AAPOR

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

May 18–21, 2017
Sheraton New Orleans
New Orleans, Louisiana

72nd
Annual
Conference

Embracing Change and Diversity in Public Opinion and Social Research

www.aapor.org/conference

#AAPOR

AAPOR 72nd Annual Conference
Embracing Change and Diversity in
Public Opinion and Social Research

May 18–21, 2017
Sheraton New Orleans
New Orleans, Louisiana

Lowering the Barriers to Capturing Questionnaire Metadata Throughout the Data Lifecycle

Barry T. Radler, PhD

University of Wisconsin-Madison Institute on Aging

Overview

- **Metadata**
  - **Definition**
  - **Importance in Survey Research**
- **DDI**
  - **Role in Survey Research**
- **MIDUS study and UWSC experience**
  - **Lessons in Lowering the Barriers to Capture Metadata**
- **Summary and Future Directions**

What is metadata?

Metadata are like punctuation

duringworldwarIIstatisticiansengagedingovernmentsamplesurv
eysforthebureauofthecensusthebureauofagriculturaleconomics
andtheresearchbranchoftheusarmysinformationandeducationd
ivisionadoptedprobabilitytheoryasthebasisforpopulationsampli
nganadvantageclaimedforprobabilitysamplingwasitscapabilityo
festimatingsamplingerrorbecausesimplerandomsamplingwasn
otpossibleforgeographicallydispersedpopulationsstatisticiansd
evisedingeniousmultistageareadesignsthatwerebothpracticala
ndinconformitywithprobabilitytheory

...for your data

During World War II statisticians engaged in government sample surveys for the Bureau of the Census, the Bureau of Agricultural Economics, and the Research Branch of the US Army's Information and Education Division adopted probability theory as the basis for population sampling.

An advantage claimed for probability sampling was its capability of estimating sampling error. Because simple random sampling was not possible for geographically dispersed populations, statisticians devised ingenious multistage area designs, that were both practical and in conformity with probability theory.

# Types of Research Metadata

- **Study-level Survey Metadata**
  - **Description of goals, methods, results**
- **Variable-level Questionnaire Metadata**
  - **Description of datasets, variables, datums**
    - full question text
    - interviewer instructions
    - skip directions, routing, or question universe
    - response options, code lists
    - variable names, variable and value labels
    - variable groupings
    - missing value specifications or schemes
    - physical format (numeric, string, data, etc.)
    - column identifiers and physical layout of data file

Redundancy in Metadata Production
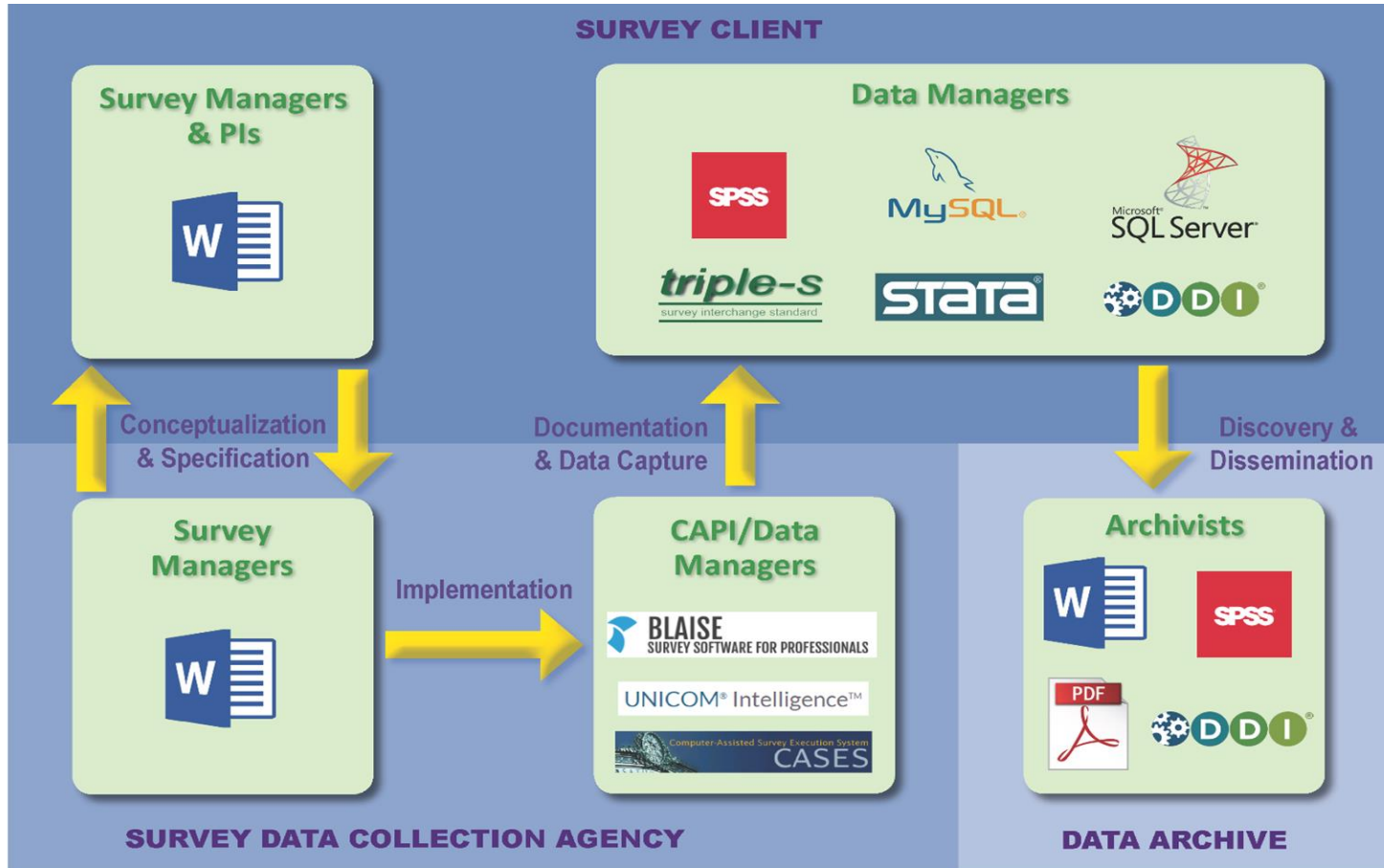
Research Data Lifecycle

Survey Metadata Desiderata

- Capture what was intended
  - What: what data were captured and why
- Capture exactly what was used in the survey implementation
  - How: the mode, logic employed and under what conditions
- Specify what the data output will be
  - That is, mirrors what was captured and its source
- Keep the connection
  - Between the survey implementation through to the data received -> data management by PIs -> to archiving
- Generalised solution
  - So that is can be actioned efficiently and is self-describing
  - So that it can be rendered in different forms for different purposes

Enter: DDI

□ The Data Documentation Initiative (DDI) is an international metadata standard for **describing the data produced by surveys.**

□ DDI is a **free standard** that can document different stages in the research data lifecycle.

- **Advantages of DDI:**
  - *A free and open web-publishing standard (XML)*
    - *Is interoperable; not limited by OS*
  - *One document → Many uses*
    - *XML plays well with other systems and software*
    - *Easily transformed into presentation languages such as HTML, PDF or plain text.*
- **Result of DDI:**
  - ***Introduces a common communication protocol to research processes***
  - ***Machine-actionable information about research processes and products***

MIDUS and UWSC

Background on MIDUS and UWSC

- **MIDUS (Midlife in the U.S.)**
  - **Longitudinal multi-disciplinary study**
    - **Large and complex amount of survey and other data**
  - **MIDUS DDI Portal – http://midus.colectica.org**

Background on UWSC

- **UWSC (University of Wisconsin Survey Center)**
  - **Goals:**
    - **Produce *one authoritative source document* that can be reused through lifecycle**
    - **Create authoring tool that clients are familiar with: MS Word**

UWSC Word Template **PROOF OF CONCEPT**

UWSC experience

☐ **Challenge 1:**

  ▪ **Describe how an instrument:**

    ▪ *Behaves* (instrument logic and variable metadata)

    ▪ *Looks (*layout, display, graphics)

      ▪ Especially useful for mixed mode surveys

      ▪ DDI is limited in documenting display issues for production

      ▪ Can reference external content (URLs)

UWSC experience

- **Challenge 2:**
  - **Different actors, different needs**
    - The data collector (**producer**/designer) wants to document the project management *processes* involved from conceptualization to fielding of final instrument.
    - The client (**user**/analyst) wants to document the *results* produced by the final instrument during fielding.

Conclusions

Conclusions and Challenges

- **Capturing metadata early**
  - **One DDI document → repurposed for multiple uses**
    - **Reduce redundancy and information loss**
- **Challenges**
  - **Technical issues**
    - **Need for tool to bridge all lifecycle stages and different systems**
    - **Instrument behavior and display in different modes of administration**
  - **Non-technical issues**
    - **Distinct and non-overlapping metadata needs across different stakeholders**
    - **Documenting Products vs Processes**

AAPOR 72nd Annual Conference
Embracing Change and Diversity in
Public Opinion and Social Research

72nd Annual Conference

AAPOR
AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH
May 18–21, 2017
Sheraton New Orleans
New Orleans, Louisiana

MIDUS

Thank you!

DDI®

bradler@wisc.edu

This presentation is offered under license CC BY-SA 4.0