

IASSIST 2002

The Metadata Toolbox:

A User's Perspective on DDI

**J.M. Eisenhauer Smith, Data Analyst/Archivist
Center for Demography of Health and Aging
University of Wisconsin - Madison**

Characteristics of Good Documentation

◆ User-Friendly

- Complete
 - ◆ should facilitate quality secondary analyses without adhoc assistance from archivist or original data producer.
- Useful as finding aid
 - ◆ Potential users should be able to quickly identify datasets that are likely to support research objectives BEFORE investing significant time and effort in learning about a dataset.
- Standardized format and content
 - ◆ so user's have to learn only one system.

Characteristics of Good Documentation

◆ Producer-Friendly

- Requires minimal programming skills.
- Offers guidance for data professionals not trained as archivists or librarians.
- Provides opportunity to economize, i.e., to consolidate external and internal documentation into a single document.

Where to Start

◆ DDI Tag Library

- <http://www.icpsr.umich.edu/DDI/CODEBOOK/code.dtd.html>

◆ Introduction to XML

- <http://www.xml.com/pub/a/98/10/guide0.html>

◆ Sample DDI-Compliant Codebooks

- <http://www.icpsr.umich.edu/DDI/SAMPLES/index.html>

Creating DDI-compliant Documentation

- ◆ Text Editor
- ◆ XML Editors
- ◆ MADDIE
- ◆ NESSTAR
 - Data Builder
 - XML Generator

Text Editors

◆ Advantages

- Because DTD is defined, XML is simple to program.
- Good way to become intimately familiar with DDI.

◆ Disadvantages

- No debugging tool (“validator”)
- Typos, typos, typos

XML Editors

◆ Advantages

- Flexible
- Customizable
- Debugging tool (“validator”) can be created

◆ Disadvantages

- Even the “simple” ones (e.g., WordPerfect) are not easy to setup for non-programmers

MADDIE: Advantages

- ◆ Template = click and type
 - Promotes understanding DDI for inexperienced
 - User is assured control over output (what you see is what you get)
 - ◆ a real XML document
 - ◆ no mystery transformation from fill-in-the-blank windows to XML text file
- ◆ Guidance for the inexperienced
 - templates for “recommended” fields
- ◆ DDI context-sensitive help included
- ◆ Tech support is fabulous: capable and timely

MADDIE: Disadvantages

- ◆ Data Description must be tagged manually
 - incredibly tedious and time consuming
 - tool to automatically generate documentation from datasets in common formats (e.g., SAS, SPSS, STATA, DBF) desperately needed.
- ◆ DDI help too thin to be really useful to the inexperienced
 - more and more and more examples

MADDIE: Technical Issues

- ◆ Needs specialized Perl modules to run software
 - Solve this problem by shipping Perl modules and instructions for installation w/ MADDIE
- ◆ Repeated field bugs need fixing
- ◆ Tree view is really great, but not helpful if can't go back and forth to document
- ◆ Windows version in development

MADDIE: Conclusions

- ◆ Advantages outweigh disadvantages for novice DDI implementers
 - DDI specifies “required” or “optional;” **MADDIE gives us “recommended.”**
 - Template-type display means non-programmers learn DDI and XML by osmosis.
- ◆ CAVEAT: More powerful tools available for creating “Data” and “Variable” Descriptions

NESSTAR

◆ Data Builder

- Available on CD

◆ XML Generator

- <http://www.nesstar.org/freesoftware/>

◆ Plan is to combine the two tools into one comprehensive utility

NESSTAR: Advantages

- ◆ XML Generator creates “data” and “variable” descriptions from SPSS, STATA and other datafiles
- ◆ XML Heirarchical Generator makes it easy to link datasets from longitudinal or other multi-file datasets

NESSTAR: Other Advantages

- ◆ Copy-paste fully implemented
 - e.g., apply common text to multiple variables by selecting all and typing once
- ◆ DDI context-sensitive help is relatively complete
- ◆ Could not be easier to use
 - programming = fill in a form
 - tech support is fabulous: capable and timely

NESSTAR: Disadvantages

◆ Documentation is thin

- Some nuances can be learned only by watching a demonstration

◆ DDI context-sensitive help

- need more and more examples

◆ ?????? NESSTAR goes .com ??????

- tremendous integrity and ingenuity vs. reality of a capitalist existence
- recommendation: software is great so go get it now while it's free

NESSTAR: Conclusions

- ◆ Advantages far outweigh the disadvantages
- ◆ The tool of choice for
 - those who prefer to avoid any direct contact with programming
 - those already intimately familiar with DDI
 - ◆ there is benefit to doing it the old-fashioned way once (i.e., w/ text file or MADDIE)
- ◆ The ONLY tool for
 - creating data and variable descriptions

Wish List: DDI Committee

- ◆ multi-language support
- ◆ “recommended” fields
- ◆ incorporate GIS metadata standards
 - <http://www.fgdc.gov/metadata/metadata.html>
- ◆ more and more and more examples

Wish List: Software Developers

context-sensitive help

- more and more and more examples
- ◆ default settings (e.g., producer)
- ◆ hyperlinks (e.g., producer email, funder websites, project websites)
- ◆ clear delineation: required vs. optional
- ◆ GIS metadata tools
 - <http://www.fgdc.gov/metadata/metadata.html>