

ANNEXE 1-B

Principes et outils de gestion des données

1. Types de données
2. Métadonnées
3. Tâches générales de gestion de données
4. Étapes de la gestion de données
5. Aspects géographiques/spatiaux

1. Types de données

- Données primaires
ex. : Enquête dans le cadre d'un projet
- Données secondaires publiées
ex. : Données de Statistique Canada
- Données secondaires non publiées
ex. : Données de fichiers administratifs
(comme le rôle d'évaluation)

1. Types de données

- Recherche
 - Sources
 - Accès
- Gestion

2. Métadonnées

- « Au-delà » des données :
Données sur les données
 - 2.1 Source
 - 2.2 Portée, univers de référence
 - 2.3 Concepts, définitions, variables
 - 2.4 Structure, organisation
 - 2.5 Méthode de collecte
 - 2.6 Évaluation de la qualité

2. Métadonnées

2.1 Source

- Quelle est l'agence, l'organisation qui a produit ces données ?
- Quelle est la crédibilité de l'émetteur ?
- Comment les données ont-elles été obtenues ?

8

Méthodes quantitatives © André Lemelin, 2002

2. Métadonnées

2.2 Portée, univers de référence

- Population couverte
 - « Population » s'entend dans le sens statistique
 - = ensemble de personnes ou de choses...
 - (ex. : « Population » d'entreprises, de villes...)
- Limites spatiales, géographiques
 - Tout le pays, une région, une municipalité...
- Période(s) de référence
 - À quelle ou à quelles années ou mois se rapportent les données ?

9

Méthodes quantitatives © André Lemelin, 2002

2. Métadonnées

2.3 Concepts, définitions, variables

- Quelles sont les unités d'observation ?
(cela fait aussi partie de la définition de la population couverte)
ex. : Familles ou individus ? « Familles », cela inclut-il les personnes vivant seules ?
- Quelle information (quelles variables) sur chaque unité d'observation ?

10

Méthodes quantitatives © André Lemelin, 2002

2. Métadonnées

2.3 Concepts, définitions, variables (suite)

- Quelle est la définition exacte de chaque variable ? De quel type de variable s'agit-il ?
ex. : Le « Revenu » inclut-il les revenus autres que de travail ?
- Quelles sont les unités de mesure de chaque variable ?
Exemples :
 - milliers ou millions ?
 - dollars canadiens ou dollars états-uniens ?
 - dollars courants ou dollars constants ? de quelle année ?

11

Méthodes quantitatives © André Lemelin, 2002

2. Métadonnées

2.3 Types de variables

- Variables *catégoriques* («nominal» en anglais)
à quelle catégorie appartient l'individu ?
 - Variable *dichotomique* : 2 catégories possibles
 - Variable *polytomique* : plus de 2 catégories
- Variables *ordinales*
classer les individus en ordre croissant ou décroissant
- Variables *d'intervalle*
comparer les différences entre individus
- Variables *rationnelles*
il y a un zéro naturel et le rapport entre deux valeurs a une signification

12

Méthodes quantitatives © André Lemelin, 2002

2. Métadonnées

2.4 Structure, organisation

- Format informatique (csv,dbf, excel...)
- Structure
Comment sont constitués les tableaux ?
Nous allons parler d'organisation plus loin

13

Méthodes quantitatives © André Lemelin, 2002

2. Métadonnées

2.5 Méthode de collecte

- Recensement ou échantillon ?
- Si échantillon, tiré comment ?
- Instrument(s) de collecte
 - questionnaire postal
 - questionnaire pour entrevue téléphonique
 - questionnaire pour entrevue en personne
 - entrevue semi-dirigée
 - etc.
- Conditions de la cueillette

14

Méthodes quantitatives © André Lemelin, 2002

2. Métadonnées

2.6 Évaluation de la qualité

- Couverture complète ?
- Données manquantes ?
- Données fiables ?

15

Méthodes quantitatives © André Lemelin, 2002

3. Tâches générales

- 3.1 Planifier
- 3.2 Exécuter
- 3.3 Mémoriser
- 3.4 Vérifier/valider
- 3.5 Documenter

3. Tâches générales

- 3.1 Planifier
 - Certaines décisions sont irréversibles
 - Planifier le « quoi ? » :
flexibilité c. économie
 - Qu'est-ce qui pourrait être utile ?
... parce qu'on ne peut pas tout prévoir
 - Qu'est-ce qui sera réellement utile ?
... parce que l'exécution coûte du travail
 - Planifier le « comment ? »
... parce que l'exécution mal planifiée
coûte davantage de travail

3. Tâches générales

3.2 Exécuter

- Les difficultés lors de l'exécution peuvent amener à revoir la planification
- Revoir la planification, plutôt que d'improviser une solution !

3. Tâches générales

3.3 Mémoriser

- Que mémoriser ? Tout !
 - Les données elles-mêmes
 - Les métadonnées
 - La structure d'organisation
 - La liste des fichiers et leur localisation
 - Les traitements ou transformations

3. Tâches générales

3.3 Mémoriser

– Pourquoi mémoriser ?

Sécurité

contre la destruction, la suppression accidentelle...

Transparence

pour que d'autres puissent les utiliser

Flexibilité

pouvoir revenir en arrière et vérifier ce que l'on a fait
pouvoir corriger les erreurs

21

Méthodes quantitatives © André Lemelin, 2002

3. Tâches générales

3.3 Mémoriser

– Comment mémoriser ?

Prendre des notes

Tenir un journal

Conserver des archives

Documents que l'on ne touche pas,
généralement sous forme de fichiers électroniques

ex. : original, original nettoyé, versions
successives, traitements, résultats...

Avec des noms différents

Avec un catalogue des fichiers et de leur description

22

Méthodes quantitatives © André Lemelin, 2002

3. Tâches générales

3.3 Mémoriser

- Méthodes de sécurité
 - Copies de sécurité
 - Disquettes verrouillées
 - Fichiers protégés contre l'écriture
 - Cellules protégées dans Excel

23

Méthodes quantitatives © André Lemelin, 2002

3. Tâches générales

3.4 Vérifier : procédures de validation

- Tests de cohérence
 - maximum, minimum, moyenne, fréquences
 - valeurs compatibles avec le type de variable
 - ...et avec son domaine de variation
- Totaux de contrôle
- Comparaison de calculs différents qui doivent logiquement donner le même résultat

**Valider, c'est pratiquer la méfiance
systématique de l'Inspecteur Colombo !**

24

Méthodes quantitatives © André Lemelin, 2002

3. Tâches générales

3.5 Documenter

- Documentation active : produire l'information
- Documentation passive : générée par les procédures de traitement de données (programmes, fichiers « log », formules d'Excel...) ...grâce à
 - la systématisation
 - la standardisation
 - l'automatisation

25

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.1 À la réception des données...

4.2 Organisation

4.3 Partage

4.4 Exploitation

27

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.1 À la réception des données...

- Sauvegarder les données originales
- Prendre connaissance des métadonnées
- Valider une première fois
 - Sont-ce bien les données que l'on attendait ?
 - Sont-elles conformes aux métadonnées ?
- Documenter le processus de réception des données

28

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.2 Organisation



29

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.2 Organisation

- Le « data modelling » : un monde !
Il existe plusieurs modèles d'organisation
- Réfléchir et planifier la structure d'organisation:
le modèle d'organisation des données adéquat est celui qui reflète le modèle *ex ante* de la réalité que l'on veut étudier
- L'organisation inclut les métadonnées !

30

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.2 Organisation

- Structure fondamentale : un tableau
 - Lignes = enregistrements, observations, objets
 - Colonnes = champs, variables, attributs
- Une clé d'identification unique pour les objets
 - éviter la confusion entre les homonymes
 - éviter les noms multiples pour un même objet (données multilingues, variations d'orthographes, erreurs...)

31

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.2 Organisation

– Structure fondamentale : un tableau

		Variables				
		X_1	X_2	X_3	...	X_k
Observations	1	x_{11}	x_{12}	x_{13}	...	x_{1k}
	2	x_{21}	x_{22}	x_{23}	...	x_{2k}
	3	x_{31}	x_{32}	x_{33}	...	x_{3k}
	4	x_{41}	x_{42}	x_{43}	...	x_{4k}

	n	x_{n1}	x_{n2}	x_{n3}	...	x_{nk}

32

4. Étapes de la gestion de données

4.2 Organisation : le cas des SIG

(Systèmes d'Information Géographique)

- Un SIG est une base de données...
- organisées en « couches » multiples d'information
- Représentation de l'espace :
fond de carte
- Représentation « dans » l'espace :
éléments géo-référencés
(associés à des lieux identifiés)

33

4. Étapes de la gestion de données

4.2 Organisation : le cas des SIG (Systèmes d'Information Géographique)

- Intégration de données de sources diverses
- Possibilités de croisements, de liens

34

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.3 Partage

- Pourquoi le partage ?
 - La recherche est de moins en moins individuelle, de plus en plus en équipe
 - Même le chercheur solitaire doit rendre compte
 - Rapport de recherche, mémoire, article scientifique...
 - Une exigence de plus en plus courante : donner accès à ses données pour que d'autres puissent confirmer les résultats en les reproduisant, selon l'exigence méthodologique de la reproductibilité
 - Partager... avec soi-même !
(quand on met de côté pour quelque temps, on oublie)

35

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.3 Partage

- Pour partager des données, il faut partager les métadonnées
- Établir des procédures pour...
 - modification des données
 - mémorisation
 - validation
 - documentation
 - etc.

36

Méthodes quantitatives © André Lemelin, 2002

4. Étapes de la gestion de données

4.4 Exploitation

Les analyses nécessitent une préparation adéquate des données

37

Méthodes quantitatives © André Lemelin, 2002

Conclusion

- L'importance des métadonnées
y compris la documentation
- L'importance de la mémoire
Tout mémoriser !