**Documentation Guidelines for IFPRI Datasets**

**IFPRI policy on datasets:** Adherence to IFPRI policy on datasets is mandatory. Before submitting the dataset for making it as public goods on IFPRI’s website, please ensure whether they comply to the IFPRI policy. IFPRI will make all primary and value-added secondary datasets collected after January 1, 1999 publicly available two (2) years after all data collection ceases or, before two years, at the time of a major publication like research reports, books, or journal articles. For complete details of IFPRI policy on datasets, please visit ([IFPRI's Policy on Dataset Management](http://intranet.ifpri.org/internal/smt-dec/InstituteWidePolicyDecisionsorPolicyChanges/2010/IFPRI_Datapolicy_OCT_20_2010.docx)).

**If the dataset is in compliance with the IFPRI policy, follow the guidelines below for submitting the dataset to CKM Division at** IFPRI-Library@CGIAR.ORG**.**

**Guidelines for Datasets**: Researchers need to furnish

1. **Metadata Template for Dataset**

|  |  |
| --- | --- |
| Title  | Full title by which dataset is known |
| Related Publication | Publications that use the data from this dataset |
| Description  | A summary describing the purpose, nature, and scope of the dataset (No word limit. we suggest to keep it to the maximum of two short paragraphs) |
| Contributor | The organization or person responsible for either collecting, managing, or otherwise contributing in some form to the development of the resource.  |
| Related Datasets | Any datasets that are related to this dataset, such as previous research on this subject |
| Production Date | Date when the data collection or other materials were produced (not distributed, published or archived) |
| Producer  | Person or organization with the financial or administrative responsibility over the dataset |
| Collaborative organizations  |  |
| Funding organizations/sources | Grant number and related acknowledgements (if available) |
| Distributor | The organization designated by the author or producer to generate copies of the particular work (dataset) including and necessary editions or revisions |
| Time period covered  | Time period to which the data refer. This reflects the time period covered by the data, not the dates the data were collected |
| Date of collection | Contains the date(s) when the data were collected |
| Geographic coverage | Information on the geographic coverage of the data. Includes the total geographic scope of the data |
| Geographic Unit | Lowest level of geographic aggregation covered by the dataset, e.g. village, county, region |
| Universe | Description of the population covered by the data in the file the group of people or other elements that are the object of the study and to which the study result refer. Age, nationality, and residence commonly help to delineate a given universe. |
| Unit of Analysis | Basic unit of analysis or observation that this dataset describes, such as individuals, families/households, institutions or organizations |
| Sample and sampling procedure | Type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction |
| Kind of data | Type of data included in the file: survey data, census data, aggregate data, process produced data, time budget diaries |

1. **Data files**
2. **Format: STATA,** SPSS, EXCEL, ACCESS and others
3. **Raw Data.** Ideally, data should be "raw" with any new variables that have been created that do not correspond specifically to a survey question dropped. If there are “constructed” variables included, details should be provided on how the variable was computed
4. **Confidentiality.** Any variables that will allow specific persons/households to be identified (i.e. name, address, phone number) should be removed**.**
5. **Documentation: Save all your documents in Dropbox**
6. Codebook: The document should (again, ideally) be submitted electronically in Word, WordPerfect, or Portable Document Format (PDF) files. The codebook should provide a list of variable names, variable labels, and label values.It should specify the data position of each variable, describe the contents of each variable, and identify the range of possible codes and the meanings of those codes
7. Data Collection Instrument:
8. One clean, unused copy of each instrument, if appropriate, including interview schedules, self-administered questionnaires, data collection forms for transcribing information from records, paper tests and scales, screening forms, and call-report forms, should be included along with a description of the circumstances in which each was used (study populations, time periods, etc.)
9. **If questionnaires were administered in a language other than English, it would be great to have both the originals and an English version. Also if available, an annotated questionnaire, i.e. a questionnaire with variable names as they appear in the dataset filled in the spaces corresponding to the appropriate questions, is preferred**
10. **If available, provide** a copy of the final project report, project summary, or other description of the project, bibliography of publications pertaining to data, and appropriate summary statistics (frequency distributions, means, etc.)
11. **Related Datasets: If any, provide names of the datasets and web links that are related to this dataset**
12. **Related Publications: If any, provide citations for publications that are related to this dataset. For information on how to cite, please visit** <http://library.ifpri.info/files/2015/07/IFPRI-Style-Manual_July2015.pdf>
13. **Related Photos/ Videos: If available, provide photos/ videos that are related to this dataset**
14. **Permissions: If collaborators are involved,** permission from the partners, who own the data and have their own release policy, is required before making the datasets public. See the [Sample Permission Letter](http://intranet.ifpri.org:8090/COM/COM%20General%20FILES/Sample%20Permission%20Letter_May_12_2009_0.docx) on the Data page
15. Data Files, Description, and Documentation: All the files info, description, and documentation related to datasets will be stored at P:/Drive. Once the dataset is made as public goods it is available to everyone from [IFPRI Datasets](https://thedata.harvard.edu/dvn/dv/IFPRI)