Improving access to and use of microdata from agricultural censuses and surveys

What role for FAO at national and global levels?

Francois Fonteneau
Statistics Division, FAO
Content

1. Rationale for improvement
2. How? Examples of what the IHSN and ADP are doing
   1. Focus: metadata
   2. Focus: confidentiality + terms of use
3. Examples from agencies: it can work
4. A possible role for FAO
Rationale for improvement
Rationale for improvement
Rationale for improvement

Complaints from users:

• “we didn’t know they had these data”   Visibility
• “they don’t want to share them”   Accessibility
• “these data are messy”   Reliability
• “this is not really what I need”   Relevance
• “how can I assess progress?”   Comparability
• “where is the documentation?”   Usability
• “there is no transparency”   Quality, credibility
Rationale for improvement

Issues faced by producers:

• “we don’t know how to deal with microdata”
  – can we disseminate them in our country?
    • If YES, how to disseminate them?
    • To whom?
    • What costs?
  – how to document them?
  – how to improve transparency on sources?

• “what do others do?”
  – where is the good practice?
Why should we improve?

1. To support research
2. To enhance the credibility of official statistics
3. To reduce duplication in data collection
4. To increase return on investment and leveraging funding for statistics
5. To comply with a contractual or legal obligation

The obligation to disseminate microdata does not conflict with the obligation to maintain confidentiality and privacy but conflicts with the aim of CD projects for surveys and censuses, when these issues are ignored!
Why it does not happen?

• It is actually happening in some countries, but still many obstacles remain, such as:
  – Legal issues: apply the law ... or change it
  – Confidentiality: maintaining respondents’ trust
    “Individual data collected by statistical agencies for statistical compilation, whether or not they refer to natural or legal persons are to be strictly confidential and used exclusively for statistical purposes.” UN 6th principle for off stat.
  – Other technical issues
  – Exposure to contradiction
    • Quality of microdata
    • Discrepancies with aggregated, published official statistics
  – Loss of exclusivity
  – Cost and technical capacity
IHSN: sharing solutions

- **IHSN Survey Catalogue**
  - 3,100+ surveys described
  - Link to access policies + datasets

- **Software Tools**
  - Microdata Management Toolkit
  - Microdata Anonymization Toolkit
  - National Data Archive (NADA)

- **Guidelines**
  - Long term preservation (ICPSR)
  - Design of access policy
ADP: bringing solutions to countries

• More survey data accessible and used
  ✓ 40 national survey catalogues established
  ✓ Use monitored
  ✓ Indicators re-calculated (Nigeria)

• More national capacity on survey data management
  ✓ 66 countries received TA
  ✓ 20 more countries involved in regional activities
    • Joint workshops with FAO

• Survey methods / instruments improved
  ✓ pilot phase with CSA-Ethiopia
1. For discovering data
   - You and your users will know about data availability by providing searchable data catalogs

2. For making data usable
   - Users need to fully understand the data: why, by whom, when, and how they were collected, processed and disseminated

3. For assessing data quality

=> metadata standards for microdata: DDI, etc.
GSBPM, DDI, SDMX

<table>
<thead>
<tr>
<th>Quality Management / Metadata Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Specify Needs</td>
</tr>
<tr>
<td>1.1 Determine needs for information</td>
</tr>
<tr>
<td>1.2 Consult &amp; confirm needs</td>
</tr>
<tr>
<td>1.3 Establish output objectives</td>
</tr>
<tr>
<td>1.4 Identify concepts</td>
</tr>
<tr>
<td>1.5 Check data availability</td>
</tr>
<tr>
<td>1.6 Prepare business case</td>
</tr>
<tr>
<td>2 Design</td>
</tr>
<tr>
<td>2.1 Design outputs</td>
</tr>
<tr>
<td>2.2 Design variable descriptions</td>
</tr>
<tr>
<td>2.3 Design data collection methodology</td>
</tr>
<tr>
<td>2.4 Design frame &amp; sample methodology</td>
</tr>
<tr>
<td>2.5 Design statistical processing</td>
</tr>
<tr>
<td>2.6 Design production systems &amp; workflow</td>
</tr>
<tr>
<td>3 Build</td>
</tr>
<tr>
<td>3.1 Build data collection instrument</td>
</tr>
<tr>
<td>3.2 Build or enhance process components</td>
</tr>
<tr>
<td>3.3 Unify workfiles</td>
</tr>
<tr>
<td>3.4 Test production system</td>
</tr>
<tr>
<td>3.5 Test statistical business process</td>
</tr>
<tr>
<td>3.6 Finalize production system</td>
</tr>
<tr>
<td>4 Collect</td>
</tr>
<tr>
<td>4.1 Select sample</td>
</tr>
<tr>
<td>4.2 Set up collection</td>
</tr>
<tr>
<td>4.3 Run collection</td>
</tr>
<tr>
<td>4.4 Finalize collection</td>
</tr>
<tr>
<td>5 Process</td>
</tr>
<tr>
<td>5.1 Integrate data</td>
</tr>
<tr>
<td>5.2 Classify &amp; code</td>
</tr>
<tr>
<td>5.3 Review, Validate &amp; edit</td>
</tr>
<tr>
<td>5.4 Impute</td>
</tr>
<tr>
<td>5.5 Derive new variables &amp; statistical units</td>
</tr>
<tr>
<td>5.6 Calculate weights</td>
</tr>
<tr>
<td>5.7 Calculate aggregates</td>
</tr>
<tr>
<td>5.8 Finalize data files</td>
</tr>
<tr>
<td>5.9 Finalize</td>
</tr>
<tr>
<td>6 Analyse</td>
</tr>
<tr>
<td>6.1 Prepare draft outputs</td>
</tr>
<tr>
<td>6.2 Validate outputs</td>
</tr>
<tr>
<td>6.3 Scrutinize &amp; value</td>
</tr>
<tr>
<td>6.4 Apply disclosure control</td>
</tr>
<tr>
<td>6.5 Finalize outputs</td>
</tr>
<tr>
<td>6.6 Finalize</td>
</tr>
<tr>
<td>6.7 Promote dissemination products</td>
</tr>
<tr>
<td>6.8 Disseminate</td>
</tr>
<tr>
<td>6.9 Manage user support</td>
</tr>
<tr>
<td>7 Disseminate</td>
</tr>
<tr>
<td>7.1 Update output systems</td>
</tr>
<tr>
<td>7.2 Produce dissemination products</td>
</tr>
<tr>
<td>7.3 Approve dissemination products</td>
</tr>
<tr>
<td>7.4 Promote dissemination products</td>
</tr>
<tr>
<td>7.5 Manage user support</td>
</tr>
<tr>
<td>8 Archive</td>
</tr>
<tr>
<td>8.1 Define archive rules</td>
</tr>
<tr>
<td>8.2 Manage archive repository</td>
</tr>
<tr>
<td>8.3 Preserve data and associated metadata</td>
</tr>
<tr>
<td>8.4 Dispose of data &amp; associated metadata</td>
</tr>
<tr>
<td>9 Evaluate</td>
</tr>
<tr>
<td>9.1 Gather evaluation inputs</td>
</tr>
<tr>
<td>9.2 Conduct evaluation</td>
</tr>
<tr>
<td>9.3 Agree action plan</td>
</tr>
</tbody>
</table>

From UNECE
focus: IHSN toolkit

1. IHSN Microdata Management Toolkit (DDI metadata editor)
2. NADA: on line searchable survey cataloguing system
3. IHSN Microdata Anonymization Toolbox (R-based sdcMicro)

Risk measurement / risk reduction / information loss
focus: confidentiality

- Microdata does not mean confidential data
- Different access modes
  - Remote execution
  - On-site
  - File distribution
- Screen users and usage
- Progressive approach
focus: confidentiality

- Reports, tables (PDF)
- Indicators
- On-line interactive tabulation, analysis and mapping tools
- Microdata (n% sample) and detailed metadata

What outputs?
What tools?

No issue of confidentiality
Disclosure risk managed by the application
Requires data anonymization

Any web development tool
CountrySTAT/DevInfo/Prog noz, etc.
REDATAM/IMIS, SuperStar, Nesstar, etc.
IHSN MMT + NADA
Formulating an **access policy** and procedures

---

Central Data Catalog

Found 33 studies out of 33

Sort results by: Year | Title | Popularity

Showing 1-15 of 33 studies

Aquaculture Production Survey 2009
2009
By: Bureau of Agricultural Statistics (BAS), Department of Agriculture

Avian Population Survey in the Philippines 2006, Inventory of Avian Species
2006
By: Bureau of Agricultural (BAS) - Department of Agriculture

PHI, MoA
Data Catalog

Show studies conducted between [1987] and [2010]

Find: in study description

Find: in variable description

Variable description includes: Name, Label, Question, Classification

Sort results by: Year | Title

Found 44 studies out of 44

Annual Survey of Industries 1997-98
By: Central Statistics Office (Industrial Statistics Wing), MOSPI, Government of India

Annual Survey of Industries 1998-99
By: Central Statistics Office (Industrial Statistics Wing), MOSPI, Government of India

Annual Survey of Industries 1999-2000
By: Central Statistics Office (Industrial Statistics Wing), MOSPI, Government of India

Annual Survey of Industries 2000-01
By: Central Statistics Office (Industrial Statistics Wing), MOSPI, Government of India
Data Catalog

Filter by Country

Find: vanuatu

Variable description includes: Name, Label, Question, Classification

Sort results by: Country ▲ | Year | Title

Found 4 studies out of 11

Vanuatu - Vanuatu National Population and Housing Census, Pilot Census 2008
Vanuatu National Statistics Office, Vanuatu Government, DDI-VAN-VNSO-NPHC-v1

- Browse metadata
- Access policy

Vanuatu - Vanuatu Agriculture Census 2006-2008

- Browse metadata
- Access policy

Vanuatu - Vanuatu Household Income and Expenditure Survey 2006

- Browse metadata
- Access policy

Vanuatu - Vanuatu Multiple Indicators Cluster Survey 2007
Ministry of Health, Government of Vanuatu, Ministry of Health, VUT-MICS-2007-1.0

- Browse metadata
- Access policy

Site Navigation

- Home
- New Releases
  - Trades Summary
  - Consumer Price Index (CPI)
  - Tourism
  - Quarterly Statistical Indicators
- Current Surveys
  - Census 2009
  - Census 2009 Tables
  - HIES 2010
- Data Catalog

LIVE POPULATION OF VANUATU
Total Live* Population of Vanuatu at 20:31:07

257,137
Annual Figures:
- Births: 7,911
- Deaths: 1,344

Population Clock Outside VNSO Building
A possible role for FAO

• We are committed to serve our clients: data users (internal and external) and data producers (internal and external)
  – Provision of financial and technical assistance to NSOs and MoAs
  – Facilitate access to microdata, central repository

• Leveraging on:
  – Global Strategy
  – World Census of Agriculture
  – Field projects
  – Membership in IHSN
Thank you!

more?

francois.fonteneau@fao.org
www.ihsn.org
Time is ripe for action

• “The fact that others are starting to create output we used to have a monopoly on, means that we are losing relevance. Does this mean that we be defensive about it and try to maintain our foothold by legislation etc.? We don't think so. We propose to actively pursue a course in which we use our strengths and resources to create the statistical outputs that will give our stakeholders a clear perspective on what is happening in the world today.”

(CES High-Level Group for the Modernisation of Statistical Production and Services,
http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+HLG)

• Lack of comparability between sources, lack of transparency = issue in the MDG measurement process.
  – Shaping the post-2015 agenda: need for greater access to sources, incl microdata

• DHS, MICS surveys or the IPUMS initiative provide a strong evidence of the large demand for Public-use microdatasets, and that a culture of data sharing and collaboration should result in a wealth of new knowledge.
how some NSOs describe their role on disseminating microdata

• **Australia**: “Access to microdata assists and encourages informed decision making through enabling wider use of ABS data for social and economic research and analysis. [...] An inability to meet these demands will increasingly become a disadvantage to ABS core business, the relevance of the ABS and ultimately to the coherence of the NSS. This, along with a number of other drivers for change including the growing risk of identification, has led the ABS to propose a new strategy for accessing ABS microdata into the future.” (ABS, [http://www.abs.gov.au/](http://www.abs.gov.au/))

• **Ireland**: “The primary objective of the CSO in providing access to microdata is to support the research community and to ensure that maximum usage is made of the data collected by the CSO. This approach supports the move towards evidence-based policy-making, has the potential to reduce the cost of research and also helps to avoid duplicate data collections.” (CSO - [http://www.cso.ie/](http://www.cso.ie/)).

• **Sri Lanka**: “The mission of the National Microdata Archive of Sri Lanka is to satisfy the data needs of the National and International research community who are striving hard to find answers to the socioeconomic problems across the world. Backed by the Department of Census and Statistics – the central statistical agency in Sri Lanka, LankaDatta disseminates relevant, reliable and up-to-date statistical information being produced by the Agencies in the National Statistical System of Sri Lanka while preserving strict confidentiality of the respondents.” (Department of Census and Statistics, Sri Lanka [http://statistics.sltidc.lk/](http://statistics.sltidc.lk/))