



Getting data into the repository: things to ponder

Indira Yerramareddy, Manager of Knowledge Management and Web
Nilam Prasai, Senior Data Curator

International Food Policy Research Institute (IFPRI)

May 16, 2019

Research data management

- Care and maintenance of the data that is produced during the research
- Describes collection, organization, storage, documentation, dissemination and preservation of research data
- Defines the quality and after-life of research data
- Should be considered in each and every steps of research

Research data lifecycle



Based on UK Data Archive Lifecycle

Fig. Research Data Lifecycle

FAIR principles (what)

- set of guiding principles for scientific data management and stewardship to ensure that data are
 - FINDABLE (F)
 - ACCESSIBLE (A)
 - INTEROPERABLE (I)
 - REUSABLE (R)
- developed and endorsed by researchers, publishers, funding agencies and others
- put emphasis on machine-actionability on datasets

FAIR principles (why)

- help to increase the visibility of research
- help to save time and resources by minimizing duplicate data collection and research
- help to innovate and discover by providing data needed to answer research questions
- help to attain maximum potential of any data assets
- help to achieve maximum impact of research
- help to improve the reproducibility of research
- help to follow international standards, protocols and best practices

FAIR principles (how)

Findable

- Metadata
- Persistent identifier
- Repository

Accessible

- Machine and human readable metadata and data
- Long term storage and preservation

Interoperable

- Metadata schema
- Standard vocabulary and ontology
- Interoperable repository

Reusable

- Data use license
- Metadata with accurate and relevant attributes

<https://www.force11.org/group/fairgroup/fairprinciples>

Key things to ponder when publishing research data

- Informed consent
- Data anonymization
- Intellectual property, ownership and copyrights
- Contractual and legal compliance
- Data files, documentations and metadata

Do I have informed consent to share data?

- Scientists are required to protect the confidentiality of the respondents
- Language in informed consent can limit scientist's ability to share data
- Use language in informed consent: such as “We may share de-identified data we collect from you...”
- Identified and sensitive datasets can be shared only under restrictive settings

Is the data completely deidentified?

- Should not compromise confidentiality/privacy of respondents
- Direct and indirect identifiers should not be present in datasets that are openly public
- Controlled sharing should be done for datasets that can't be deidentified completely
- Always use non-disclosure agreement while sharing identified datasets

Who owns the data? Intellectual property and copyrights

- Confirm who has the rights and ownership to the data
- Confirm what intellectual property and copyrights applies to data
- Confirm if you plan to apply patent for the dataset, data collection methodology etc

Is there any legal and contractual provision related to data publishing?

- Provisions in the contract
 - Data security, protection, retention, transfer, sharing
 - Confirm if permission is needed from collaborators for data publishing
- Legal regulations
 - Data security, protection, retention, transfer, sharing

Preparing data for publishing

- Preparing data files
- Preparing documentations
- Preparing metadata

Preparing data files

- File format
- File name
- Variable name
- Data values and codes
- Special types of data
- Anonymization

Preparing documentation: key elements to consider

- Context
- Data collection methods
- Structure of data
- Quality controls
- Versions
- Access and use conditions

Preparing documentation: key documents to include

- Codebook(s)
- Data collection instruments
- Sampling methodology
- Interviewer manual
- Summary report
- Working papers
- Project reports and publications

Preparing metadata: minimum metadata elements

- Title
- Authors/Contributors
- Description
- Funder
- Keywords
- Topics
- Sample and sampling procedure
- Data collection date
- Time period covered in the datasets
- Geographic coverage
- Geographic unit
- Producer
- Production place
- Distributor

Preparing metadata: optional metadata elements

- Related publications
- Related datasets
- Universe
- Unit of analysis
- Time method
- Data collector
- Collector training
- Frequency
- Target sample size
- Major deviations from the sample design
- Data collection mode
- Types of research instrument
- Characteristics of data collection situation
- Weighting
- Response rate
- Error notes
- Estimates of sampling error
- Cleaning operations

Preparing metadata: don't forget

- Persistent identifier
- Terms of use
- Data license
- Access conditions (if any)

Data citation and Metadata Example from IFPRI Dataverse

Harvard Dataverse > International Food Policy Research Institute (IFPRI) Dataverse > Malawi Agriculture and Food Security Policy Processes Endline Survey, 2017/18

Metrics

3 Downloads

Contact Share

Link Edit

Malawi Agriculture and Food Security Policy Processes Endline Survey, 2017/18 Version 1.0

International Food Policy Research Institute (IFPRI), 2019, "Malawi Agriculture and Food Security Policy Processes Endline Survey, 2017/18", <https://doi.org/10.7910/DVN/9PQCET>, Harvard Dataverse, V1, UNF:6:hloWvrxXk613FFCwhAqemw== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

Description

Several initiatives in Malawi have sought to strengthen the processes through which the design and content of policies, strategies,

Files

Metadata

Terms

Versions

Add + Edit Metadata

Export Metadata

Citation Metadata

Dataset Persistent ID

doi:10.7910/DVN/9PQCET

Publication Date

2019-03-05

Title

Malawi Agriculture and Food Security Policy Processes Endline Survey, 2017/18

Author

International Food Policy Research Institute (IFPRI)

Contact

Use email button above to contact.

IFPRI-Data (International Food Policy Research Institute (IFPRI))

Description

Several initiatives in Malawi have sought to strengthen the processes through which the design and content of policies, strategies, and programs in the agriculture sector that affect the nation's food security are established. A two-part study was done under the New Alliance Policy Acceleration Support: Malawi Project (NAPAS: Malawi) to assess the quality of these policy processes and the institutional framework through which they are conducted and how perceptions of their quality have changed over time. The study is based on a two-round survey of national stakeholders in Malawi on issues centered on agriculture or food security that was conducted in 2015 and 2017/18.

This study contains data from the endline survey. The endline survey was conducted in late 2017 and early 2018. The 86 who made up the analytical sample for the baseline survey were contacted again and asked to complete an online question

Feedback

Thank you: Questions and comments(?)