

Les données à caractère personnel pseudonymisées sont des données qui ne peuvent être mises en relation avec des personnes physiques identifiées ou identifiables qu'au moyen d'un code (ou des données qui ne peuvent être mises en rapport avec des personnes spécifiques sans avoir recours à des données complémentaires qui sont conservées ailleurs). L'organisation qui les traite à des fins de recherche scientifique n'est pas en mesure de les mettre en relation avec les personnes physiques concernées.

L'organisation qui traite les données à caractère personnel à des fins de recherche scientifique dispose uniquement d'un code sans signification pour l'identification des personnes concernées. Il ne suffit d'ailleurs pas de supprimer le numéro d'identification, le nom, le prénom et l'adresse de la personne concernée pour qu'il s'agisse de données à caractère personnel pseudonymisées. Il faut éviter que leurs caractéristiques soient disponibles de manière trop précise.

La BCSS veillera, dès lors, en particulier à minimiser les possibilités de réidentification des personnes concernées auxquelles les données se rapportent. Ainsi, les données fournies doivent être suffisamment généralisées (en utilisant par exemple des classes de valeur suffisamment larges) et la quantité de critères de sélection à combiner ne peut être telle que le nombre de personnes qui y satisfont devient minime (et que les personnes deviennent par conséquent identifiables).

La BCSS applique à cet effet les règles suivantes:

Pas de dates exactes

Aucune date exacte n'est communiquée. Les dates sont de préférence exprimées en années ou en années et trimestres. Le mois et le trimestre sont communiqués lorsque l'étude le requiert. Une autre possibilité consiste à définir le nombre de jours entre 2 dates et à diviser le nombre en classes.

Pas de montants exacts

Aucun montant exact n'est communiqué. Les montants doivent toujours être divisés en classes.

Pas de domiciles précis

Aucune adresse de domicile n'est communiquée. Cette donnée est de préférence agrégée à un niveau supérieur. Plus le niveau est bas, plus il faut prouver que ce niveau est nécessaire.

Application de la règle 1-3 ou de la règle 1-9 (origine)

Cette règle est appliquée lorsque des tableaux croisés sont demandés. Si une cellule déterminée (une certaine combinaison de variables) représente un trop petit nombre de personnes, c'est-à-dire moins de 3 ou moins de 9, alors le nombre effectif est remplacé par 1-3, respectivement 1-9.

Divisions de variables en classes plus larges

Cette règle s'applique non seulement aux données à caractère personnel pseudonymisées mais aussi aux données anonymes (agrégées). Les classes d'âge, les classes de salaire, etc. peuvent être élargies.

Créer plusieurs tableaux à partir d'un seul tableau croisé dans lesquels plusieurs variables sont croisées

Lorsqu'il s'avère qu'un tableau croisé contient trop de petites cellules avec de petites valeurs, une solution consiste à créer plusieurs tableaux à partir d'un seul tableau croisé dans lesquels plusieurs variables sont croisées.

Domicile	sexe	classe d'âge	nationalité	position socio-économique	nombre
Bruges	1	18-24	belge	Occupé	1
Gand	2	25-34	UE	Chômeur	2
Courtrai	2	25-34	UE	Occupé	1
Alost	1	35-49	Afrique	Chômeur	1
...					



Domicile	sexe	classe d'âge	nombre	domicile	nationalité	position socio-économique	nombre
Bruges	1	18-24	14	Bruges	belge	Occupé	37
Gand	2	25-34	19	Gand	UE	Chômeur	26
Courtrai	2	25-34	24	Courtrai	UE	Occupé	18
Alost	1	35-49	20	Alost	Afrique	Chômeur	54
...							

Méthode en 2 phases

Cette méthode s'applique aux données à caractère personnel pseudonymisées lorsqu'une grande banque de données doit être consultée. Cette banque de données est placée sur un serveur de la BCSS dans un environnement sécurisé. Le chercheur reçoit un accès à cette banque de données dans les locaux de la BCSS. À titre de préparation, un petit échantillon issu de cette banque de données est transmis au chercheur. L'idée est que le chercheur développe et teste ses applications sur cet échantillon et qu'il les exécute ensuite sur la banque de données complète dans les locaux de la BCSS. Le chercheur emporte uniquement les résultats de cette exécution, et ce sous la forme de données agrégées.

Brouillage de données

Le brouillage est une technique consistant à remplacer les données réelles par des données ne représentant plus de situation réelle. Les valeurs peuvent être remplacées par des valeurs fictives ou les valeurs peuvent être interchangées (par exemple, le domicile Anvers est remplacé par le domicile Hasselt). Plusieurs conditions peuvent aussi être imposées (p.ex. les valeurs doivent se situer dans une limite donnée, certaines valeurs sont interdites, etc.). Des conditions peuvent aussi être imposées entre certaines variables de sorte que certains liens soient maintenus (p.ex. une valeur X de la variable A va toujours de pair avec une valeur Y de la variable B). Les ensembles de données brouillés ne servent pas de base pour l'exécution d'analyses, mais bien à la préparation des applications qui peuvent être exécutées sur les données réelles sur un serveur sécurisé dans les locaux de la BCSS. Le chercheur emporte uniquement les résultats de cette exécution, et ce sous la forme de données agrégées.