Data archive in developing countries: preservation and dissemination of microdata as an instrument for better development results

Olivier Dupriez
Senior Economist / Statistician
The World Bank
odupriez@worldbank.org
About the World Bank

• Mandate: Millennium Development Goals

• International Development Agency (IDA): Loans and grants to 81 poorest countries, to support poverty reduction strategies (> $10 Bn per year)

• IDA14 replenishment (March 2005)
  ➔ Focus on effectiveness
  ➔ IDA Results Measurement System (input/outcomes)
  ➔ Marrakech Action Plan for Statistics
    ➔ Recommendation: International Household Survey Network
IHSN - Rationale

• In developing countries: sample surveys are a key source of data
  → needed for targeting, designing, monitoring and evaluating projects and policies

• Huge efforts to collect data, but issues of quality, timing, comparability, analytical capacity, accessibility by secondary users, and others
  → data remain under-exploited
IHSN - Mandate

IHSN created in September 2004 to:

– Coordinate donors’ activities (survey planning)
– Harmonize methods, promote standards
– Foster better use of data

requires access by secondary users

build a central survey repository
assist countries in preserving, documenting and disseminating microdata (Toolkit + Handbook)
Building a Central Survey Repository

- Started as a World Bank internal project. Now an IHSN activity
- Custom-made application (WB)
- Geographic coverage: developing countries
- Scope of collection: HH surveys / censuses / others
- > 2300 surveys/censuses listed (May 2005)
- > 600 datasets available
A web-based database of surveys, searchable by region/country, type of survey, year, etc. Advanced search (by topic) being developed.
For each survey, information is provided in four pages: description, content, documentation, and dataset.
Authorized users (identified at logon) can download data. Authorizations are granted by survey at individual level or to groups of users (e.g., all World Bank staff).

Non-authorized users must first request and obtain access.
Webservice available. Used by IHSN website, Health Metrics Network/WHO Country Logbook, and others to publicly disseminate metadata.
Survey Repository: Issues

**Acquisition**
- Many surveys poorly documented
- Much information lost by producers
- Datasets not anonymized / properly edited

**Management**
- Need a way to assess quality of datasets
- Ad-hoc / informal dissemination agreements
- Need a more formal collection, acquisition, and dissemination strategy
- Limited “search by topic” (no thesaurus)
- Users: demanding, not supplying!
Microdata Management Toolkit
Documentation / Dissemination / Preservation

• Objective: proper data archive in/by developing countries (in the past, technical assistance focused on data collection; lack of concern for preservation, documentation and dissemination)

• Toolkit: a user-friendly package based on best practice and international standards (specialized DDI/DC editor + CD-Builder)
A specialized DDI/DC metadata editor… (custom-branded Nesstar Publisher)

For more information on the Toolkit, see presentation by M. Diggory, O. Dupriez, P. Heus, J. Ryssevik (IASSIST 2005, Session E2)

... and a CD-ROM / Web Builder

CD-ROMs, DVDs

Websites

NESSTAR server: on-line analysis
Toolkit - Next steps

• Translate, document
• Roll-out program (training / technical assistance)
• Finalize a “Handbook on Microdata Documentation, Dissemination and Preservation”
  – Legal, ethical, technical and institutional guidelines to assist developing countries in establishing data archive services
  – Drafted (ICPSR contribution); more contributions needed!
• Further development and link to other tools (e.g., Argus, CsPro, DevInfo, others)
Contact, link and acknowledgments

Contact: Olivier Dupriez (World Bank / IHSN):
    odupriez@worldbank.org

IHSN website: www.surveynetwork.org

Acknowledgments:
- Pascal Heus
- Jostein Ryssevik and Nesstar Ltd.
- Mark Diggory (Harvard/MIT Data Center)