RE-USING QUALITATIVE DATA: QUALITATIVE RESEARCHERS’ UNDERSTANDING OF AND THEIR EXPERIENCE ON DATA REUSE

SESSION BLOCK 4: THURSDAY, JUNE 5, 2014
AYOUNG YOON
SCHOOL OF INFORMATION SCIENCE
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
AGENDA

• Qualitative data reuse (secondary analysis) landscape
• Research motivations and research questions
• Research method
• Preliminary findings
• Discussion & future study
• Questions
QUALITATIVE DATA & SECONDARY ANALYSIS (OR REUSE)

• Data reuse
  • Researchers generally understand *reuse* as the secondary use of data, not for the original purpose but for studying new problems (Karasti & Baker, 2008; Zimmerman, 2008).

• Secondary analysis
  • “Any further analysis of an existing datasets which presents interpretations, conclusions, or knowledge additional to, or different from, those presented in the first report on the inquiry as a whole and its main results.” (Hakim, 1982, p.1)
  • “A re-working of data already analyzed. As such, it may appear to offer little by way of originality and seem to be an unlikely method of revealing new and exiting findings.” (Dale et al., 1988, p. 3)
QUALITATIVE DATA REUSE

• A well-established tradition of the secondary analysis of quantitative social science data

• Skepticism about the secondary analysis of qualitative data (e.g., Van Den Berg, 2005; Corti, 2002)

• Discussions regarding qualitative data sharing and reuse have also been very active, for instance in Europe and Australia (e.g. the establishment of the Qualitative Data Archival Resource Centre in UK)

• Discussions do not seem to be as prominent in the United States, despite a long history of depositing and curating practices of scientific data.
LITERATURE REVIEW

- Benefits of data sharing/reuse (e.g., Borgman, 2010; Borgman, 2011; Hey & Trefethen, 2003; Hey, Tansley, & Tolle, 2009).
  - Validate existing results
  - Generate new findings built on the work of others
  - Ask new questions of extant data, advance solutions for complex human problems and the state of science, reproduce research, and expand the instruments and products of research to new communities
LITERATURE REVIEW

• Difficulties in qualitative data reuse (secondary data analysis) (Corti & Thompson, 2004)
  • Ethical and consent considerations
  • Representation, coverage and context of the research and fieldwork
  • Unfamiliarity with the methods
  • Lack of infrastructure for data-sharing
  • Misinterpretation of data
RESEARCH MOTIVATIONS & QUESTIONS

• Qualitative data reuse in the states? (Empirical research)
• Focusing on understanding barriers or hindrances to reusing qualitative data and the appropriate conditions for reusing qualitative data

1. What is the distinctive nature of qualitative research data, that can potentially influence to data reuse?

2. How and why do researchers reuse data?
   1. What are the advantages and disadvantages of this experience?
   2. What are the barriers or hindrances to reusing qualitative data?

3. What would be the appropriate conditions for reusing qualitative data?
METHOD

• In-depth semi-structured phone interviews with qualitative researchers who have reused data.
  • 45 to 80 min
• Study participants
  • Qualitative social science researchers
  • Data citation tracking from major databases (e.g. Sage journals, ERIC, etc.)
  • Publication since 2004
  • Total 40 for initial contact – 13 responses (response rate 32.5 %)
  • 8 interviewed, 2 scheduled.
PRELIMINARY FINDING - PARTICIPANTS

- All female
- 2 associate and full professors, 2 researcher / consultant, rest were assistant professors
- Age: 20s to 50s
- Years in qualitative research: 3 to 25 years
- Experience in data reuse: 1 to 2 times
PRELIMINARY FINDING – CHOOSING DATASET

- Reason for choosing datasets: prior connection to the original dataset and PIs of the studies
  - Advisor or advisor's co-worker
  - Colleague who met in the conference
  - Community or organization previously worked with
PRELIMINARY FINDING – TRUSTING DATA

“I would never analyze data from somebody I didn't know. [...] I would only work with somebody if I had access to them. And that would be a very credible part for me.” (P02)

“I think I would only do it with data sets from researchers that I actually knew, and knew well and that they were aligned with my philosophy. [...] The original researchers and that their philosophy was aligned with mine, I feel more confident in analyzing their data.” (P07)

“I trust my, trusted [name of researcher], knew the quality of the research that she has done in the past.” (P06)

“[It’s] to the work that's being done with the person, who's pursuing the secondary analysis, it's to look at, "Okay, is this someone who is strong in qualitative methods? Is this someone who was experienced?" Just because you can't speak to that yourself, because you haven't gone through the process.” (P03)
PRELIMINARY FINDING – GETTING DATASET

• Process of acquiring datasets
  • Download from organization / community's website with permission
  • From original investigators
  • Formal agreement form (P03)
    • “It kind of outlined my overall purpose for the secondary analysis, what I was requesting, and kind of like the intended use. So, really a lot of what you might see when you, you know, you're looking to use a quantitative database from somewhere that you might have, a user's agreement form or something like that.”

• Datasets: transcripts / observation notes and study descriptions
  • PIs’ information, interviewers’ information, audio tapes, original investigators’ memos, notes, and coded transcripts, pseudonyms or code numbers, the characteristics of the sample
PRELIMINARY FINDING - UNDERSTANDING

• Process of understanding
  • “It was a super learning curve for me, just to figure out exactly the parameters and the context where the other data was collected. (P01)
  • Interaction with original data collectors/investigators
    “My committee, as a team, could help bring me up to speed in terms of how the data was collected. And so I can really understand the original context of the data collection, and then they could answer questions for me. I think, if there was nobody on the committee who had actually collected the data, it would have been really hard for me to... So that's I think, one of the key things that was important to me is that, I have access to people who had been involved in that original study.” (P06)
  • Infer from own experience
PRELIMINARY FINDING - CHALLENGES

- Different purpose
  - “[The data] was not [created for] my number one intended purpose, I was kind of making a square fit into a circle in some ways.” (P01)
  - “It's actually pretty problematic when... Because it's time, it's about time. It's about time and interpretation. And so if you're going to do something like that you really need to indicate that the person who is being interviewed doesn't have control over interpretation. Right? And so I had some people that disagreed with my interpretation of what was going on and they made some demands that I could not meet. They wanted me to reinterpret a document entirely and that just wasn't possible.” (P05)
PRELIMINARY FINDING - CHALLENGES

• Missing information and context: typo, abbreviation, tone of voice, probe, etc.

• Restriction
  • “Only concern that they would have had, had I asked to see them[original dataset], is that the anonymity and confidentiality of the people they were interviewing, family members they were interviewing would continue to be protected.” (P04)
PRELIMINARY FINDING - CHALLENGES

• Differences in fields
  • “The people who organize the data collection are coming primarily from a museum perspective and an oral history perspective, they also want to preserve the names of the people who are being interviewed and that's certainly not part of my research design. [...] I'm coming from more of a social science perspective. [...] The way that I was trained was that usually you try to protect your study participants. [...] So, what I had to do is figure out how do I do this in a way that is ethical, how do I do this in a way... There were whole bunch of issues that you really wouldn't think about for qualitative datasets.” (P05)
  • “It wasn't just the challenge of secondary data analysis, but it was that the ways... It was a disciplinary difference in how the data was collected and what was included in the set.” (P05)
PRELIMINARY FINDING - CHALLENGES

• Publishing

  • “Quantitative secondary analysis happens all the time, but qualitative secondary analysis is very unique and I think it takes a lot of over explaining. We ended up publishing the pieces, you know, but it took a while in order to pitch it in a way that it would be deemed acceptable. At first we wrote a very over explanatory piece talking about the valid points of secondary analysis, and that just was over burdening the whole argument, and so we took that out. So we really had to structure it.” (P08)

  • “There just weren't a lot of publishing outlets that I thought were very workable. That's a huge challenge and I think it should be explored a little bit more. But sometimes that's also part of how journals are constructed. So I think that especially if it's a methodological journal or a qualitative journal, that the mission statements of those journals also encourage authors to talk about their data collection process and things like that. [...] those journals really would not justifiably have room for secondary analysis.” (P06)
DISCUSSION & FUTURE STUDY

• Personal connection and/or interaction is the key of qualitative data reuse experience.
  • (How) Can this personal interaction be replaced?

• Infrastructure / disciplinary support

• Future study
  • Continue interviewing
  • Large scale study
THANK YOU!

Ayoung Yoon (ayyoon@email.unc.edu)
Doctoral candidate
School of Information and Library Science
University of North Carolina at Chapel Hill