Sécurité et Protection des Données

Quelles données protéger et comment ?

Roland Trouville, LPP - 2023

Les 3 différents types de données

Types de données

Il existe 3 types de données:

	Non sensibles	Sensibles	Critiques
Définition	• Données ne présentant aucun risque / ne contenant aucune information confidentielle ou personnelle	• Données à caractère personnel • Résultats non encore publiés	• Données personnelles pouvant porter préjudice • Données nominatives / biométriques
Général	Todo list, article déjà publié, script non breveté, etc…	Article en cours de rédaction, méthode brevetée, etc …	Identifiants informatiques, dossiers médicaux, etc…
Parole	• Enregistrements locuteurs neurotypiques en parole lue / pseudomots	• Enregistrements locuteurs neurotypiques en parole élicitée	• Enregistrements locuteurs pathologiques • Enregistrement parole spontanée • Données médicales (IRM, Scanner, etc …) • Fiches de renseignement locuteur

Pour la parole en général

Il faut se poser une question: Si cet enregistrement venait à être public, est ce que cela pourrait porter préjudice ou stigmatiser la personne enregistrée ?

- En aucun cas : Donnée non sensible (pseudo mots, parole lue)
- Peut être : Donnée sensible (hésitation, baffouillement, réponse un peu stupide à une élicitation, etc ...)
- Totalement : Donnée critique (Personne avec un trouble de la parole, parole spontanée abordant des sujets très privés, de famille / de couple, ou apportant un jugement sur une personne tierce, etc...)

Et ensuite il y a une deuxième question : L'enregistrement permet-il de reconnaître la personne enregistrée ?

- Oui : Le locuteur donne des détails personnels (nom, ville, etc) ou a un trouble de la parole → la donnée devient immédiatement critique indépendamment de la question 1

Anonymisation des données

Les données doivent impérativement être anonymisées, même pour les données non sensibles.

Procédure d'anonymisation :

- Attribuer un code à la fiche de renseignement du locuteur (par ex **LOC01**)
- Attribuer un code **différent, non dérivé,** au fichier d'enregistrement (en général dans le nom de fichier) (par ex **H_1809-1**)
- Gérer un tableau de correspondance entre les deux codes, si possible sur papier, qui sera stocké ailleurs que les données

LOC01	H_1809-1
LOC02	H_2209-1

Le stockage des données

Stockage non pérenne

Pour stocker de manière non pérenne

- Les données non sensibles : où vous voulez (ordinateur perso, clefs usb, google drive, etc ..)
- Les données sensibles : sur un support chiffré non public (ordinateur, disque externe, ...)
- Les données critiques : sur un serveur de données sécurisé, non public, hébergé en Europe.

Stockage pérenne

Pour s'assurer que les données soient disponibles dans le futur (plusieurs années, décennies, voire siècles), il faut mettre les données sur un stockage dit pérenne.

En france, pour les SHS, une solution: Huma-Num (et sa branche Nakala). TGIR d'archivage. Il est possible d'y poser des données (avec un compte HumanID) qui ne seront rendues publiques qu'au bout de X années.

Attention, y mettre des données qui ne bougent plus dans le temps. Ce ne sont pas des espaces de travail.

Au LPP

Stockage non pérenne au LPP

Pour les données sensibles et critiques:

- Serveur nextcloud du laboratoire (https://nextcloud.laboratoirephonetiquephonologie.fr) : serveur hébergé en France et géré par le LPP
- NAS Data : Serveur en raid 1 3 Disques, accessible uniquement depuis le réseau filaire au laboratoire.
- Votre ordinateur et vos stockages externes doivent être chiffrés selon le règlement et la PSSI du CNRS.
    - Ne chiffrez pas tout seul, il faut vous rapprocher de moi pour que je puisse archiver les clefs de récupération et éviter des pertes de données.