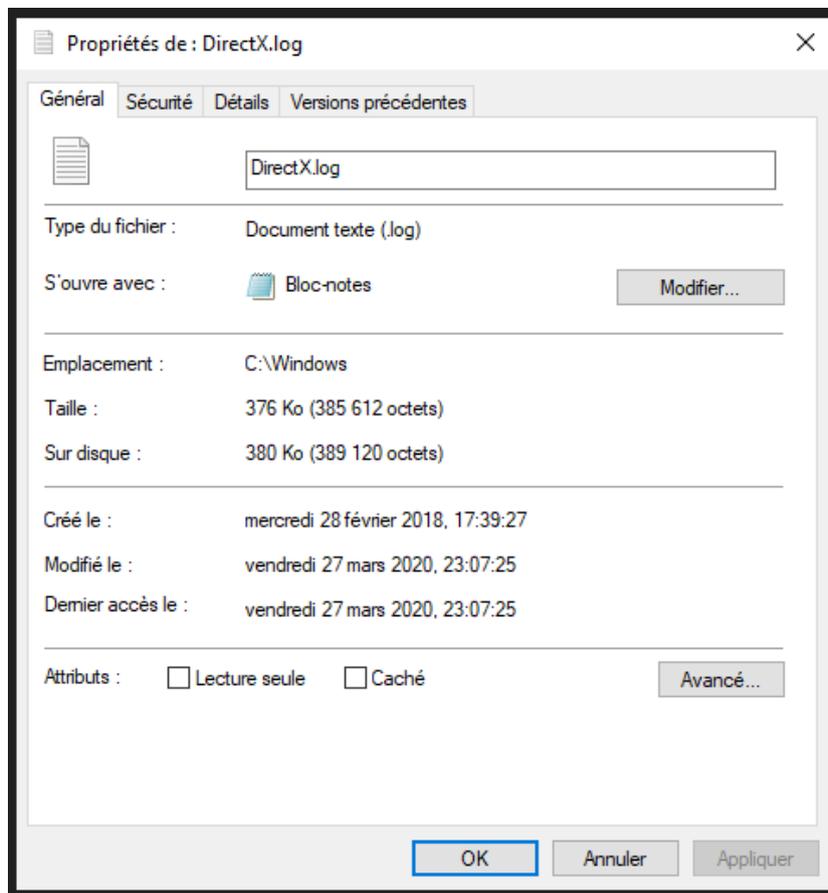
LES DONNÉES ET LEURS STRUCTURES

Métadonnées

Le terme méta (qui vient du grec meta, signifiant « au-delà de » ou « avec ») permet de désigner ce que l'on peut dégager « à propos » d'une chose. En l'occurrence, lorsque l'on parle de métadonnées, on évoque donc des informations relatives à des données. Oui, des données sur des données ! Mais à quoi cela peut-il servir ? Eh bien, à mieux identifier, classer, retrouver nos informations... Lorsque vous utilisez l'explorateur de fichiers de votre ordinateur, et qu'apparaissent des colonnes du type « nom », « modifié le », « type », « taille » ou ce genre de choses... Il s'agit là de données relatives à vos données ! Des informations qui complètent vos fichiers. Ainsi, l'extension d'un fichier est relative à son type : un fichier au format PDF aura une extension « .pdf » ce qui vous permet à la fois de l'ouvrir automatiquement avec le bon logiciel (Acrobat Reader ou Sumatra, par exemple), mais aussi de trier les fichiers d'un dossier par type afin de distinguer plus facilement les PDF du reste.

Métadonnées d'un fichier standard

Prenons l'exemple d'un fichier au hasard d'un système Windows. En cliquant dessus avec le bouton droit de la souris, on obtient ce qu'on appelle le menu contextuel (qui porte ce nom car ce qu'il affiche diffère suivant l'endroit d'où vous l'appellez). En bas de ce menu, vous avez à votre disposition le choix « Propriétés » qui peut donc afficher ce genre d'informations :



EXERCICE

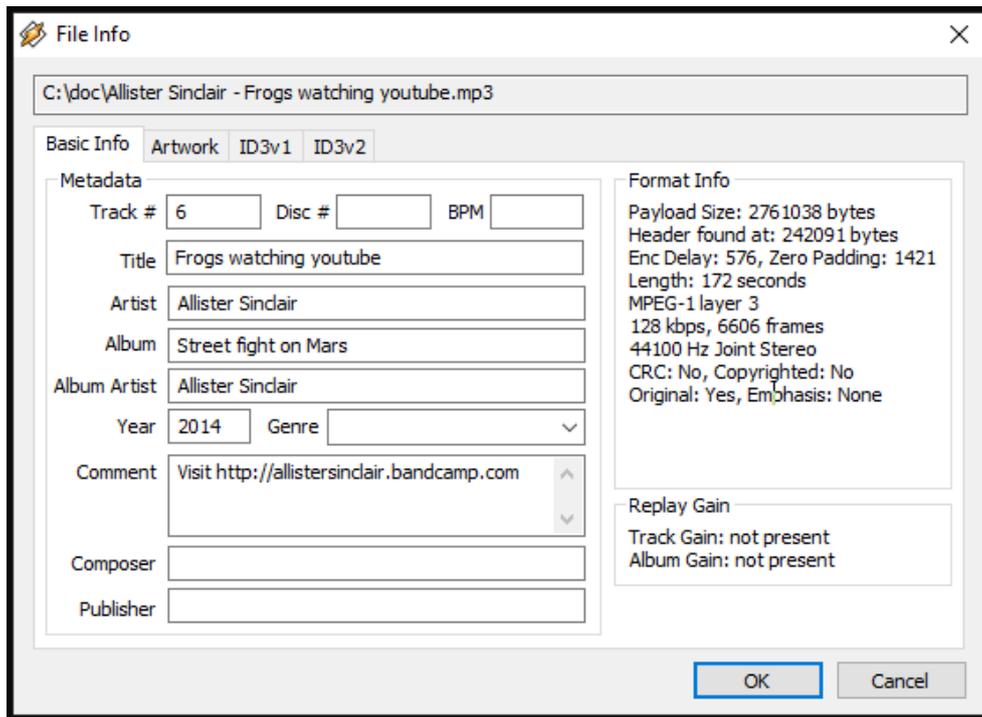
02 Exploration de métadonnées

Utilisez l'explorateur de fichiers Windows (que vous pouvez appeler avec le raccourci « touche Windows » + « touche E ») ainsi que le menu contextuel pour afficher les propriétés de certains de vos fichiers. Que contient l'onglet « Détails » ? Voyez-vous des différences entre fichiers textes et fichiers d'images (par exemple, d'autres comparaisons peuvent être intéressantes) ?

Métadonnées d'un mp3

Examinons un autre exemple de métadonnées, celle qui accompagnent un fichier son.

Prenons l'exemple d'un fichier mp3 ouvert avec le logiciel Winamp. Ce dernier (comme la plupart des autres lectures de musique) permet d'afficher les « informations » relatives à une piste. Cela donne par exemple ceci :



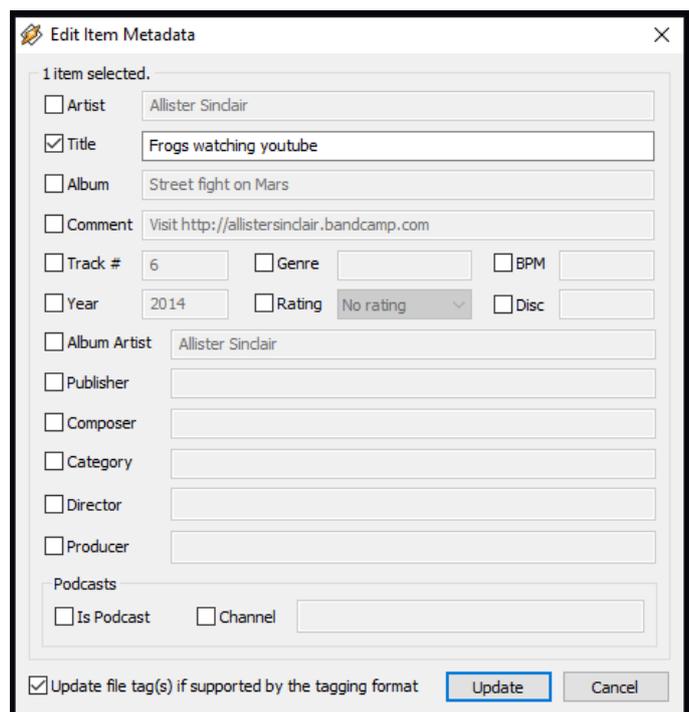
Si l'on avait ouvert les propriétés du fichier son avec le menu contextuel vu auparavant, on aurait vu les métadonnées classiques rattachées au système de fichiers de Windows : date de création ou de modification du fichier, etc. Ici, nous sommes dans un logiciel musical qui sait exploiter de manière pratique les métadonnées spécifiques aux fichiers sons. En effet, si les données principales d'un fichier mp3 sont les ondes sonores qui permettent de l'écouter, d'autres données sont disponibles : titre, artiste, album, année de sortie, genre, commentaires...

Modifier des métadonnées

Les métadonnées ne sont pas immuables ! Elles « vivent » avec le fichier qu'elles accompagnent, et évoluent suivant les modifications automatiques ou manuelles qu'il subit.

Par exemple, une information de « date de création » ou « date de modification » est gérée automatiquement par le système d'exploitation (même s'il est possible de les modifier manuellement sous certaines conditions). Par contre, les données accompagnent un fichier média, à savoir le nom de l'artiste ou la date de première publication, sont faites pour être mises à jour soit par des bases de données publiques (exemple : CDDb pour la musique) soit manuellement par le ou la propriétaire du fichier. Reprenons l'exemple du fichier pris plus haut. Winamp ne permet pas seulement de lire les métadonnées d'un fichier son (mp3 ou autre), il permet aussi de les modifier.

Cela se fait via cet écran :



Toujours avec l'explorateur de fichiers Windows et le menu contextuel, affichez les propriétés d'un fichier musical (wav, mp3, ogg, etc.).

Que contient l'onglet « Détails » ? Retrouvez-vous les informations proposées par Winamp (montré en exemple plus haut) ? Pouvez-vous les modifier ?



LES DONNÉES ET LEURS STRUCTURES

Données dans le nuage

Il existe différentes façons d'enregistrer des données à distance. Au sein de locaux d'entreprise, par exemple, on peut tout simplement disposer d'un serveur permettant de partager des dossiers depuis l'ensemble des postes clients du bâtiment. Même au sein d'un foyer, à vrai dire, il est de plus en plus courant de voir de mini-ordinateurs, des box internet ou des télévisions « intelligentes » disposer de disques durs internes destinés à recevoir les films des offres VOD (Video On Demand). Dans ce genre de cas, différentes machines dans le foyer peuvent aller chercher ces informations centralisées.

Il s'agit bien là d'un stockage distant, d'un stockage en réseau, mais le concept de cloud va encore plus loin.



Le « data center » ARSAT en Argentine.

Imaginé dans les années 60 et expérimenté par CompuServe et AT&T dans les années 80 puis 90, c'est à partir de 2005 que de grands noms du « cloud » (comme Amazon ou Dropbox) ont ouvert la voie à ce nouvel usage de plus en plus courant.

On parle désormais de cloud pour désigner les données stockées à distance (et nécessitant donc une connexion interne relativement rapide et fiable). Ceci dit, même si cela signifie que nos machines personnelles ont moins besoin de stockage physique, les données doivent toujours être enregistrées physiquement quelque part !

Cela se passe dans de grandes salles contenant des ordinateurs configurés en tant que « serveurs » (dans le sens où ils répondent à des requêtes effectuées par des ordinateurs « clients »). Ces serveurs disposent de vastes espaces de stockage. Pour les entreprises ayant besoin de stocker de très grandes quantités de données, l'enregistrement est souvent effectué sur des disques durs magnétiques (moins coûteux). Mais pour certains clouds misant sur la rapidité, des disques flashs (type SSD) peuvent être utilisés. Dans tous les cas, les serveurs de ce type « doublent » (de manière littérale ou suivant un radio adapté) souvent les disques par sécurité : les données sont enregistrées sur un disque ou plusieurs disques, et un disque est dédié en parallèle à la sauvegarde ou à la vérification d'erreurs (via des systèmes RAID, par exemple). L'idée étant ici de permettre la récupération des données au cas où l'un des disques tomberait en panne (ce qui ne manque pas d'arriver).

Les gigantesques salles serveurs nécessaire pour le maintien de ce genre d'architecture portent le nom, en anglais, de **data centers**. Ce genre d'infrastructure, de plus en plus courante, n'est pas sans conséquence pour l'environnement et la consommation énergétique (deux concepts étroitement liés). Notons que, d'après l'association française *The Shift Project* (un laboratoire d'idées ayant pour objectif l'atténuation du changement climatique), le numérique dans son ensemble était responsable en 2018 de 3,7 % des émissions mondiales en CO₂.



EN PASSANT, SUR LE WEB...

Le cloud vu par l'INRIA

Visionnez la vidéo « Qu'est-ce que le Cloud ? » disponible sur YouTube (chaîne « pixees Science participative »). On peut y entendre Serge Abiteboul, membre de l'INRIA (Institut National de Recherche en Informatique et en Automatique) expliquer les grands contours du cloud.

Question : Quels avantages et inconvénients du cloud avez-vous relevé ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

stockez chez eux. C'est le cas par exemple de Google : des scripts automatiques passent en revue toutes les informations que les utilisateurs confient à la firme d'une manière ou d'une autre. En 2017, des documents stockés sur Google Drive ont été bloqués parce qu'ils enfreignaient les conditions d'utilisation de Google, ce qui prouve que ce « cloud » ne se contente pas de stocker ce que vous y déposez sans en explorer le contenu. Il est très difficile de savoir si l'entreprise repère aussi des mots-clés récurrents dans ce que vous stockez afin de mieux déterminer vos goûts et revendre des informations de profil à des partenaires commerciaux. Quoiqu'il en soit, si vous estimez que les avantages d'un tel stockage distant sont supérieurs aux inconvénients réels ou supposés, vous pouvez vous inscrire auprès de ces services. Souvent, il est alors requis d'installer un logiciel sur votre machine, qui s'occupera d'effectuer les synchronisations.

Différents types de paramétrages sont envisageables. Parmi ceux-ci, on peut noter :

- le simple stockage à distance ; dans ce cas c'est comme un disque dur externe, qui n'est bien sûr accessible que lorsqu'on est connecté à Internet ;
- le stockage **synchronisé unidirectionnel** ; on a alors un espace sur notre disque local, quand on le modifie en ajoutant ou modifiant un fichier, c'est envoyé vers le cloud, qui sert alors soit de simple sauvegarde soit permet d'accéder à ces mêmes données depuis une autre machine ;
- le stockage **synchronisé bidirectionnel** ; même chose que précédemment, sauf que lorsque l'on modifie les données sur le cloud (toujours avec le logiciel fourni ou un accès web), notre espace local est impacté et mis à jour.



POUR ALLER PLUS LOIN

[Big Data : la révolution des mégadonnées—documentaire](#)

Avec la prolifération exceptionnelle des appareils numériques, une force invisible et inéluctable est en train de bouleverser la vie des humains. À l'heure actuelle, l'espèce humaine génère en deux jours seulement la même quantité de données qui a été générée entre son apparition sur Terre et l'an 2003.

L'accumulation massive et l'analyse de ces données en temps réel permettent d'affronter d'importantes problématiques, comme la pollution, la faim dans le monde et les maladies. Toutefois, Edward Snowden et le coulage de documents de la NSA (National Security Administration) ont démontré qu'il y a un prix à payer pour l'accessibilité de toutes ces données. Ce documentaire illustre de façon captivante les merveilles et dangers associés à la révolution des données.

[Disponible en DVD ou en replay sur internet](#)



Vous pouvez maintenant
faire et envoyer le **devoir n°1**

