Protecting the Privacy of Participants: Lessons Learned from WILIS 3

SUSAN RATHBUN-GRUBB, PHD
Assistant Professor
School of Library and Information Science,
University of South Carolina
Co-investigator, WILIS 3

CHERYL A. THOMPSON, MSIS
Research Data Coordinator
UNC Odum Institute Archive
Graduate research assistant, WILIS 3

JENNIFER CRAFT MORGAN, PHD
Research Scientist
UNC Institute on Aging
Co-principal investigator, WILIS 3

JOANNE GARD MARSHALL, PHD
Alumni Distinguished Professor
School of Information and Library Science,
University of North Carolina
Co-principal investigator, WILIS 3
Agenda

- Workforce Issues in Library and Information Science (WILIS) projects (http://wilis.unc.edu/)
  - Funded by the Institute of Museum and Library Services (IMLS)

- Protecting privacy of participants
  - Methods
  - Challenges
  - Lessons learned
  - Future directions
WILIS projects – Context

- **WILIS 1**
  - Retrospective career survey of graduates of 5 LIS programs in North Carolina (1964-2007)

- **WILIS 2**
  - Survey of recent graduates of 39 LIS programs in North America (2000-2009)

- **WILIS 3**
  - Genesis: this is important, useful data, and a lot of it... what do we do now?
Importance of Archiving Research Data

- Promoting transparency in the use of public funds for scientific research
- Supporting ongoing analysis of research data beyond the initial investigation
- Preserving raw data as part of the historical record
- Integrating LIS researchers in the development of models for archiving data
WILIS 3 – Goals

1. Create datasets from the WILIS 1 and WILIS 2 projects for public and investigator use

2. Develop and document a collaborative approach to data sharing

3. Produce a best practices toolkit for data archiving

4. Develop a model for other IMLS grant holders to support future LIS research and data sharing
Collaboration

- Partners – represent three perspectives on research process
  - Researcher perspective: WILIS research team
  - Data collection perspective: Survey Sciences Group, Inc.
  - Data archivist perspective: Odum Institute Data Archive
Collaboration

- Researcher perspective
  - Project design, conceptual framing, research questions informed by literature, survey/measurement quality, needs of LIS stakeholders, human subjects protection, analysis needs

- Data collection perspective
  - Survey software, data structure, sample management, methodological issues

- Data archivist perspective
  - Standards for data documentation, Dataverse interface and capabilities, ingesting process
Privacy concerns

DE-IDENTIFICATION:

METHODS, CHALLENGES, AND LESSONS LEARNED
## WILIS data sets

<table>
<thead>
<tr>
<th>WILIS 1</th>
<th>WILIS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contains educational and work histories, periods of unemployment, career motivations, continuing education needs, retirement intentions, career satisfaction and LIS program satisfaction</td>
<td>Contains educational history, brief work history, periods of unemployment, career motivations, continuing education needs, career satisfaction and extended LIS program assessment</td>
</tr>
<tr>
<td>Over 1700 variables, structured and textual</td>
<td>About 500 variables, structured and textual</td>
</tr>
<tr>
<td>2,653 cases representing 5 LIS programs</td>
<td>3,507 cases representing 39 LIS programs</td>
</tr>
</tbody>
</table>
Methods

- Review variables in WILIS 1 and 2 datasets to de-identify participants
  - Three-pass process
    - 2 Odum passes
    - 1 final WILIS researcher pass
  - Developed rules for redacting, collapsing, and deleting
  - Tracked edits to data sets

- Create public use and restricted use versions of data
Challenges

- Getting on the same page (collaborators)
- Anticipating researcher needs and use of data
- Balancing participant protection and data usability
- Working with unstructured or open ended data
Challenges

- **Data set problems**
  - File formats
  - Variable labels/naming conventions
  - File versioning
  - Constructed variables

- **Timing** – it takes longer than anticipated
Lessons Learned

- Consult a data archivist early in the research process to minimize challenges

- Build in staff time and resources for de-identification at the outset of the project

- Identify survey questions that could collect personal information
  - Consider which single pieces of information, and what combination of variables are identifying
Lessons Learned

- Think about how you will deal with open-ended responses ahead of time (redacting takes time)

- Consider the value of multiple passes by archivist and researchers

- Be careful with data file versioning and format conversions

- Keep notes on the de-identification process - challenges, rules, and decisions. Include it in your submission package.
Future Directions

- Finalize archive in Dataverse and publicize availability.

- Finalize user tools (e.g. short videos) to help the public and other researchers use the data.

- Document the process, costs, and challenges of archiving the WILIS data.

- Finalize the toolkit to help other researchers plan for, cost out, and successfully archive their data.
Thank you

• Questions?

• Susan Rathbun-Grubb (srathbun@mailbox.sc.edu)
• Cheryl A. Thompson (ct.annette@gmail.com)
• Jennifer Craft Morgan (jcraftmorgan@gmail.com)
• Joanne Gard Marshall (marshall@ils.unc.edu)