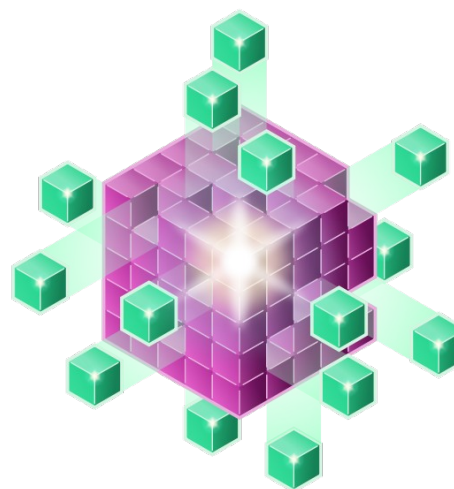
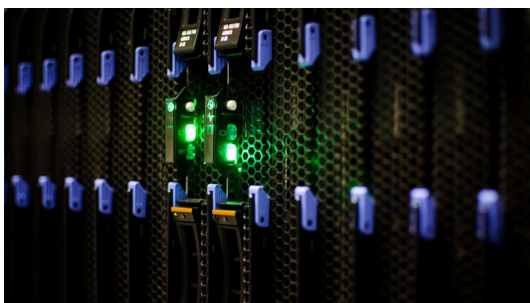




Bien stocker ses données



Petit sondage rapide : levez vos post'its !



- Où stockez-vous vos données de thèse ?



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
- **Où sauvegardez-vous vos données de thèse ?**



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
- **Où sauvegardez-vous vos données de thèse ?**
 - Sur mon ordinateur



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !
- Je n'ai pas de sauvegarde



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
- **Où sauvegardez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
 - Je n'ai pas de sauvegarde
- **A quelle fréquence sauvegardez-vous vos données de thèse ?**



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
- **Où sauvegardez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
 - Je n'ai pas de sauvegarde
- **A quelle fréquence sauvegardez-vous vos données de thèse ?**
 - Jamais



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !
- Je n'ai pas de sauvegarde

- **A quelle fréquence sauvegardez-vous vos données de thèse ?**

- Jamais
- De temps en temps



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
- **Où sauvegardez-vous vos données de thèse ?**
 - Sur mon ordinateur
 - Sur un DD externe
 - Sur un serveur du labo
 - Sur un cloud académique
 - Sur un cloud commercial (google, AWS, dropbox ...)
 - Ailleurs !
 - Je n'ai pas de sauvegarde
- **A quelle fréquence sauvegardez-vous vos données de thèse ?**
 - Jamais
 - De temps en temps
 - Toutes les semaines



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !
- Je n'ai pas de sauvegarde

- **A quelle fréquence sauvegardez-vous vos données de thèse ?**

- Jamais
- De temps en temps
- Toutes les semaines
- Tous les jours



Petit sondage rapide : levez vos post'its !



- **Où stockez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !

- **Où sauvegardez-vous vos données de thèse ?**

- Sur mon ordinateur
- Sur un DD externe
- Sur un serveur du labo
- Sur un cloud académique
- Sur un cloud commercial (google, AWS, dropbox ...)
- Ailleurs !
- Je n'ai pas de sauvegarde

- **A quelle fréquence sauvegardez-vous vos données de thèse ?**

- Jamais
- De temps en temps
- Toutes les semaines
- Tous les jours
- Plusieurs fois par jour (sauvegarde automatique par exemple)



- **Vos pratiques sont adaptées ??**



De quoi parle-t-on ? Stockage



- **Stockage** = enregistrement d'une information sur un support physique
- Ce support physique peut avoir des **caractéristiques très variées** :
 - En fonction du **matériel utilisé**
 - Par exemple, quand vous achetez un portable vous pouvez choisir un disque SSD ou SATA.
 - En fonction de la **technologie utilisée** pour accéder à ce support physique
 - Par exemple, vous ne pouvez pas ouvrir sous Windows un disque dur qui a été formaté sous Linux car les technologies utilisées pour organiser le stockage ne sont pas les mêmes.

De quoi parle-t-on ? Sauvegarde



- Rappel : **stockage** = enregistrement d'une information sur un support physique
- **Sauvegarde** = dupliquer des données pour les mettre en sécurité sur des **supports de stockage différents**
 - Recopie des données à l'**identique**
 - Sur des supports différents et localisés en général dans des **endroits différents**
 - ➔ Par exemple : je fais régulièrement des copies des fichiers de mon portable sur un DD externe
 - Objectif : pouvoir facilement **récupérer des données** en cas de perte ou de mauvaise manipulation
 - ➔ Attention : ne permet que de récupérer les données à la date de la dernière sauvegarde !

De quoi parle-t-on ? Archivage

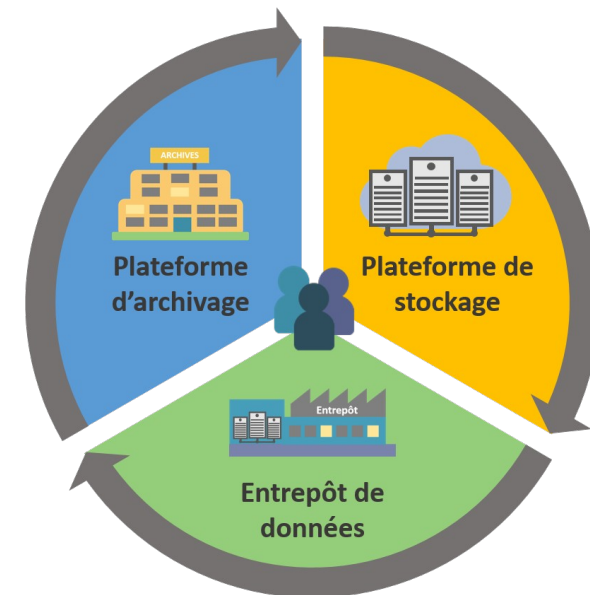


- **Archivage** = ensemble d'actions qui a pour but de garantir l'accessibilité sur le long terme d'informations (dossiers, documents, données) que l'on doit ou souhaite conserver pour des raisons juridiques, historiques ou culturelles. Il comprend à la fois des règles (procédures), des compétences et des infrastructures. (wikipedia)
- Dans notre cas : archivage à **long terme de données numériques**
- Donc **ce n'est pas seulement** du stockage sur une longue durée !
- Nécessité :
 - D'assurer la **pérennisation des supports** de stockage sur du long terme
 - D'assurer l'**accès au contenu** même quand les formats des données deviennent obsolètes
 - D'assurer l'**intégrité** des données
- Cadre **juridique propre**
- En France : **un seul opérateur** pour l'archivage des données de l'ESR : le CINES (Centre Informatique National de l'Enseignement Supérieur)

Plateformes de stockage, archivage, diffusion, entrepôts de données ...



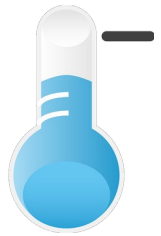
- Différentiées par les **usages**, et la/les **phases du cycle de vie** des données :
 - **Plateforme de stockage** : infrastructure proposant un stockage des données, avec des fonctionnalités de gestion des accès et éventuellement de sauvegarde intégrée
 - **Plateforme d'archivage** : infrastructure qui intègre dans son fonctionnement tout le processus nécessaire à l'archivage des données
 - **Entrepôt de données, plateforme de diffusion** : Réservoir de données de recherche, brutes ou dérivées, qui peuvent être retrouvées et réutilisées grâce à une description par des métadonnées. Un identifiant pérenne ou numéro d'accès est attribué à chaque jeu de données. Il peut être disciplinaire ou thématique, être institutionnel ou centralisé.



Données chaudes, données froides

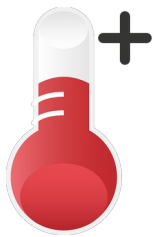


- **Données froides** : très peu utilisées et donc très peu accédées.



- Un carton de photos anciennes dans un grenier : on sait qu'il est là mais on ne va pas l'ouvrir tous les jours. Par contre, on a envie de pouvoir sortir une photo lors d'une occasion particulière
- En général, on peut se permettre des temps d'accès un peu long sur ces données
- Un DD sur lequel on a stocké des données d'un vieux projet

- **Données chaudes** : actives, accédées souvent voir de façon très intensive



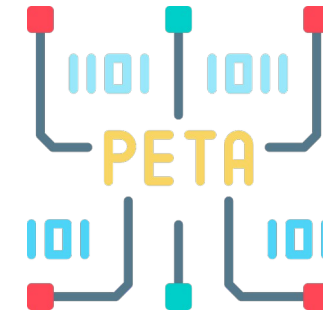
- Par exemple, des données issues d'une expérimentation en cours et que l'on doit analyser
- Dans ce cas, on souhaite que les traitements soient rapides et donc que les temps d'accès aux données soient performants
- Il existe des espaces de stockage spécifiques pour les traitements intensifs de données permettant de faire des calculs efficacement

Volumes de données



- **1 Po = 1 000 To ; 1 To = 1 000 Go ; 1 Go = 1 000 Mo**

- Capacité DD externe « moyen » : 2 To
- Plateforme de stockage de site : quelques Po
- Fichier texte < fichier bureautique ~ fichier audio ~ fichier image < fichier vidéo



- On peut commencer à parler de gros volumes à partir de **quelques To à quelques dizaines de To** : en fait à partir du moment où le temps de transfert ou le temps de chargement devient prohibitif

- Conséquences d'une **volume important** :

- Problématique du transfert des données
- Problématique du chargement en mémoire pour traitement / analyse / calcul
- Problématique de la capacité de stockage nécessaire et de son coût
- Problématique de la politique de sauvegarde ...





- **Le coût financier**

- Le stockage a un coût qu'il ne faut pas négliger
- Par sa nature, contrairement au calcul par exemple, le stockage n'est pas vraiment mutualisable
- On peut optimiser en partageant une plateforme et les coûts humains associés
- Plus un stockage est performant (on accède rapidement aux données), plus il est cher
- Exemple de coût :
 - DD externe 2 To : 35 à 70 € environ
 - 1 To sur la plateforme Summer : de 25 € à 80€ annuel en fonction du type de stockage
 - 1 To sur la plateforme Microsoft Azur (équivalent Summer sur technologie NetApp) : de 165 à 442 € environ par mois en fonction du type de stockage (cf <https://azure.microsoft.com/fr-fr/pricing/details/netapp/>)

- **Le coût humain**

- Le stockage personnel n'est pas scalable et peu sécurisé
- L'exploitation d'une plateforme de stockage à grande échelle, sécurisée et pérenne nécessite des compétences techniques de pointe qu'il est important de mutualiser

- **Le coût environnemental**

- Le stockage, et encore plus l'archivage a un coût environnemental non négligeable
- Exemple : pour la plateforme Bettik à l'UGA (stockage performant pour le calcul) : 1 Go.an émet 12 g CO2e - en 2019 : 1.4 Po.an → 17 t CO2e



- **Mes données sont sur un DD externe**
 - Je branche le DD et j'accède en direct sur mon portable → le système d'exploitation reconnaît et « monte » mon DD comme celui qui est interne
 - Je maîtrise complètement le stockage
 - Je ne peux pas partager mes données
 - Stockage non scalable, peu sécurisé
- **Mes données sont sur un cloud**
 - J'accède à mes données via un navigateur internet ou un client cloud que j'installe sur mon portable
 - Je peux aussi accéder à des données sur le cloud via une url qu'on m'a envoyée
 - Je ne sais pas où sont mes données la plupart du temps sauf si j'utilise un cloud académique. Il est aussi parfois difficile de savoir qui a les droits d'accès
 - Je peux partager mes données
 - Stockage peu scalable, sécurisé si plateforme académique
 - Exemple : cloud UGA, MyCore CNRS



- **Mes données sont sur une plateforme de stockage**
 - J'accède à mes données : soit via un montage réseau (VPN souvent nécessaire pour assurer la sécurité), soit via un protocole d'accès distant (comme scp par exemple)
 - La plateforme peut aussi proposer un accès par le web
 - Si j'utilise une plateforme académique, je sais où sont les données et qui les gèrent. J'ai la maîtrise de la gestion des droits d'accès mais cela peut nécessiter un savoir-faire technique. Je peux partager mes données.
 - Stockage hautement scalable et sécurisé
 - Exemple : plateforme Summer
- **Mes données sont sur une plateforme de stockage associée à une infrastructure de calcul académique**
 - J'accède à mes données à travers mon utilisation de l'infrastructure de calcul
 - Je sais où sont mes données et j'ai la maîtrise des accès. Je peux partager mes données.
 - Stockage scalable, non ou peu sécurisé
 - Exemple : Bettik, Silenus, Mantis, stockage des centres nationaux de calcul



Nos données sont des **éléments précieux** de notre vie professionnelle et personnelle.

- **Quelles sont vos données les plus critiques ?**
- **Quels événements redoutez-vous le plus ?**

Perte, destruction de données

Vol de données

Autre ?





Les questions à se poser :

- **Que se passe-t-il si je perds cette donnée / information / fichier / dossier ?**
 - Quelles conséquences, quels impacts, quels préjudices ?
 - Pour moi ? Pour mon équipe ? Mon unité ? Mon employeur ? L'État ?
- **Que se passe-t-il si cette donnée / information / fichier / dossier fuite et se retrouve dans les mains d'autrui ?**
(Collègues, partenaires, presse, grand public, concurrents, ...)
 - Quelles conséquences, quels impacts, quels préjudices ?
 - Pour moi ? Pour mon équipe ? Mon unité ? Mon employeur ? L'Etat ?

Sécurisation : éviter les pertes de données



Plusieurs façons de perdre ses données :

- **Problème matériel** : un disque externe abîmé, une clé USB détériorée, un portable qui ne démarre plus, un mot de passe oublié, une coupure de courant sur un serveur ...
- **Problème sur les fichiers** : format obsolète, logiciel pour ouvrir ce type de fichier défaillant, ...
- **Problème sur la maîtrise du contenu** : fichiers non documentés, données incompréhensibles car non explicites ...
- **Problème de vol** matériel ou virtuel





- **Sécurisation matérielle interne :**

- Au niveau d'un **serveur de stockage** : techniques permettant de répartir les données sur plusieurs disques d'un même serveur afin d'améliorer la tolérance aux pannes lorsqu'un des disques a un problème
- Au niveau d'une **plateforme de stockage** : les fonctionnalités d'une plateforme de stockage constituée de plusieurs serveurs peuvent intégrer des sauvegardes ou des synchronisations permettant de sécuriser les données stockées lorsque l'un des serveurs a un problème ou devient inaccessible.

Exemple des plateformes Summer ou Mantis

- **Sécurisation via l'utilisateur :**

- règle de la sauvegarde 3-2-1 (3 copies sur 2 supports différents, avec au moins 1 une copie à distance), sauvegarde quotidienne
- Sécuriser son poste de travail (chiffrement, mises à jour, protection adaptée ...)
- Eviter d'oublier son ordi dans le train ...



- **Mots de passe :**
 - Personnel, robuste
 - Ne pas laisser son ordinateur avec une session ouverte
- **Gestion des droits d'accès :**
 - toutes les solutions de stockage offrent une gestion des droits d'accès des utilisateurs plus ou moins fine, plus ou moins sécurisée.
 - Depuis un accès ouvert via une url pour un stockage de document sur le cloud jusqu'à la gestion de droits unix pour des espaces de stockages associés à des machines de calcul par exemple
 - Cette gestion peut nécessiter une **certaine maîtrise technique**
- **Partage des données :**
 - Se poser les bonnes questions : qui doit pouvoir accéder à quelles données et avec quels droits (lecture, écriture, lecture-écriture ?)
 - Mettre en œuvre les droits minimaux, surtout si les données sont sensibles !



- **Intégrité :**
 - Pouvoir s'assurer que les fichiers de données n'ont pas été modifiés ou altérés
 - Lors de leur stockage, d'un traitement ou d'une copie par exemple
- **Mécanismes de contrôle de l'intégrité**
 - Utilisation de **fonction de hachage** qui convertit une valeur numérique en entrée en une valeur numérique de taille fixe en sortie
 - Déterministe et à sens unique : on ne peut pas retrouver le contenu original
 - Le hash correspond à l'empreinte digitale d'un fichier
 - Elle est unique
 - Exemple : calcul de la valeur de la fonction de hachage sur un fichier avant sa copie et après sa copie permet de s'assurer qu'on a exactement le même fichier
 - Utilisation de fonctions **SHA-256 ou supérieur** (MD5 et SHA1 obsolètes).
 - Exemple d'outil : Hashdeep (disponible sur tous les systèmes, <http://md5deep.sourceforge.net/start-hashdeep.html>)

Sécurisation : chiffrement des données



- **Chiffrement** : procédé de cryptographie grâce auquel on souhaite rendre la compréhension d'un document impossible à toute personne qui n'a pas la clé de chiffrement.
- Pour **augmenter la sécurisation des accès** aux données, il est possible de chiffrer tout un disque, des répertoires ou des fichiers. Ils ne seront alors lisibles qu'avec la clé de déchiffrement.
- Attention à ne pas perdre la **clé de déchiffrement** car il serait alors impossible de récupérer les données !
- Certaines tutelles exigent le chiffrement des **postes utilisateurs** (ordinateurs portables mais aussi tablettes et smartphones professionnels)
- Exemple d'outil : 7-ZIP (voir <https://www.cnil.fr/fr/comment-chiffrer-ses-documents-et-ses-repertoires>)





- **CHIFFREMENT DE MES SUPPORTS DE DONNÉES**

- **Top !**

Tous mes supports (PC fixe et mobile, tablettes, téléphones...) sont chiffrés

- **Oui pas mal**

Mon poste de travail principal est chiffré

- **Bof**

Je ne sais pas si mes supports sont chiffrés ou ne suis pas sûr ou suis sûr que non (mais je vais y travailler !!)

- **MES SAUVEGARDES**

- **Top !**

Je dispose d'une sauvegarde récente (moins de 7 jours)

- **Oui pas mal**

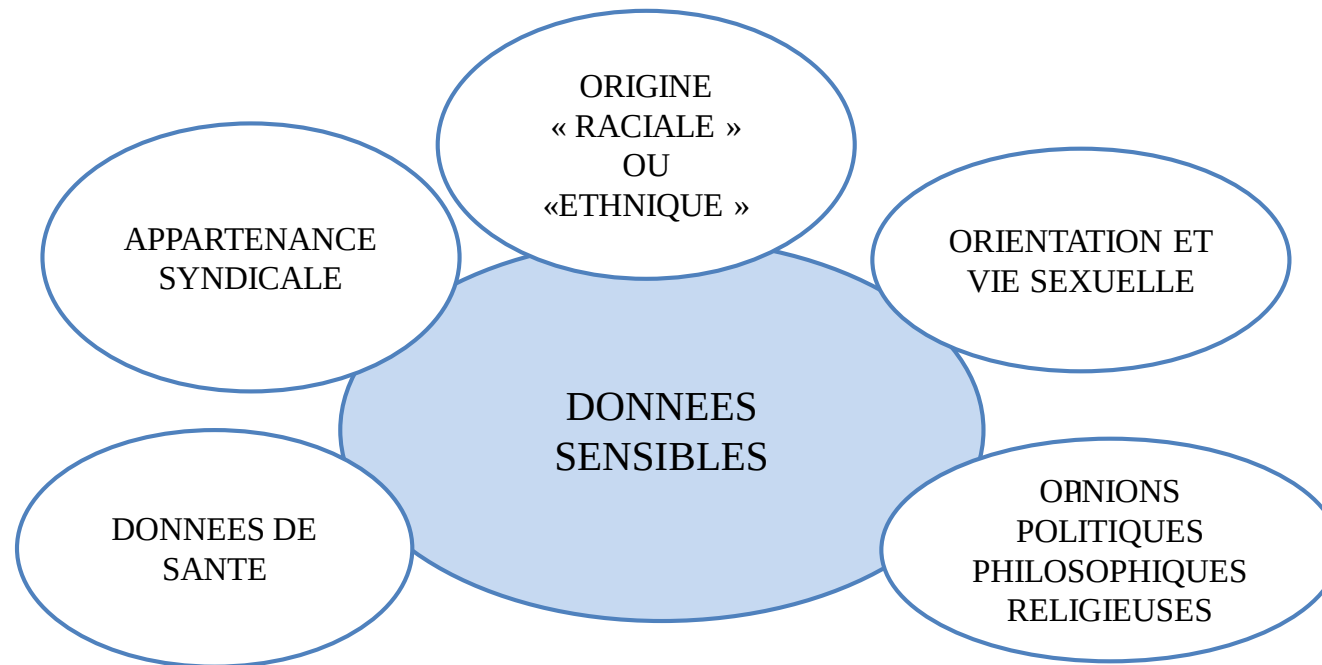
Je crois avoir une sauvegarde, je ne sais pas exactement de quand elle date

- **Bof**

Je ne dispose pas de sauvegarde fiable (mais je vais y travailler !!)



LES DONNEES PERSONNELLES SENSIBLES



Par principe la loi interdit le traitement des données sensibles sous peine de 5 ans d'emprisonnement et 300.000 euros d'amende.



- L'utilisation en recherche de données dites sensibles ne s'improvise pas
- Outre les obligations réglementaires (**module RGPD de ce cours le 13 juin**), il y a des contraintes techniques à respecter :
 - Mise en œuvre de techniques d'anonymisation ou de pseudonymisation
 - Cloisonnement des données et chiffrement
 - Protection particulière des tables de correspondance
 - Gestion fine des droits d'accès aux données

Se faire accompagner sur RGPD et données sensibles

A QUI DEMANDER ? QUAND DEMANDER ? QUE DEMANDER ?

Réponses lors de la séance du 13 juin

Contacter la cellule data : sos-data@univ-grenoble-alpes.fr



- <https://scienceouverte.univ-grenoble-alpes.fr/donnees/stocker/>

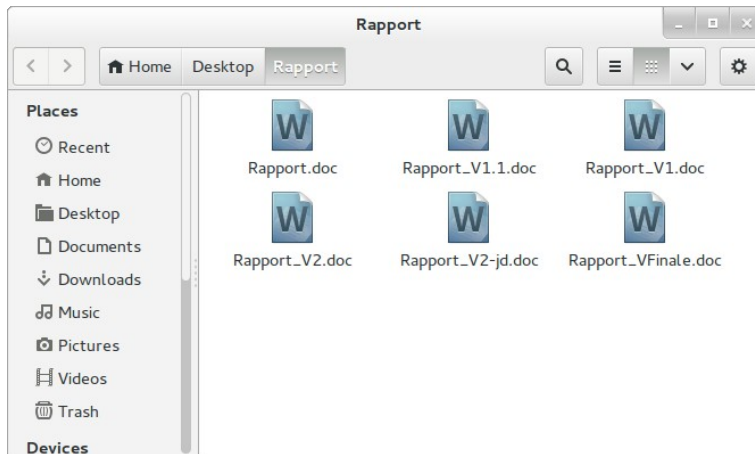
Description des différentes plateformes disponibles localement (opérées par l'UGA et hébergées dans les datacentres UGA) et de leurs caractéristiques

- **SUMMER** : stockage généraliste
 - **MANTIS, BETTIK, SILENUS** : stockage pour le calcul et le traitement de données
 - **NextCloud** : stockage partagé et collaboratif
 - **Gitlab** : gestion de version, travail collaboratif sur fichiers texte
- Si vous avez des questions sur ce sujet : contacter la Cellule Data Grenoble Alpes sos-data@univ-grenoble-alpes.fr



- Certains documents peuvent être amenés à évoluer, à avoir plusieurs versions avant une version finale, définitive.
- Deux solutions :
 - préciser la version dans le **nom du fichier**
 - Utiliser un **gestionnaire de version** (important quand plusieurs personnes sont susceptibles de modifier les fichiers)
- Pourquoi ?
 - Pouvoir rétablir une **version antérieure** en cas d'erreur
 - Pouvoir accéder à l'**historique des modifications** et donc de la vie du fichier
 - Pouvoir identifier les **auteurs des modifications** (utilisation d'un gestionnaire de version)
 - Eviter de travailler sur une **version périmée** du fichier

Pourquoi utiliser un gestionnaire de version ?



- La version la plus à jour est-elle Rapport.doc ou Rapport_VFinale.doc ?
- Les versions n'apparaissent pas dans l'ordre (1.1, 1, 2)
- La version 2-jd vient-elle avant ou après la version 2 ?

Source : <https://perso.liris.cnrs.fr/pierre-antoine.champin/enseignement/intro-git/>

- La gestion des versions est un travail fastidieux et méthodique.
- Les humains ne sont pas doués pour les travaux fastidieux et méthodiques.
- Laissons cela à l'ordinateur ...



Fichiers texte : fichiers markdown, fichiers source, fichiers latex, ...



Fichiers images, pdf, exécutable, ...

Gestion de version avec git



- Initialement dédiés à la **gestion de code source** pour les projets logiciels mais également :
 - Documentation
 - site web (wiki, git pages)
 - travail collaboratif : facilité d'échange, traçabilité, gestion des conflits
- **Principe** : consiste à conserver/maintenir toutes les versions d'un ensemble de fichiers.
- Deux fonctions principales :
 - permet le **travail simultané de plusieurs personnes** sur un ensemble de fichiers :
 - ✓ synchronisation des modifications, fusion automatique des fichiers,
 - ✓ détection et résolution (plus ou moins automatique ...) des conflits,
 - une gestion de l'**historique du projet** :
 - ✓ accès à n'importe quelle version archivée,
 - ✓ information sur qui a fait une modification, quand, où, pourquoi ...
 - ✓ notifications automatiques aux participants
- Pour aller plus loin :
 - <https://pole-calcul-formation.gricad-pages.univ-grenoble-alpes.fr/ced/>





- **PPST : Protection du Potentiel Scientifique et Technique**
 - risque autour de l'intelligence économique
 - risque terroriste
 - risque lié à la prolifération des armes de destruction massive
 - risque lié à la défense et l'armement
- **GAFAM** = essentiellement des services gratuits. **L'utilisateur est le produit**
 - Pas de cession de droits sur les données mais elles sont exploitées (ciblage publicitaire, statistiques ...) → perte de la maîtrise sur ces données, il est impossible de savoir exactement à quoi elles vont servir.
 - Par exemple, utilisation de gmail → Google scanne tous ces mails en particulier à des fins de ciblage publicitaires (mais pas que)



- US, Russie, Chine : **directives extra-territoriales**, cad textes qui s'appliquent même en dehors de leur territoire propre.
 - **Patriot Act** (2001) : permet aux services de sécurité américains d'accéder et d'analyser toutes les données issues de toutes les entreprises sans autorisation préalable et sans en informer les utilisateurs.
 - **Cloud Act** (Clarifying Lawful Overseas Use of Data Act – 2018) : permet au gouvernement américain d'obliger les prestataires de service à divulguer les données personnelles de leurs utilisateurs, dès lors que les autorités américaines (police, justice et administration) le leur demandent.
 - → géants du Web comme Microsoft, Google, Facebook ... **ne peuvent garantir** à leurs utilisateurs la confidentialité de leurs données personnelles, même si ces dernières sont stockées en Europe.
- **Privacy shield** (2016) : réglemente le transfert de données entre l'UE et les USA.
 - **invalidée en juillet 2020** par la cour de justice de l'UE en raison des pratiques de surveillance de masse en vigueur aux USA



- **Attention à l'utilisation de cloud commerciaux**, en particulier hors Europe
 - Vigilance vis à vis des problématiques d'intelligence économique
 - Vigilance vis à vis des données personnelles (sociétés de e-commerce, assureurs, ... exploitent les données personnelles).
- En particulier, sur les **questions RGPD**, et en raison des réglementations décrites ci-dessus, le chercheur se met dans ce cas dans l'illégalité.
 - Réglementation RGPD : règles très précises sur la sous-traitance du stockage : **pas de transfert en dehors de l'UE** sauf garanties contractuelles et vérifiables.





- Depuis juillet 2020, directives concernant l'ESR et les EPST
 - utilisation d'un cloud privé éventuellement possible mais :
 - ➔ **qualification SecNumCloud** de l'ANSSI
 - ➔ Ou se baser sur le **guide externalisation de l'ANSSI** : exploitation par des équipes européennes, conformité PSSIE (dont hébergement sur le territoire national), RGPD ...
- Dans le cas des **IRR / ZRR** : usage des clouds privés strictement interdit.
 - Relève du pénal en cas d'infraction
- **Solutions alternatives** :
 - solutions académiques
 - opérateurs privés mais coûteux : nécessité d'appels d'offre et de marchés publics. Bien inclure des clauses de sécurité dans ce cas.
- Dans le cas de solution comme AWS : règlement par **carte achat ou par CB personnelle**.
 - responsabilité du porteur de la carte achat engagée, ainsi que celle du chercheur : usage illégal de la carte achat et transgression de la réglementation PPST.
- Si aucune solution ne peut être trouvée, remonter aux **RSSI des établissements** concernés.



- Utiliser un / des **stockages adaptés à vos besoins**
- **Sauvegarder** régulièrement !
- Eviter les services **non académiques** pour vos données professionnelles
- Si vous avez des interrogations :
 - Échanger avec votre **directeur/directrice de thèse**
 - Contacter la **cellule data** : sos-data@univ-grenoble-alpes.fr

