Starting in the Middle: The Wisconsin Longitudinal Study and DDI

Carol Roan, Stan Howald, Vicky Chang, & Joe Savard

Middle?

Beginning

Now

Future

Middle?

WLS methods of data sharing and providing metadata have changed with technology and with emerging data standards.

Our current metadata is good but we know we can improve.

What is the WLS?

- Panel Study of Wisconsin High Schools' class of 1957
- Six Waves of Interview Data

What is the WLS?

Wave	Year(s)	Participants	Mode
1	1957	WI High School Seniors in April 1957 (Graduates)	Self-Administered Questionnaire (SAQ)
2	1964	Parent	Postcard SAQ via mail
3	1975 / 1977	Graduates / Partial sample of Siblings	Telephone Suvey (Interviewers used paper instrument)
4	1993 / 1994	Graduates / Siblings	CATI with post mail-out SAQ; SAQ-ONLY for some Siblings
5	2004 / 2005	Graduates and Graduates' Spouses / Siblings and Siblings' Spouses	CATI with post mail-out SAQ for Graduates and Siblings (Not spouses); SAQ-only for participants who refused CATI
6	2011	Graduates & Siblings; Proxies	CAPI with SAQ hand-off; CATI for CAPI refusers; SAQ-only for CATI & CAPI Refusers

What is the WLS? (continued)

- Panel Study of Wisconsin High Schools' class of 1957
- Six Waves of Interview Data

- Administrative Data
- Genetic Data for Graduates and Siblings
- Microbiome Data (Pilot Sample)
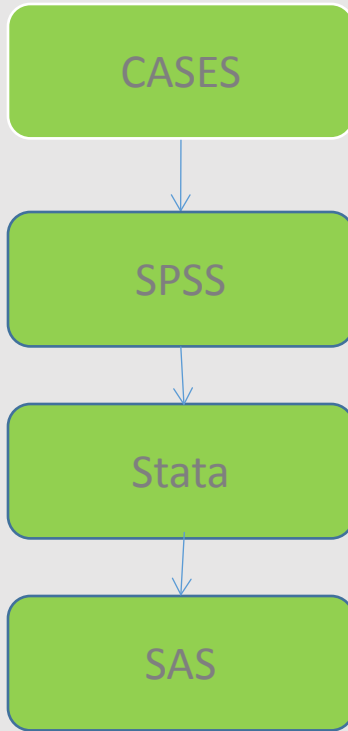
Survey Content

- "WLS is the study of Everything" – Bob Hauser
- Questions/Topics evolved over time

- Google "Wisconsin Longitudinal Study"
- www.ssc.wisc.edu/wlsresearch

WLS Metadata

- From 2004 onward it's a DDI 2-compliant XML datafile.
  - Thanks Jeremy Iverson!

- Our XML is a middle layer not a permanent repository

- WLS metadata are developed and maintained as text files.

DATA

CASES

↓

SPSS

↓

Stata

↓

SAS

↓

TXT → **XML** → **postgreSQL PHP HTML**

METADATA

Downloadable datasets (SAS, Stata, SPSS) contain basic documentation:
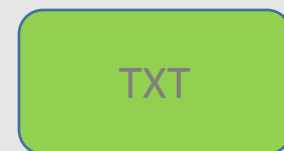
Variable labels
Value labels

WWW Documentation

Flowcharts
Module overviews
Variable labels
Value labels
Notes
Appendices

Search some or all collection waves by topic or variable name or "label contains XYZ"

WLS Metadata a closer look

- Hierarchical:
  - Module-level
    - Overviews including brief variable descriptions
  - Variable-level
    - Variable label
    - Respondent Type
    - Mode
    - Source Items (question on instrument)
    - Year of collection
    - Notes
  - Value-level
    - Value labels

{some} variables are also grouped into topics and subtopics

Why DDI Lifecycle? WLS goals

- Richer metadata allows more user interface possibilities
  - multiple variable contexts (hierarchical vs topical vs longitudinal)
- Better interface between instrument and variables
  - Link between original question wording and analysis variable
- Versions management
- Increase usability
  - Harmonization
    - Within wave across participants
    - Across waves for the same person
    - Link WLS to similar studies via common measures/concepts
- Long-term study viability (curation using best practices)

Also

Less proprietary more standards-based!

(what if Stan gets hit by a bus?)

Nuts & Bolts: how do we do this?

- We're planning to use Colectica software/expertise
- Learn from other studies using DDI (e.g. MIDUS)

- Convert current metadata to a DDI 3 repository
- Fill in the blanks
- Add more layers
  - By concept: closely-associated variables (same question)
  - By topic, subtopic

Challenge #1 Data security

- Abridged public data vs unabridged (protected) data
  - Documentation
    - Top/bottom coding
    - Anonymized (de-identified) data records
  - Access
    - Are variables flagged public vs non-public?
      - General web app: Only *public* variables can be downloaded

More Questions and Challenges

- 28,000 variables
  - Value labels (code sets) duplicated; no central library
    - Fix before or after this mess lands in DDI3?
  - Measures meant to be longitudinal are "close" but not exact
    - Can/should these differences be noted in metadata?
  - Changes: long-time data users may need to update their older programs
    - BUT fixing data errors and improving metadata are not equivalent