Building National Data Archives and Using Metadata in Developing Countries

IHSN Microdata Management Toolkit
2007 Update

Pascal Heus
International Household Survey Network
pascal.heus@gmail.com
http://www.surveynetwork.org
Microdata Management Toolkit Overview

1. Import data and compile metadata
   - Metadata Editor
   - Server
   - WebView

2. Import metadata and prepare CD-ROM
   - CD-ROM Builder

3. Generate HTML based CD-ROM

http://www.surveynetwork.org
Toolkit new developments

• Metadata Editor
  – Next release should coincide with Nesstar Publisher 4.0 (Summer)
  – Support for multilingual metadata
  – Support for Data Cubes
  – Fixes and minor enhancements

• CD-ROM Builder
  – HTML Editor
  – Outline validation
  – Fixes and minor enhancements

• Support for English, Spanish, French, Russian, Arabic
Toolkit deployment / training

• 25 countries in 2007, 50+ countries by end of 2008
• Africa:
  – Cameroon, Ethiopia, Gambia, Mozambique, Niger, Nigeria, Uganda,…
• Latin America:
  – In partnership with Inter-American Development Bank (IADB)
  – Honduras, Peru, …
• Asia/Pacific
  – In partnership with UN Economic and Social Commission for Asia and the Pacific (ESCAP)
• MICS 3 Surveys
New components under development

[Diagram showing relationships between components]

- Data Archive Web Site
  - Packaging
    - Stata
    - SPSS
    - SAS
  - ASCII
- Data Disclosure
  - Stata
    - C plug-in
- Quality Assurance
- Cataloging
  - Oracle
    - Berkeley DB
  - Tomcat
    - Java
What is CSPro (1)

- CSPro stands for Census and Survey Processing
- Public-domain software package for entering, editing, tabulating and mapping census and survey data.
- Joint effort from developers of IMPS and ISSA:
  - International Programs Center of the US Census Bureau
  - Macro International
  - Serpro, S.A.
- Funding:
  - US Agency for International Development (USAID)
What is CSPro (2)

• Features
  – File descriptions (data dictionary)
  – Data entry applications
  – Edit applications
  – Tabulations
  – Dissemination Products (Table Viewer, Table Retrieval System, Map Viewer)

• CSPro Utilities:
  – Sort Data File
  – Tabulate Frequencies
  – Text, Table, & Map Viewer
  – Concatenate Data Files
  – Reformat Data
  – Compare Data Files
  – Export Data (delimited text, SAS, SPSS, Stata)
  – Various Conversions (IMPS, ISSA)
What is CSPro (3)

- First release in May 2002, latest version is 3.2 (Feb 2007)
  - Expertise dates back to 70’s-80’s (CONCOR, IMPS, ISSA)
  - Mature product
- Widely used by national statistical agencies in developing countries
- More than 30,000 download from over 150 countries
- Free support provided by the US Census Bureau

http://www.surveynetwork.org
IHSN Toolkit / CSPro bridge using DDI

- **Integration**
  - Utility to convert the CSPro data dictionary (DCF) into a DDI document
  - Export data to fixed ASCII
  - Import in Metadata Editor

- **Status**
  - Currently beta-testing stand alone application
  - Integration in CSPro Exporter expected in release 3.4 (second half of 2007)
Other potential CSPro developments

- Table retrieval system now produces output in XHTML. Plan to add table level descriptive metadata (possibly SDMX based)
- Add survey level metadata in CSPro
- Include additional metadata such as weight variables, relationships and question text when available
- Take advantage of XML technologies to produce reports and other outputs
Data Disclosure / Anonymization

• Protection of the respondent
• Based on data reduction and perturbation techniques
  – Address discrete and continuous variables
  – Suppression, recoding, micro-aggregation, etc.
• Build open sources solutions:
  – Stand alone or integrate in as Stata plug-in (and possibly other packages)
  – High performance C code for complex algorithms
  – Build practical tools based on theoretical algorithms
• Development in progress / Open Source
Metadata Quality Assurance

• Having tools available is not sufficient, we need to ensure high quality content

• Toolkit Users Guide

• Drafting new guidelines:
  – Performing a National Inventory of Microdata and Metadata
  – Quick reference for Archivist: elements definitions and recommended content
  – IHSN DDI reviewers’ feedback

• Handbook on the Documentation, Dissemination, and Preservation of Microdata
  – Policy Recommendations and Technical Guidelines
National Data Archive (NADA) web site

- Specialized web site for national data archive
- Easy to use configure and maintain
- Fully customizable
- Web based content management system (CMS)
- Default recommended structure: home, mission, activities, policies & procedures, catalog, contacts
- Selection of “skins” for look and feel
- Open source
National Data Archive (NADA) web site

• **DDI based catalog**
  – Upload DDI documents (output of Toolkit)
  – Search in title, variable, documentation
  – Automated access request forms (public use, licensed datasets)
  – Limited number of surveys (no database)

• **Requires minimal infrastructure**
  – Web server with pHp 5, no database required

• **Availability June/July**

• **Advanced cataloging tools**
  – Support for large catalogs and more complex queries
  – Based on Oracle Berkeley DB-XML
  – Requires Java/Tomcat environment
  – Under development
Thank you!

Demos / Questions / Comments

For further information
http://www.surveynetwork.org
info@surveynetwork.org