



Organisation des Nations Unies
pour l'alimentation et l'agriculture

Ouvrir l'accès aux microdonnées d'enquête agricole

4 – Documentation de fichiers de microdonnées

Webinaire régional - Afrique

10h00 – 13h00 (GMT+0), 30 Novembre 2021



Contenu

- **Avantages** de la documentation des microdonnées
- **Localisation de la phase de documentation** dans le Modèle générique du processus de production statistique
- Principales caractéristiques d'une **bonne documentation**
- **Normes internationales de métadonnées** (DDI et Dublin Core)
- L'éditeur de métadonnées **Nesstar Publisher** (DDI et Dublin Core)

Avantages de la documentation des microdonnées

Documenter les microdonnées :

- Aide à **comprendre** ce que les données mesurent et comment les données ont été créées.
De bonnes métadonnées réduisent le risque d'une compréhension erronée et d'utilisation inadéquate des microdonnées.
- Aide les utilisateurs à évaluer la **qualité des données**
Les chercheurs peuvent comprendre si les données sont pertinentes pour une question de recherche particulière.
- Aide les utilisateurs à **localiser** les données qui les intéressent. Tout système de catalogage et de localisation de ressources (manuel ou numérique) est basé sur les métadonnées.
- Aide à créer de la **transparence** et une crédibilité accrue des données

Avantages de la documentation des microdonnées

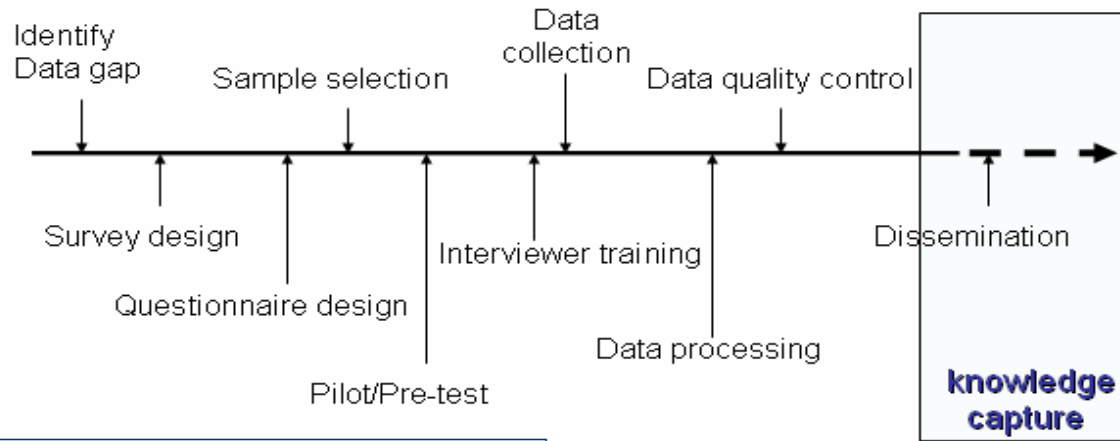
La documentation des microdonnées permet également :

- **Préserver** la mémoire institutionnelle
- **Générer** du matériel de formation pour les futurs statisticiens
- **Réplication** de la collecte et de l'analyse des données
- **Réutilisation** ou **réaffectation** des métadonnées

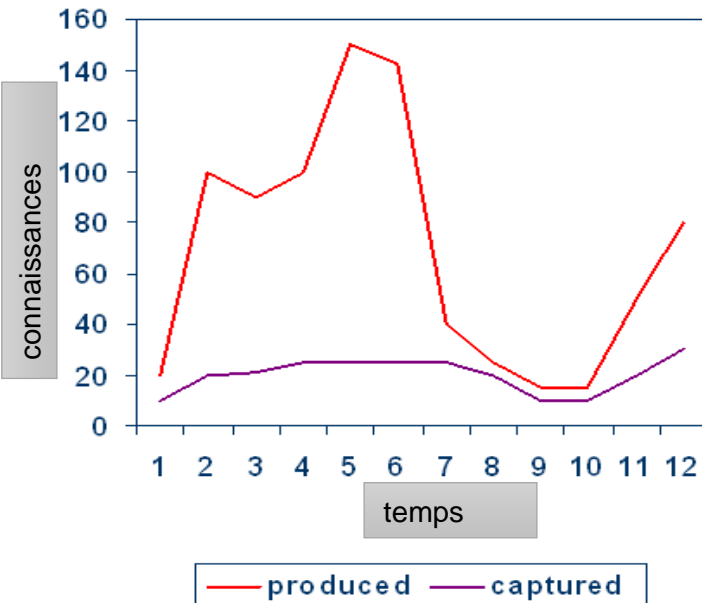
« *La norme de réplication stipule qu'il existe suffisamment d'informations pour comprendre, évaluer et s'appuyer sur un travail antérieur si un tiers peut reproduire les résultats sans aucune information supplémentaire de la part de l'auteur* »

<http://gking.harvard.edu/files/replication.pdf>

Quand initier la documentation les microdonnées?

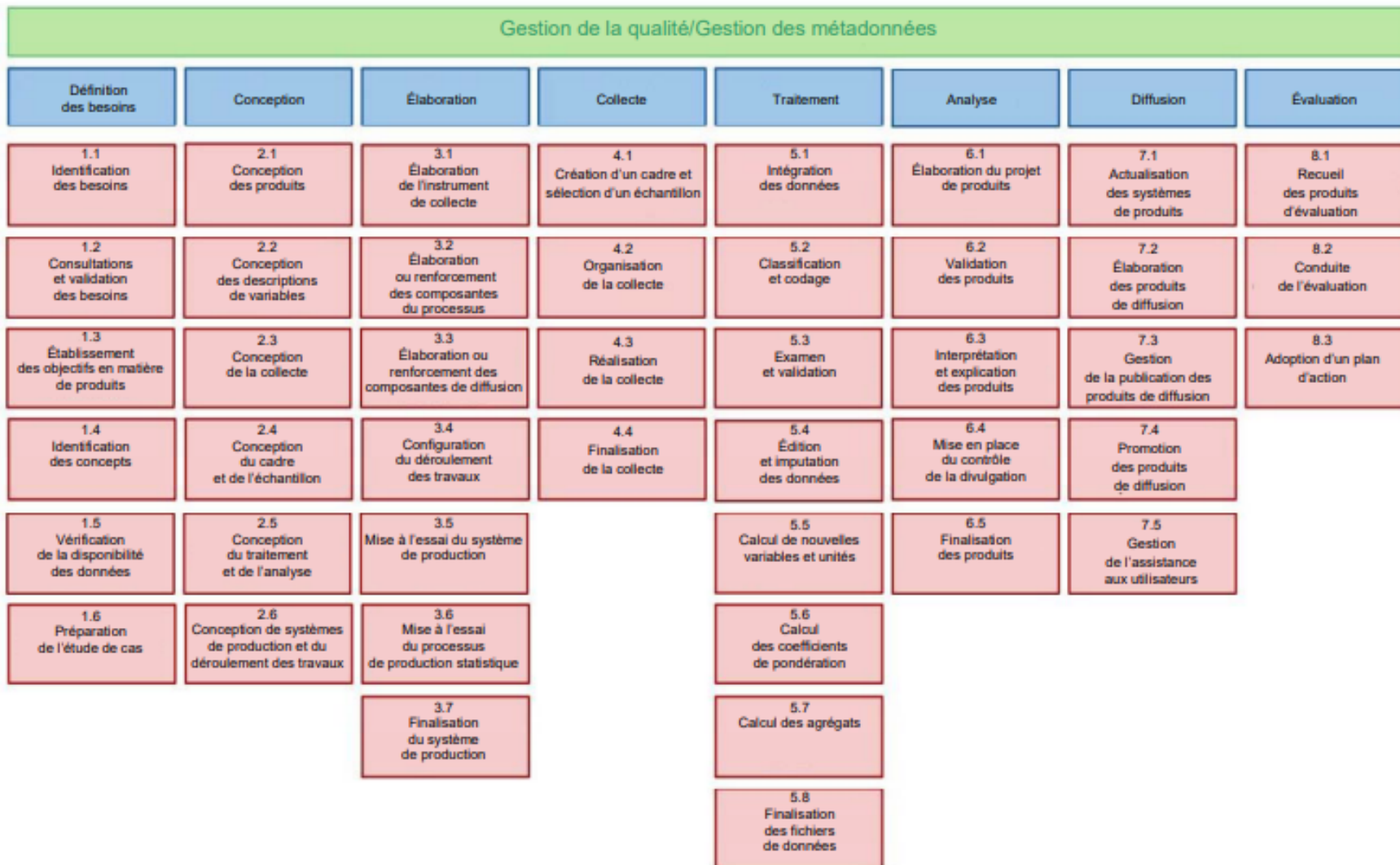


Document
As
You
Survey



En règle générale, les ensembles de données sont documentés après l'achèvement de l'ensemble de l'enquête ou de l'opération de recensement. À ce stade, il est cependant souvent trop tard pour saisir toutes les informations de métadonnées produites au cours du cycle de vie d'une activité de collecte de données.

Modèle générique du processus de production statistique (GSBPM)



Planifiez et initiez la documentation dès le début du programme d'enquête !

Évitez que des informations importantes ne soient enterrées et perdues sur un disque dur d'un membre du personnel. Centralisez toutes les données et métadonnées associées !

Qu'est-ce qu'une bonne documentation ?

Métadonnées : généralement définies comme des « données sur les données ». De bonnes métadonnées doivent contenir bien plus qu'un dictionnaire de données.

« Du point de vue de l'archiviste et de l'utilisateur final, un bon jeu de données est un jeu de données facile à utiliser. Sa documentation est claire et facile à comprendre, les données ne contiennent aucune surprise et les utilisateurs peuvent accéder à l'ensemble de données avec un temps de démarrage relativement court. »

Guide de préparation et d'archivage des données en sciences sociales, ICPSR

Du point de vue du producteur de données, une bonne documentation est celle qui apporte crédibilité et transparence aux résultats (tableaux et analyses); et qui permettrait de reconduire la collecte de données de telle façon à garantir la comparabilité des données.

Norme de métadonnées de l'Initiative de Documentation des Données (DDI)

Ensemble de normes recommandées pour documenter les microdonnées.

- Le DDI est une **norme de métadonnées** utilisée par la plupart des archives de données de sciences sociales dans le monde pour documenter les microdonnées.
- Le DDI fournit une **liste structurée et complète d'éléments (ou champs)** qui peuvent être utilisés pour documenter un ensemble de données.
- Le DDI a été conçu pour englober les types de données générées par les enquêtes, les recensements, les dossiers administratifs, etc.

Autre norme de métadonnées : Dublin Core

- Basé sur les mêmes principes que la spécification DDI.
- Se compose d'un ensemble d'éléments étiquetés (ou « balises »), organisés pour former un fichier XML.
- Comprend un nombre limité d'étiquettes (« balises »), telles que le titre, la description, le type, le créateur, l'éditeur, la date, le format et la langue.

What is XML?

- XML est l'acronyme de eXtensible Markup Language. Un langage de structuration des informations utilisé notamment pour la gestion et l'échange d'informations sur le Web ou à échanger entre des systèmes logiciels.
- XML est un format de fichier, lisible par n'importe quel éditeur de texte (par exemple, le Bloc-notes).
- Texte de balises XML pour le sens.
- Dans un fichier XML, les informations sont encapsulées entre une étiquette d'ouverture et une étiquette de fermeture. Le nom de l'étiquette indique son contenu.

Qu'est-ce que XML ? Un exemple

« L'INS de Popstan a mené une enquête agricole annuelle avec le soutien financier de la FAO en 2019. 5 000 ménages, représentant l'ensemble de la population du pays, ont été sélectionnés sur base aléatoire pour participer à l'enquête, selon une méthodologie d'échantillonnage stratifié en deux étapes ». 4 900 de ces ménages ont fourni des informations ».

En langage XML, cela pourrait ressembler à ceci :

```
<titl> Enquête agricole annuelle 2019</titl>
<altTitl> EAA 2019 </altTitl>
<AuthEnty> Institut national des statistiques (INS)
</AuthEnty>
<fundAg abbr= "FAO">Organisation des Nations Unies
pour l'alimentation et l'agriculture </fundAg>
<nation> Popstan </nation>
<geogCover> National </geogCover>
<sampProc> 5 000 ménages, stratifiés en deux étapes
</sampProc>
<respRate> 98 % </respRate>
```

Cela n'a pas l'air très facile... mais heureusement nous n'avons pas besoin d'apprendre à écrire le langage XML

Génération de DDI à l'aide d'un éditeur de métadonnées

- Les fichiers DDI peuvent être générés facilement à l'aide d'un éditeur de métadonnées spécialisé.
- Le réseau international pour les enquêtes auprès des ménages (*International Household Survey Network* (IHSN)) et la Banque mondiale ont développé une boîte à outils de gestion des microdonnées conforme à la DDI basée sur l'éditeur Nesstar (logiciel gratuit du Norwegian Center for Research data)
- Cette boîte à outils propose un éditeur de métadonnées convivial ainsi que d'autres outils pour convertir les métadonnées DDI en catalogues d'enquêtes en ligne.



Des outils pour documenter et publier

Un éditeur de métadonnées DDI spécialisé (Nesstar)

The screenshot shows the Nesstar DDI metadata editor interface. It features a menu bar (File, Edit, Documentation, Variables, Data, Publishing, Tools, Help) and a toolbar. The main workspace is divided into several panels:

- Variables:** A table listing variables with columns for Number, Name, Label, Width, StartCol, EndCol, and Data Type. The table contains several rows of variable definitions.
- Data Entry:** A table showing data entry for variables v1-hid, v2-hi6, v3-hi3d, v4-tbegun, and v5-tend. The table has 6 rows of data.
- Projects:** A tree view on the left showing a project structure with folders for Document Description, Study Description, Other Study Materials, Datasets, File Description, Key Variables & Relations, and External Resources.
- Metadata Fields:** A list of metadata fields on the right, including Description of Scope, Topics Classifications, Coverage, Country, Producers and Sponsors, Primary Investigator, Funding, Other Acknowledgments, Sampling, Data Collection, Dates of Collection, Time Periods, Mode of Data Collection, Notes on Data Collection, Questionnaires, Data Processing, Data Entry and Editing, Other Processing, Data Appraisal, Estimates of Sampling Error, Data Access, and Confidentiality.

The screenshot shows the NADA catalog/diffusion solution interface. It features a menu bar (DOCUMENTATION, STUDY DESCRIPTION, DATA DESCRIPTION, GET MICRODATA) and a toolbar. The main workspace is divided into several panels:

- Overview:** A panel showing the overview of the study, including the title, country, and series information.
- Data Description:** A panel showing the data description, including the data file name, content, cases, variables, and notes.
- Documentation:** A panel showing the documentation, including the questionnaire and reports.
- Questionnaires:** A panel showing the questionnaires, including the questionnaire and reports.
- Reports:** A panel showing the reports, including the report and reports.

Below the main workspace, there is a section titled "Et une solution de catalogage/diffusion (NADA)" with the URL <http://nada.ihsn.org/>. An orange arrow points to the right.

Conclusions et recommandations

Document As You Survey

Diffuser

Adopter et se conformer aux normes internationales de métadonnées

Faire de l'assurance qualité de la documentation une activité régulière

MA / INS

Merci

Références et liens:

- Logiciel:
 - Nesstar publisher: <http://ihsn.org/software/ddi-metadata-editor>
- Documents techniques:
 - Quick Reference Guide for Data Archivists: http://ihsn.org/sites/default/files/resources/DDI_IHSN_Checklist_OD_06152007.pdf
 - Principles and Good Practice for Preserving Data: <http://ihsn.org/sites/default/files/resources/IHSN-WP003.pdf>