

# LA PRÉPARATION DES DONNÉES EN SCIENCES SOCIALES POUR SWISSUBASE : EN BREF

Pour déposer vos données dans [SWISSUbase](#) et les rendre accessibles à d'autres chercheurs·euses, vous devez décrire votre projet de recherche et y ajouter un (ou plusieurs) jeu(x) de données. Par ailleurs, les données doivent répondre à certaines exigences.

En effet, un jeu de données doit être composé au minimum d'un **fichier de données nettoyées, anonymisées** dans un des formats listés plus bas (cf. p. 2), ainsi que de la **documentation** nécessaire à la bonne interprétation et à la réutilisation des données.

## Nettoyage des données

### Données quantitatives :

- Les labels de variables et de valeurs doivent être inclus dans le fichier de données. Les labels peuvent être constitués à partir de questions d'enquête raccourcies. Les labels peuvent être formés à partir des questions de l'enquête raccourcies. Les liens entre les variables dans le fichier de données et les questions correspondantes du questionnaire doivent être clairs.
- Des codes explicites doivent être attribués aux valeurs manquantes (par ex. ne sait pas, refus, filtre, non réponse).
- Les fréquences doivent être contrôlées et les inconsistances ou erreurs détectées doivent être corrigées ou supprimées (par ex. valeurs invalides ou hautement improbables).

### Données qualitatives :

- Des informations sur la récolte des données – telles que la date, le lieu, l'identifiant unique de l'interviewer et de la personne interviewée – doivent être disponibles soit dans le fichier (par ex. transcriptions), soit dans un document séparé.

## SWISSUbase

[SWISSUbase](#) est une plateforme en ligne dédiée à l'archivage et à la dissémination des données de recherche et de leurs métadonnées. Différents services pour la gestion des projets de recherche sont également proposés.

Issue de la collaboration de plusieurs institutions, SWISSUbase s'appuie sur des normes d'archivage internationales pour garantir la préservation et l'accessibilité des données à long terme et répondre aux besoins des universités et autres organismes de recherche.

**SWISS**  **base**

## Anonymisation

**Il est de la responsabilité des chercheurs·euses d’anonymiser leurs données avant de le dépôt** de manière à ce qu’aucun individu ne puisse être identifié sans “efforts disproportionnés”. Ce processus d’anonymisation nécessite un traitement rigoureux des identifiants directs et indirects :

- **Les identifiants directs** correspondent à toutes les variables ou autres éléments qui permettraient d’identifier directement des individus (par ex. noms, adresses, numéros de téléphone). Ils doivent être supprimés ou remplacés par des pseudonymes ou des catégories.
- **Les identifiants indirects** correspondent à toutes les variables ou autres éléments qui, combinés, pourraient permettre d’identifier un individu (par ex. âge, profession, code postal). Ils doivent également être contrôlés et recodés ou supprimés si nécessaire.

Si’il n’est pas possible d’anonymiser les données, un **consentement** explicite des participant·e·s est nécessaire pour l’archivage et le partage des données. Un exemplaire du formulaire de consentement doit alors être fourni.

SWISSUbase offre différentes options pour garder le contrôle sur l’accès aux données : accord préalable, embargo, restrictions d’utilisation.

## Formats

En fonction du type de données, les formats suivants sont à privilégier dans SWISSUbase :

- **Données en tableaux** : SPSS data file & portable format (.sav; .por), MS Excel (.xls; .xlsx), Comma-separated value file (.csv).  
Les fichiers tels que les fichiers de données Stata (.dta), Statistical Analysis System (.sas), R Data Format (.RData ; .rds), OpenDocument spreadsheet (.ods) sont également acceptés, mais doivent être convertis en SPSS, Excel ou CSV par le producteur de données afin que des contrôles de qualité puissent être effectués par le service de données de FORS.
- **Données textuelles** : Portable Document Format (.pdf), Rich Text Format (.rtf), ASCII (.txt), MS Word (.doc; docx), OpenDocument Text (.odt).
- **Données constituées d’images** : JPEG (.jpeg; .jpg; .jp2), TIFF (.tif; .tiff), RAW image format (.raw), Bitmap Image File (.bmp), Portable Network Graphics (.png), Portable Document Format (.pdf).
- **Données audio** : MP Format (.mp3), FLAC (.flac), Waveform Audio Format (.wav), Advanced Audio Coding (.aac).
- **Données vidéo** : MP Format (.mp4), QuickTime video (.mov), Fichier AVI (.avi), Motion JPEG 2000 (.mj2).

Les dossiers et les fichiers dans les formats susmentionnés peuvent être déposés au format ZIP (.zip ; .rar ; .7z).

## Documentation

Le dépôt devrait inclure toute documentation nécessaire à la bonne interprétation des données et permettant d’assurer leur qualité, à savoir :

**Un rapport décrivant les aspects suivants :**

- Le projet : contexte, problématique, objectifs, hypothèses.
- La méthode : population, échantillonnage (taille, taux de réponse), mode de collecte des données.
- Les données : nettoyage, anonymisation, codage, validation, le cas échéant description des variables recodées ou construites, pondération.

**Les documents relatifs à la récolte des données :**

- Le matériel envoyé à l’avance aux répondant·e·s (par ex. courriers préliminaires) ;
- Les instructions ou matériel pour les enquêteur·trice·s ;
- Le matériel présenté aux répondant·e·s durant les entretiens ;
- Les instruments utilisés pour la récolte des données (questionnaires, grilles, protocoles).

**Si disponibles, les documents suivants sont aussi à ajouter :**

- Les publications et rapports finaux ;
- Le codebook ;
- Les instructions de codage ;
- Les syntaxes ;
- Le Data Management Plan.

Les fichiers de documentation doivent être soumis au **format PDF** (à l’exception des syntaxes) et dans les différentes langues disponibles.

### Information et ressources

- [Ressources SWISSUbase](#)
- [Data Management](#) (Site Internet de FORS)
- [FORS Data protection policy](#) (Site Internet de FORS)
- [The informed consent as legal and ethical basis of research data production](#) (FORS Guide n°5)
- [How to draft a DMP from the perspective of the social sciences, using the SNSF template](#) (FORS Guide N°07)
- [Data anonymisation: legal, ethical, and strategic considerations](#) (FORS Guide n°11)
- [Data Management Expert Guide: Anonymisation](#) (CESSDA recommendations for data anonymisation)

### Besoin d’aide?

[dataservice@fors.unil.ch](mailto:dataservice@fors.unil.ch)