Opening Access to Agricultural Microdata

1 – Microdata dissemination programs

Regional webinar - Africa
10:00 to 13:00 (GMT+0), November 29, 2021
Objective for Today

• Provide in-depth introduction to the key components of a microdata dissemination program:
  • The access model, access policy and terms of use,
  • Microdata documentation (including the DDI-standard),
  • Microdata anonymization or Statistical Disclosure Control (SDC)
  • the National Data Archive (NADA) as data cataloguing tool.

• Discuss how these components interlace and if approached jointly may offer solutions to ensure confidentiality safeguards are in place as well as to ensure microdata dissemination programs align on best international standards.
# Program for Today

<table>
<thead>
<tr>
<th>Day</th>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Opening Introduction to survey microdata dissemination: Benefits, costs and risks</td>
<td>FAO/Christophe Duhamel (ESS-FAO)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marie Vander Donckt (ESS - FAO)</td>
</tr>
<tr>
<td>2</td>
<td>Defining microdata access policies and terms of use</td>
<td>Marie Vander Donckt (ESS - FAO)</td>
</tr>
<tr>
<td>3</td>
<td>Statistical Disclosure Control (SDC)</td>
<td>Amsata Niang (ESS - FAO)</td>
</tr>
<tr>
<td>4</td>
<td>Microdata documentation and the DID standard</td>
<td>Gaye Parcon, International expert on microdata documentation and DDI standard</td>
</tr>
<tr>
<td>5</td>
<td>Archiving and Cataloguing Tool: The National Data Archive (NADA)</td>
<td>Somnath Sambhudas (ESS - FAO)</td>
</tr>
<tr>
<td>6</td>
<td>The FAM Catalogue</td>
<td>Yakob Seid (OCS – FAO)</td>
</tr>
<tr>
<td>7</td>
<td>Discussions and Closing</td>
<td>All/FAO</td>
</tr>
</tbody>
</table>
Outline

1. Before we start - The AGRISurvey Program and 50x2030 Initiative
2. Defining microdata
3. Benefits of microdata dissemination
4. Costs and risks of microdata dissemination
5. Enabling environment: legal, ethical, and technical considerations
6. Tools, guidelines, and good practice
7. AGRISurvey – A multidimensional approach to microdata dissemination
Overview of AGRISurvey

**Objective:**
Accelerate and improve the collection and dissemination of agricultural data.

**Format:**
Training and technical assistance program geared towards the implementation of an integrated system of agricultural surveys (i.e., the AGRIS model).

**Methodology:**
The AGRIS method was developed in 2016-2017 under the framework of the Global Strategy to Improve Agricultural and Rural Statistics.

**Financial Support:**
Bill and Melinda Gates Foundation
USAID
Accelerate and improve the collection, dissemination and use of agricultural data.

AGRISurvey also strives to:

- Strengthen data dissemination programs, including micro-data
  
  → **Data Dissemination component**

- Foster the effective use of agricultural data for decision-making
  
  → **Data Use component**

*AGRISurvey now part of the global 50x2030 Initiative!*
Agricultural Integrated Surveys – AGRISurvey - Program


50x2030 Initiative to Close the Agricultural Data Gap

Visit: https://www.50x2030.org/
What is Microdata?

Microdata are individual record or respondent data. It is not aggregate data or summary data.

Microdata are data that hold information collected on individual units, such as people, households or enterprises.

A set of records containing information on individual respondents or on economic entities. Such records may contain responses to a survey questionnaire or administrative forms.
Benefits of microdata dissemination

• Support and diversify research work (tabular data only answer predefined questions)
  • Maximize use and usefulness of data
  • Increased return on investments; more use → funding
• Reduced duplication of data collection
• Improved data quality, higher credibility
  • Relevance, reliability through feedback from users
  • Transparency and replicability as scientific safeguards
• Satisfy a legal requirement, and/or a contractual obligation with sponsors
Costs and risks of microdata dissemination

- Confidentiality / privacy protection are key to maintain trust from respondents
- Need to preserve reputation
- Need to comply with the law
- Exposure to criticism
- Non-official estimates vs. official estimates
- May have to respond to inquiries
- Financial cost (data preparation, anonymization, dissemination)
- Building and maintaining technical expertise
Enabling conditions: legal, ethical

- Fundamental Principles of Official Statistics
  - “Making available microdata for research is not in contradiction with the sixth UN Fundamental Principle as long as it is not possible to identify data referring to an individual.”
  - Provision of microdata must comply with legal and other arrangements that ensure confidentiality of the released microdata.
  - The procedures for access to microdata must be transparent and publicly available “to increase public confidence that microdata are being used appropriately and to show that decisions about microdata release are taken on an objective basis.”
African charter on official statistics

Principle 4: Dissemination

- **Accessibility:** African statistics shall not be made inaccessible in any way whatsoever. This concomitant right of access for all users without restriction shall be guaranteed by domestic law. Micro-data may be made available to users on condition that the pertinent laws and procedures are respected and confidentiality is maintained;

- **Dialogue with users:** Mechanisms for consultation with all African statistics users without discrimination shall be put in place with a view to ensuring that the statistical information offered are commensurate with their needs;

- **Clarity and understanding:** Statistics shall be presented in a clear and comprehensible form. They shall be disseminated in a practical and appropriate manner, be available and accessible to all and accompanied by the requisite metadata and analytical commentaries;

- **Simultaneity:** African Statistics shall be disseminated in a manner that ensures that all users are able to use them simultaneously. Where certain authorities receive advance information under embargo, to allow them time to respond to possible questions, public announcement shall be made indicating the nature of such information, the identity of the recipients and the set timeframe before its public dissemination;

- **Correction:** Statistics authorities shall correct publications containing significant errors using standard statistical practices or, for very serious cases, suspend dissemination of such statistics. In that event, the users shall be informed in clear terms of the reasons for such corrections or suspension.
Enabling conditions: technical

- Protocols and procedures (transparent policy)
- Detailed metadata (to ensure usability of data)
- Catalog (to support discoverability and access)
- Staff skills (in data curation, anonymization)
- Manuals, training, tools are available
Microdata Dissemination

• Documentation, including DDI-metadata
  • DDI metadata standard and editor training

• Catalog (preservation and release)
  • Support for upgrade to or setting up of NADA 5
  • Catalogue management

• Microdata access policy
  • Drafting microdata access policy (release type, terms of use, contract)
  • Establishing institutional protocol and workflow to handle access requests

• Anonymization or Statistical Disclosure Control (SDC)
  • Anonymization training
  • Technical support in anonymizing AGRISurvey microdata and preparing SDC report
Generic Activity Model for Statistical Organizations (GAMSO)
### Generic Statistical Business Process Model (GSBPM)

<table>
<thead>
<tr>
<th>Overarching Processes</th>
<th>Specify needs</th>
<th>Design</th>
<th>Build</th>
<th>Collect</th>
<th>Process</th>
<th>Analyse</th>
<th>Disseminate</th>
<th>Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1.1</td>
<td>2.1</td>
<td>3.1</td>
<td>4.1</td>
<td>5.1</td>
<td>6.1</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identify needs</td>
<td>Design outputs</td>
<td>Build collection instruments</td>
<td>Collect Cause frame and select sample</td>
<td>Process Integrate data</td>
<td>Analyse Prepare draft outputs</td>
<td>Disseminate Update output systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2</td>
<td>2.2</td>
<td>3.2</td>
<td>4.2</td>
<td>5.2</td>
<td>6.2</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consult and confirm needs</td>
<td>Design variable descriptions</td>
<td>Build or build processing and analysis components</td>
<td>Collect Set up collection</td>
<td>Process Classify and code</td>
<td>Analyse Validate outputs</td>
<td>Disseminate Produce dissemination products</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.3</td>
<td>2.3</td>
<td>3.3</td>
<td>4.3</td>
<td>5.3</td>
<td>6.3</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Establish output objectives</td>
<td>Collect aggregation components</td>
<td>Collect Run collection</td>
<td>Process Review and validate</td>
<td>Analyse Interpret and explain outputs</td>
<td>Disseminate Manage release of dissemination products</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4</td>
<td>2.4</td>
<td>3.4</td>
<td>4.4</td>
<td>5.4</td>
<td>6.4</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identify concepts</td>
<td>Build frame and sample</td>
<td>Collect Configure workflows</td>
<td>Process Finalise collection</td>
<td>Analyse Edit and update</td>
<td>Disseminate Agree an action plan</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5</td>
<td>2.5</td>
<td>3.5</td>
<td>4.5</td>
<td>5.5</td>
<td>6.5</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Check data quality</td>
<td>Design processing and analysis</td>
<td>Collect Test production systems</td>
<td>Process Apply disclosure control</td>
<td>Analyse Devise new variables and units</td>
<td>Disseminate Manage user support</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.6</td>
<td>2.6</td>
<td>3.6</td>
<td>4.6</td>
<td>5.6</td>
<td>6.6</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design production systems and workflow</td>
<td>Design processing and analysis</td>
<td>Develop statistical business process</td>
<td>Collect Test statistical business process</td>
<td>Analyse Calculate weights</td>
<td>Disseminate Manage user support</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7</td>
<td>2.7</td>
<td>3.7</td>
<td>4.7</td>
<td>5.7</td>
<td>6.7</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Finalise production systems</td>
<td>Design production systems and workflow</td>
<td>Finalise statistical business process</td>
<td>Collect Calculate aggregates</td>
<td>Analyse Finalise data files</td>
<td>Disseminate Manage user support</td>
<td></td>
</tr>
</tbody>
</table>

**Quality management**

**Data management**

**Metadata management**
Opening Access to Agricultural Microdata

2 – Microdata Access Models

Regional webinar - Africa
10:00 to 13:00 (GMT+0), November 29, 2021
1. Modalities of microdata dissemination
2. Microdata Access Policy
3. Procedures and Conditions on the Release of Microdata Files
4. Types of Microdata Files
5. Access policy – Public Use Files (PUFs)
6. Access policy – Licensed Files/Scientific Use Files (SUFs)
7. Access policy – Data enclave
Modalities of microdata dissemination

• Data enclave / research data centers → Expensive, inconvenient
• Remote access
  • Remote submission → Slow, rigid, inconvenient to users, can be expensive to implement
  • On-line analysis → Technical limitations, software can be expensive
• Sharing microdata files
  • Scientific Use Files (SUF) → Anonymized microdata (with lower risk-aversion than PUF), confidentiality pledge, terms for use
  • Public use files (PUF) → Microdata anonymized for low disclosure risk

• Release type influences disclosure scenarios (different users have different data available)
• One dataset can be released in different release modes
• Only single file per release mode
Microdata Dissemination  
- Microdata Access Policy -

**Microdata access policy:** set of access modalities and conditions holding for all microdata datasets released by an Agency.

The microdata access policy should specify:

- **The audience:** who can access microdata;
- **The conditions:** what conditions need to be fulfilled to access the data;
- **The access model:** PUF, SUF, remote access, on-site access
- **The application protocol and a clear set of rules:** who decides on access applications and contract template;
- **The sanction and enforcement mechanism** in case of breach of confidentiality.
Dissemination Policy

Components of a dissemination policy:

• Objective of the policy: nature of microdata files to be released; and intended use and conditions of release
• Statement on the reason for microdata dissemination
• Legal basis for microdata dissemination
• Principles applied to preserve privacy and confidentiality of data
• Statement on the timing of data release
• Statement on the pricing policy
• Key roles and responsibilities for the definition and implementation of the policy
Procedures and Conditions on the Release of Microdata Files

- How users can request access to the data
- Permissions and restrictions attached to the various types of datasets
- Who is responsible for decisions to grant access and other information on the review process
- Statistical disclosure control methods in place
- Information required from the researchers, and what can be done with the information
- Detailed pricing policy
- Technical support available to users
- Other practical information
Types of Microdata Files

• **Public-Use Files (PUFs)** – open to anyone with access to an NSO website

• **Licensed Files or Scientific-Use Files (SUFs)** – terms and conditions must include the basic common principles plus some additional ones applying to the researcher’s organization

• **Files in Data Enclaves** – particularly sensitive data or for more detailed data for which sufficient anonymization to release them outside the NSO premises is not possible
Access Policy
Conditions set by the NSO on the use of the microdata files

Public-Use Files (PUF)

• Include statements defining suitable uses for and precautions to be adopted in using the data.
• These are not legally binding but they serve to sensitize the user.
• Prohibitions such as attempts to link the data to other sources can be part of the “use statement” to which the user must agree, on-line, before the data can be downloaded.
• User must register in the system. This is done to enable users to be notified about changes to files and availability of new ones.
Access Policy
Conditions set by the NSO on the use of the microdata files

Public-Use Files (PUF)

Global Financial Inclusion (Global Findex) Database 2017
Afghanistan, 2017
Development Research Group, Finance and Private Sector Development Unit
Collection: Global Financial Inclusion (Global Findex) Database

* Requires user registration
Access Policy
Conditions set by the NSO on the use of the microdata files

Public-Use Files (PUF)

Access policy

CONTACTS

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Email</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development Research Group</td>
<td>World Bank</td>
<td><a href="mailto:research@worldbank.org">research@worldbank.org</a></td>
<td><a href="http://go.worldbank.org/B9W4QTDHR0">http://go.worldbank.org/B9W4QTDHR0</a></td>
</tr>
</tbody>
</table>

CITATION REQUIREMENTS

Use of the dataset must be acknowledged using a citation which would include:
- the identification of the Primary Investigator
- the title of the survey (including country, acronym and year of implementation)
- the survey reference number
- the source and date of download

Example:

Access Policy
Conditions set by the NSO on the use of the microdata files

Public-Use Files (PUF)

* Requires user registration
Access Policy

Conditions set by the NSO on the use of the microdata files

Public-Use Files (PUF)
Access Policy
Conditions set by the NSO on the use of the microdata files

Licensed Files / SUFs
Data are provided to a researcher or a team for a specific purpose
• Interested users must complete a formal request to access the data.
• User will specify that the files will not be shared outside the organization and that data will be stored securely.
• To the extent possible, user must identify:
  • Intended use of data
  • List of expected outputs
  • Organization’s dissemination policy
• There is a legally-registered sponsoring agency
Access Policy

Licensed Files / SUFs

Application for Access to a Licensed Dataset

First Name: Gaye
Last Name: Parcon
Organization: gayeparcon@gmail.com


This form must be filled and submitted by the Lead Researcher. Lead Researcher refers to the person involving this agreement. Access to licensed datasets will only be granted when the Lead Researcher is a company, research centre, national or international organization, etc., on behalf of which access is to be granted for compliance with all terms of this Data Access Agreement by employees of the receiving organization.

This request will be reviewed by a data release committee, who may decide to approve the request for the Lead Researcher. A signed copy of this request form may also be requested.

Receiving organization name:

Telephone (with country code):

Intended use of the data:
Please provide a short description of your research project (project question, objectives, methods, expected outcomes, etc.)

Data and other material provided by the Microdata Library will be used solely by the user and shall not be redistributed or sold to other individuals, institutions or organizations.
Access Policy
Conditions set by the NSO on the use of the microdata files

Licensed Files / SUFs
Blanket agreement to an organization
• The license should ensure compliance with a named individual assuming responsibility for this
• Each additional user must be made aware of the terms and conditions that apply to data files – have them sign an affidavit

Refer to ISHN Working Paper No.5 for Examples of Access Conditions
Access Policy

Conditions set by the NSO on the use of the microdata files

Data Enclave

• May be located at the NSO headquarters or in major centers such as universities close to the research community.
• Used to give researchers access to complete data files without the risk of releasing confidential data.
• Staff supervise access and use of the data.
• The computers must not be able to communicate outside the enclave.
• Results obtained by the researchers must be screened for confidentiality before taken out.
Access Policy
Conditions set by the NSO on the use of the microdata files

Data Enclave

Data Available in an Enclave: under this policy **no data is shared through the application**. Users submit an application to access the data at a secure facility physically located on the premises of the data producer or a facility nominated by the data producer.
Thank You

References:

