

Integrated Online Analysis: Evaluating Nesstar and SDA

Marc Maynard

*The Roper Center for Public Opinion Research
University of Connecticut*

IASSIST Conference 2004, Madison, WI

www.RoperCenter.uconn.edu

Integrated Online Analysis

- Background
- Scope
- Evaluation Criteria
- Evaluation Results
- Conclusions

Background

- Why?
 - To supplement the data discovery process
 - To provide *basic* exploratory analysis and possibly some graphical reporting capabilities
- iPOLL
 - holds nearly 500,000 questions (60% linked to datasets)
- Previous attempts
 - Static cross-tabulation tables for some 20k questions (early 1990s)
- Dissimilar data formats and structure
 - from a variety of data sources and time periods

Scope

- Focused on SDA 1.2b and Nesstar 2.32b
 - SDA 1.3 released April, 2004
 - Nesstar 3.0 just released
- Integrated product not stand-alone
- Off-the-shelf versions
- Other models for analysis systems
 - UNC-Chapel Hill (SAS)
 - P. Yott workshop from IASSIST 1998 (SPSS)
 - Asteroid from Roy Morgan Research (Australia)

Evaluation Criteria

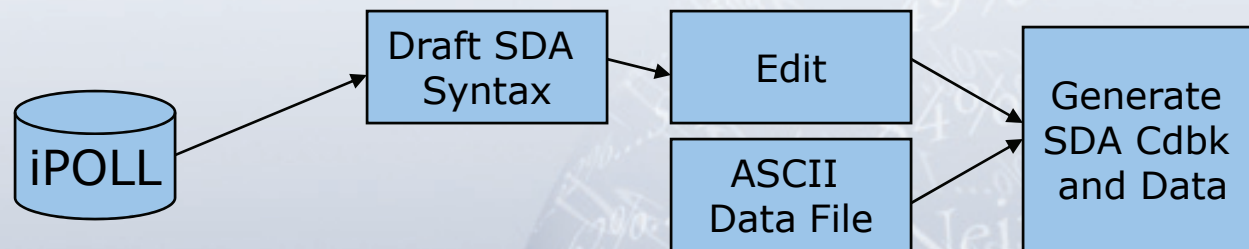
- File preparation
- System maintenance
- Ease of integration
- Performance issues
- Presentation features
- Standards (DDI)
- Other features

File Preparation

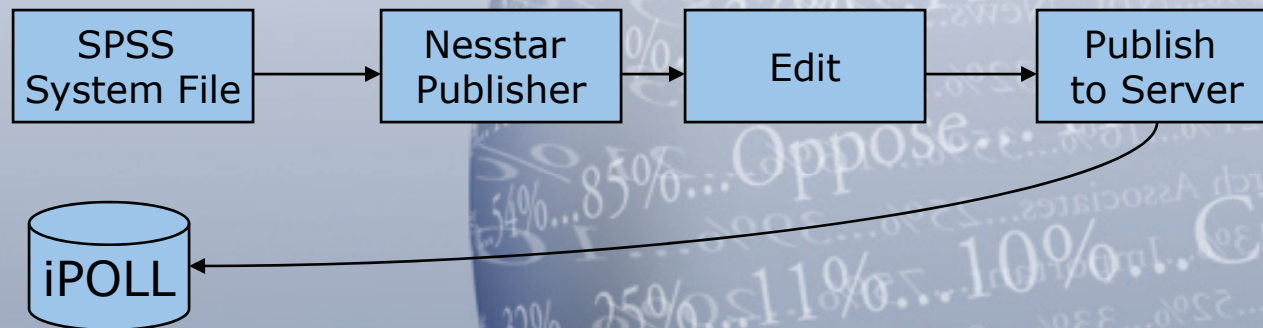
- SDA 1.2b:
 - DOS-based
 - Employs a text data definition file in a specified format
 - Requires a standard ASCII data file for processing
 - Utilities to convert SPSS/SAS/Stata data definition statements to SDA-DDL syntax
- Nesstar 2.32b:
 - Windows-based
 - Publisher handles virtually all file preparation
 - imports variety of formats directly
 - Provides a nice interface based on DDI elements to compile study metadata and publish to the server

File Preparation

SDA:

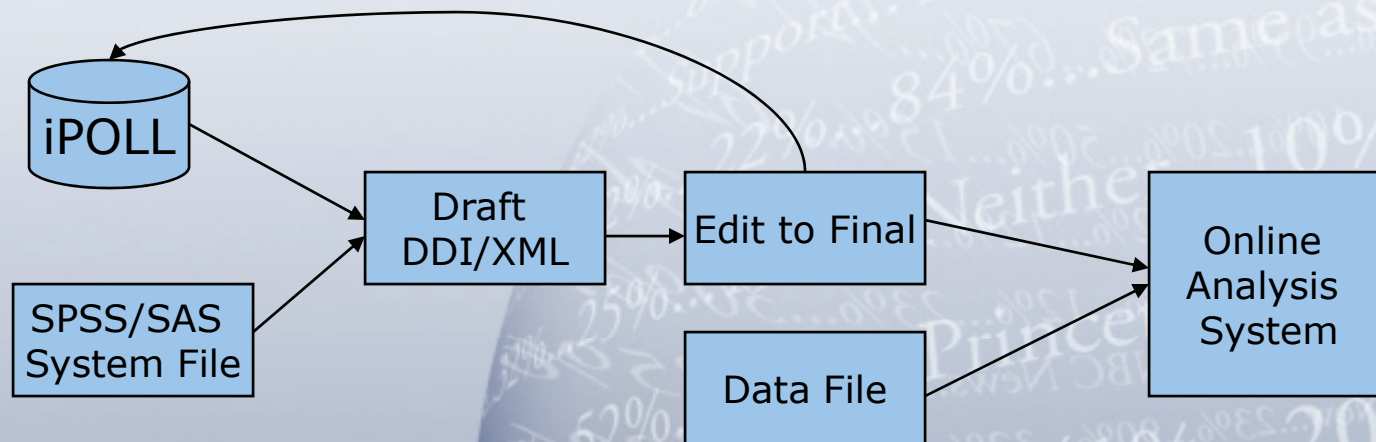


Nesstar:



File Preparation

DDI/XML:



System Maintenance

- SDA 1.2b:
 - CGI script-based operation
 - SDA syntax text files must be retained (and versioned) along with standard documentation
 - Codebooks must be recreated and stored separate from data
- Nesstar 2.32b:
 - Java-based application
 - Metadata stored in MySQL database
 - Documentation and data modifications handled as one operation

Ease of Integration

- Dynamically generate links or calls to the analysis system from within iPOLL
 - iPOLL need to first recognize studies or questions available for online analysis
 - Generated link needs to provide the analysis system with the required information to process the request
- Little formal documentation on this process
- Variable referencing issues (Nesstar)

Ease of Integration

- SDA 1.2b:
 - ended up decomposing the default forms
 - developing our own default parameter values
 - many parameters could be held constant and set by the iPOLL system in hidden form elements
 - simplified the form, allowing basic options to the user
 - handles one “transaction” at a time

Ease of Integration

- Nesstar 2.32b:
 - based on object-model (catalog, study, variable)
 - fairly simple in its basic implementation
 - because of its use of frames and relative addressing, it seems to get complex pretty quickly when going beyond the simply accessing a specific object
 - opens the object within context, allowing for a richer experience for the user

Performance

- Comparing apples and oranges
- SDA 1.2b:
 - Light and stable CGI scripts called directly for single operations
- Nesstar 2.32b :
 - Heavier web client (including context frames) requiring more processing effort but provides context for the analysis

Presentation Features

- SDA 1.2b:
 - Simple table reports
 - Limited options (color coded vs. not, question text vs. not, etc.)
 - Little direct customization available
- Nesstar 2.32b :
 - More detailed report options
 - Graphics available
 - Moderately customizable (header graphics, etc.)

Standards (DDI)

- SDA 1.2b:
 - SDA 1.2b was not directly-based on DDI
 - In July 2003 offered DDI to DDL conversion service
 - Version 1.3 offers more direct integration of DDI/XML to data processing
- Nesstar 2.32b:
 - Has integrated the DDI in all aspects of data processing
- Both store data in their own internal formats:
 - SDA using data vectors (a file for each variable)
 - Nesstar using NSDStat format for data storage and manipulation

Other Features

- SDA allows for user-driven variable recoding and computation of new variables
- Nesstar provides context for all analyses
- Both provide subsetting (variable and case)
- Both allow dataset downloads:
 - SDA: Customized data download requests are restricted to ASCII, CSV or space delimited formats with SPSS/SAS/DDI XML DDL syntax
 - Nesstar: Data stored in one format (NSDstat) and converted 'on-the-fly' to other formats

Conclusions

SDA

- Established Technology
- Focused Functionality
- Moving to DDI
- Established User Community
- Integration:
 - Form-based
 - Fairly simple

Nesstar

- New Technology
- Expanding Functionality
- Based on DDI
- Developing User Community
- Integration:
 - URL-based
 - Some complexity
 - Variable reference issues

Conclusions

While both systems may have their place depending upon the specific application, some questions still remain:

- ? Cost of ownership
- ? Scalability
- ? Resource availability
- ? New release improvements

References

- iPOLL Databank:
 - <http://www.ropercenter.uconn.edu/ipoll.html>
- SDA:
 - <http://sda.berkeley.edu>
- Nesstar:
 - <http://www.nesstar.org>
- Email: Marc.Maynard@UConn.edu