

Improving access to survey microdata in Asia and the Pacific

Seminar on Data Dissemination:
Emerging Trends and Issues
1-3 August 2007
Kuala Lumpur, Malaysia

Jan Smit

Overview

- The Project and the ADP
- Barriers to microdata dissemination
- Project assistance, services
- Expected achievements

The Project and the ADP

- Improving access to survey microdata in Asia and the Pacific project part of worldwide Accelerate Data Program
- <http://www.surveynetwork.org/adp/>
- Better data needed to monitor development outcomes (MDGs) and assess effectiveness of policies
- Improving survey programs in participating countries

Barriers to microdata access

- Data processing
- Documentation
- Archiving and storage
- Dissemination
- Confidentiality
- Note
 - Partial picture of some AP countries
 - Big differences between countries

Asia-Pacific microdata project overview

4
2 Aug 2007

Data processing

- Different tools used for data entry and editing & imputation across surveys and over time within NSOs
 - Especially when subject-matter units are responsible for data processing
 - Potential for internal learning
- Manual editing & coding labor-intensive and time-consuming
 - Effects timeliness and relevance of data

Asia-Pacific microdata project overview

5
2 Aug 2007

Documentation

- Some NSOs are developing, maintaining a database of census and survey metadata
 - But without using international standards
- Most, however, do not
- Quality and usefulness of survey data can be greatly enhanced through better documentation
 - Data only as good as the metadata

Asia-Pacific microdata project overview

6
2 Aug 2007

Archiving and storage

- No centralized data archives
 - Datasets floating around on PCs of individuals in subject-matter units
- Data formats
 - Older datasets stored on outdated, inaccessible formats (magnetic tapes)
 - Limited (or no) capacity to produce datasets (with metadata) in user-preferred formats

Asia-Pacific microdata project overview

7
2 Aug 2007

Dissemination

- Focus on production of paper versions of survey reports
 - Electronic copies of especially older surveys often not available
 - No systematic efforts to preserve and disseminate other metadata resources (questionnaires, manuals, programs)
- Legislation and policies for microdata dissemination lacking or vague

Asia-Pacific microdata project overview

8
2 Aug 2007

Confidentiality

- Inevitably cited as a key concern, barrier for the dissemination of microdata
 - Protection of the identity of respondents
 - Data are the “property” of NSO, government
- When data are disseminated, anonymization only through the removal of direct identifiers

Asia-Pacific microdata project overview

9
2 Aug 2007

Project support possibilities

- Data processing
- Documentation of surveys
- Anonymization
- Formulating microdata dissemination policy
- Establishment of a National Data Archive service
- Some hardware and software

Data processing

- Training and other assistance in the implementation of tools, such as:
 - Integrated survey processing software
 - CSPro (U.S. Census Bureau)
 - Blaise (Statistics Netherlands)
 - Dedicated editing & imputation software
 - Various tools developed by Statistics Italy and Statistics Canada

Documentation

- IHSN Microdata Management Toolkit
 - <http://www.surveynetwork.org/home/?lvl1=tools&lvl2=documentation&lvl3=toolkit>
- Set of tools for documentation, archiving, dissemination and preservation of microdata
- Assessment of quality of metadata generated for each dataset

Toolkit components

- Metadata Editor
 - Documents survey data in accordance with international standards
- CD-Rom Builder
 - Generates CD-Rom and web output
- Nesstar Explorer
 - For viewing metadata and re-exporting data to various formats

Anonymization

- Removing or modifying identifying variables contained in datasets
 - Describe characteristics of persons that are observable, registered (identification numbers, etc.), or otherwise can be known to other persons
- Direct identifiers
 - Names, addresses, ID card numbers
- Indirect identifiers

Indirect identifiers

- Characteristics that may be shared by several respondents, and whose combination could lead to the re-identification of one of them
 - E.g., the combination of variables district, age, sex and profession is identifying if only one individual of that particular sex, age and profession lived in that particular district

Anonymization tools

- μ -Argus software
 - EU-sponsored Computational Aspects of Statistical Confidentiality (CASC) project
 - Imports ASCII and SPSS files
- Stata plug-ins
 - Currently being developed by IHSN
- sdcMicro package for the R language
- Data Swapping Toolkit
 - National Institute of Statistical Sciences

Asia-Pacific microdata project overview

16
2 Aug 2007

Microdata dissemination policy

- International practices
- Different levels of accessibility
 - Public use datasets
 - Very low disclosure risk (highly anonymized), but info loss relatively high
 - Licensed datasets
 - Semi-anonymized; signed agreement between (trusted) user and provider
 - Datasets for use in data enclave only
 - Not anonymized; data not to be taken out

Asia-Pacific microdata project overview

17
2 Aug 2007

Dissemination of data

- Participation in the Project/ADP does not oblige countries to publicly disseminate micro datasets
 - Producer keeps full ownership of data and decides on most appropriate dissemination policy, within framework of national legislation
- However, objective is to foster better use of data and ADP therefore promotes wider accessibility to data

Asia-Pacific microdata project overview

18
2 Aug 2007

National Data Archive

- NADA: pre-packaged data archive website (< 50 surveys)
- Customized XML database + web interface solutions
- Training in the required IT skills

Achievements expected

- Documentation, in accordance with international standards, of censuses and household surveys after 1990
 - As many as possible, feasible
 - Incl. censuses and surveys produced by other organizations than NSO
- Documentation is treated as an ongoing part of survey activity
 - Not as last step of survey process

Achievements expected (cont.)

- Availability of census and survey metadata
 - Via websites of NSOs, ESCAP and IHSN
- Availability of quality, anonymized micro datasets to bona fide users
 - Via regional archive (at ESCAP)
 - Ultimate decision on access remains with countries, if they wish
 - Via national archives

Achievements expected (cont.)

- Procedures for access to and conditions to use microdata are publicly available
 - Via websites of NSOs (& other producers) and ESCAP
- Regional group of experts available to assist countries in census and household survey microdata management

Asia-Pacific microdata project overview

22
2 Aug 2007

Participating countries

- Practical assistance
 - Bangladesh, Fiji, Indonesia, Mongolia, Viet Nam, Sri Lanka (with UNICEF)
 - Several more to be decided
- 30 countries in region to participate in concluding workshop (2nd half 2008)

Asia-Pacific microdata project overview

23
2 Aug 2007

ADP contacts

- Asia-Pacific: ESCAP
 - Jan Smit, smit1@un.org, +6622881689
- Latin America and Caribbean: IADB
 - Jose Antonio Mejia, JOSEAM@iadb.org
- Elsewhere: World Bank
 - Olivier Dupriez, odupriez@worldbank.org

Asia-Pacific microdata project overview

24
2 Aug 2007
