

Les dix commandements de la saisie

Remarque liminaire : la plupart de ces recommandations visent à segmenter au maximum l'information lors de la saisie. Il est en effet techniquement plus simple, ensuite, de regrouper (de façon en général automatique) que de scinder (de façon en général manuelle) le contenu des cases. Éclater les informations facilite les opérations de tri (qui se font sur le premier caractère de la case). Cela conduit à multiplier les colonnes, ce qui ne doit pas, en soi, poser de problème. Si l'on dépasse les limites en la matière d'un tableur, il est toujours possible d'ouvrir une nouvelle « feuille de calcul », tant qu'elle renvoie bien aux mêmes individus.

1. Réserver une et une seule ligne, la première, aux intitulés des variables.

2. Utiliser une colonne pour attribuer des identifiants numériques arbitraires aux individus étudiés (individu n° 1, n° 2...). Ils servent à éviter les

ambiguïtés liées à la typographie (accents...) ou à l'homonymie et parfois à relier des informations issues de différentes bases. Il n'est pas gênant que certains chiffres ne soient pas ou plus attribués comme identifiants (suite par exemple à la découverte que deux individus ne font en fait qu'un).

3. Conserver le plus possible les formulations de la source. S'il faut simplifier, moderniser l'orthographe, etc., le faire *a posteriori*, dans une autre colonne ou un autre fichier.

4. Conserver le lien avec la source. Selon qu'une source est à l'origine de toute la base, de toutes les données pour une personne, pour une variable, d'une seule donnée, etc., les solutions varient, mais la référence de la source doit apparaître. On peut ainsi avoir une source par *classeur*, une source par *feuille d'un classeur*, une source par ligne ou colonne (indiquée au début de celle-ci), une colonne « source des renseignements de la colonne précédente », etc. L'essentiel est de conserver l'information précise

dans au moins une version de la base de données, quitte ensuite à procéder à des regroupements et simplifications.

5. Utiliser les fonctions d'annotation ou de commentaire des logiciels, par exemple pour mentionner la source particulière d'un renseignement, un choix de transcription ou de codage, un doute sur une graphie...

6. Scinder le plus possible l'information, par exemple avec des colonnes « M./Mme/Mlle », « nom », « prénom 1 », « prénom 2 », « prénom 3 », « titre de noblesse », « nom de jeune fille » et « pseudonyme » plutôt qu'une seule colonne « nom ». De même pour les adresses : non que le numéro soit assez intéressant pour constituer une variable, mais parce que le mettre dans une colonne à part permet de trier par nom de rue.

7. Conserver le lien avec la date de l'information. Si la profession est connue à plusieurs dates, faire une colonne par date, ou des colonnes « profession 1 », « date de la

profession 1 », « profession 2 », « date de la profession 2 », etc. Si les dates sont en fait des intervalles, faire une colonne pour la date de début et une pour la date de fin.

8. Éviter, sauf si on travaille seulement sur le xx^e siècle, le format « date » des logiciels, difficilement exportable. Préférer le codage en trois colonnes : jour, mois et année (en format « nombre »), qui permet des tris plus fins.

9. Aller explorer les fonctions suivantes dans l'aide du tableur : figer les volets, ajustement automatique de la largeur des colonnes, sélections multiples, copie automatique ou incrémentation, collage spécial, remplacer, insertion de fonction, concaténer, mise en pages, zone d'impression, tri, filtre, tableau croisé dynamique ou pilote de données ; et regarder tout ce que l'on peut faire avec un clic droit sur un PC (ou Ctrl + clic sur Macintosh).

10. Sauvegarder aussi souvent que possible.