Sécuriser ses données

Certaines catégories de données produites ou collectées dans un cadre de recherche requièrent un niveau de sécurisation plus élevé et une attention particulière portée à leurs conditions de stockage et de partage.

Pour assurer la sécurité physique de vos données face aux risques de perte, il convient de recourir à de bonnes pratiques de stockage.

Si l'exploitation des données à caractère personnel et sensibles (renvoi vers démarche 22) dans un cadre de recherche scientifique est autorisée par le RGPD (Règlement général sur la protection des données), elle nécessite néanmoins la mise en œuvre de précautions particulières.

Outre le choix d'un stockage sécurisé, il est notamment recommandé de procéder au **chiffrement** de ces données pendant le projet. En effet, le chiffrement garantit la confidentialité des données en permettant seulement aux personnes légitimes d'avoir accès à ces données et à leur contenu. Les fichiers sont chiffrés à l'aide d'une clé : cette clé est ensuite nécessaire pour déchiffrer le document, que ce soit par les humains ou les machines. Sans clé, la donnée reste illisible. Plusieurs logiciels permettent de chiffrer soi-même simplement ses données et/ou répertoires de fichiers.

Pendant le projet il est également conseillé de **pseudonymiser** les données, c'est-à-dire de remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro séquentiel, etc.).

La pseudonymisation permet ainsi de traiter les données d'individus sans pouvoir identifier ceux-ci de façon directe. Mais, dans la mesure où il est toutefois bien souvent possible de retrouver l'identité de ceux-ci grâce à des données tierces, et où l'opération est réversible,les données concernées conservent donc un caractère personnel.

Afin de permettre une exploitation plus large de ces données et une réutilisation par des tiers, une autre option est possible : **l'anonymisation**. Ce procédé permet de supprimer le caractère identifiant d'un ensemble de données, et donc de diffuser ces données sans porter atteinte à la vie privée des personnes. Il repose sur un ensemble de techniques à mettre en œuvre : randomisation, généralisation.

Le chiffrement et l'anonymisation des données sont des actions irréversibles. La décision de choisir l'une ou l'autre de ces démarches dépend des usages que l'on envisage pour ses données et doit être anticipée le plus tôt possible dans le cadre d'un projet de recherche : l'anonymisation implique une nécessaire perte de qualité des données mais permet une diffusion plus large des jeux de données ; le chiffrement préserve l'intégralité des informations mais restreint de manière drastique les possibilités d'accès aux données et de réutilisation ultérieure.

Contact
Pour toute question liée aux données de recherche
guichet-ardoise@groupes.renater.fr
Fiona Edmond

Chargée des données de la recherche - SCD

fiona.edmond@univ-rennes2.fr
Oanez Helary

Animatrice Datalab MSHB Rennes

oanez.helary@mshb.fr
Outils

- Chiffrement de données :
  - 7-Zip (Windows)
  - Filevault (macOS)
  - Zed! (payant)
- Chiffrement de répertoires de fichiers :
  - VeraCrypt (Windows, macOS ou GNU/Linux). Un tutoriel de ce logiciel réalisé par la CNIL est disponible en ligne : https://youtu.be/fMpzmkzAliE

Aller plus loin

- CNIL, Donnée personnelle
- CNIL, Donnée sensible
- ANSSI, Guide d'hygiène informatique, 2017
- CNIL, Les conseils de la CNIL pour choisir un bon mot de passe
- CNIL, Comment chiffrer ses documents et ses répertoires ?
- CNIL, Comprendre les grands principes de la cryptologie et du chiffrement
- CNIL, L'anonymisation de données personnelles