The changing international statistical landscape - implications for local statisticians:
An Introduction to the Data Documentation Initiative (DDI)

Arofan Gregory
Metadata Technology NA
Background/History

• DDI is an XML standard coming from the social science data archival community
  – Originally grant-funded development more than 10 years ago
  – Now developed and maintained by a member-based consortia, the DDI Alliance (www.ddialliance.org)

• DDI has two major production streams:
  – DDI Codebook (the early versions, but still in production use)
  – DDI Lifecycle (the more recent versions, in increasing use)
DDI Codebook

- This version of DDI was produced to have good, processable (XML) documentation for individual studies
  - Good for search and dissemination
  - Good for documentation
  - Very detailed, down to the variable level
  - Also provides good study-level information
- Widely used in the developing world as a result of the International Household Survey Network’s (IHSN) “Metadata Management Toolkit”
  - More than 80 national statistical offices and growing
- In widespread use among the European and North American data archives
DDI Lifecycle

• Designed to describe more complex studies (logitudinal, repeat cross-sectional, etc.) and to provide for comparison among data sets
• Organized around a lifecycle model of statistical production
• Can describe surveys
• Can describe complex data files
• Useful for data production, as well as data dissemination, documentation, and re-use
Why Is DDI Important?

• If you work in regional and urban statistics, why should you care about DDI?

• Two reasons:
  – You may use data produced at the national level
  – You may produce data, and report it to the national level (or to other regions)
DDI for Data Users

- Rich information regarding your data sets, typically to the variable level
- Standard tools for producing online and print documentation of data sets
- Standard tools for producing formats for use with statistical packages (SPSS, SAS, Stata, etc.)
- Many tools are free/open source
DDI for Data Producers

• DDI – especially DDI Lifecycle – allows for the re-use of metadata throughout the data lifecycle
  – If you produce data on a regular basis, this will improve the accuracy and efficiency with which you produce and document your data
  – DDI-Lifecycle allows you to automatically generate data collection instruments across modes
  – DDI broadly allows for accurate, complete documentation at the variable level using software tools
• DDI documentation makes your data easier to discover and to re-use
National-Level Use of DDI

• As we mentioned above, DDI-Codebook is widely used in the developing world among statistical agencies

• DDI-Lifecycle is now becoming popular with statistical agencies in the developed world
  – Australian Bureau of Statistics, New Zealand, Canada, France, etc.
  – Eurostat is now becoming interested
The Broader Context

• National agencies worked together under the auspices of the UN/ECE Statistical Metadata Working Group (METIS) to create:
  – The Generic Statistical Business Process Model, GSBPM, a conceptual model of statistical production processes (based in part on the DDI lifecycle model)
  – Now, they are working on the Generic Statistical Information Model, GSIM, a conceptual information model

• This work is forming the basis of cooperation among statistical agencies at the national level
Conceptual and Implementation Models

• Conceptual models require implementation models to be useful
  – Implementation models are what we have inside our software tools and databases

• Two major implementation models are being considered for use with GSBPM/GSIM
  – DDI, for data production and microdata
  – Statistical Data and Metadata Exchange (SDMX), for dissemination of aggregate data
Conclusions

• DDI provides useful tools for using and producing data today
  – Many tools are freely available
• DDI is also part of the evolving picture of collaborative developments among national statistical agencies
  – DDI will become more significant in the future as a result
Questions?

[Ask Wendy Thomas]