Drowning in data: How to prioritize

Jennifer Watling Neal (jneal@msu.edu)
Zachary Neal (zpneal@msu.edu)
Michigan State University

Research as map making
Research as map making

You are here

Michigan School Program Information Project

Research as map making

You are here

Any good map shows the scale.

In this map, an inch represents five miles.
Research as map making

You are here

Any good map shows the scale.

In this map, an inch represents five miles.

What makes maps useful?
Research as map making

Would a map be useful if it had a 1:1 scale?

If an inch on the map represents an inch in reality?
Research as map making

Would a map be **useful** if it had a 1:1 scale?

If an inch on the map **represents** an inch in reality?

"What a useful thing a pocket-map is!" I remarked.

"That's another thing we've learned from your Nation," said Mein Herr, "map-making. But we've carried it much further than you... We actually made a map of the country, on the scale of a mile to the mile!"

"Have you used it much?" I enquired.

"It has never been spread out, yet," said Mein Herr. "The farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well."

– Lewis Carroll (1895)
Research as map making

We can wind up drowning in data when we try to reproduce reality.

But, we can make research useful when we only try to represent reality.

Remember –
“All models are wrong, but some models are useful”

Everything is connected. But, some connections matter more.

– Every network researcher ever

What is MiSPI?

What is **MiSPI**?

**Quantitative Data**
Networks linking 382 school administrators and 472 information brokers.

Surveys from all 854 participants

70 indicators for 3569 school entities

**Qualitative Data**
Interviews with 90 administrators, 50 brokers, & 12 researchers: 75+ hours audio, 2000+ transcript pages

What is **MiSPI**?


**Data used in this paper**

Interviews with 19 school administrators in 2 districts (~17 hours)

Surveys from 19 school administrators
We’ve avoided drowning in data by following six simple rules!
The Rules of Thumb

(1) Only collect the data you need

(2) Transcribe as you go

(3) Break analysis into single papers

(4) Code for one paper at a time

(5) Code w/ consistent unit of analysis

(6) Write for a specific audience
Rule 1: Only collect the data you need

Start with a well-planned overall design

Articulate clear research questions

Stick to your design and questions!

Bottom Line – If there is not a use for a piece of data in your design or questions, do not collect it!

Don’t collect data just because “it could be interesting to know…”

Rule 1: Only collect the data you need

We asked administrators to tell us about a school program or practice they recently considered adopting. We also asked how they learned about it.

It would have been interesting, but we didn’t ask them:
- If the program worked
- If the students liked the program
- What parents thought of the program
Rule 2: Transcribe as you go

Train a transcription team before you start interviews.

As interviews start, assign transcriptions to team.

Review transcripts while still collecting data.

Bottom Line – By the time you finish data collection, transcription should be nearing completion.
Rule 2: Transcribe as you go

When an interview is finished:
(1) We put the audio file on a secure shared drive, and add it to a signup sheet.

(2) A team member signs up to transcribe it, then emails us when they’re finished.

Our interviews are typically transcribed within 1-2 weeks.
Rule 3: Break analysis into single papers

Avoid analyzing everything at once

Conduct analyses in manageable chunks

Bottom Line –
A good strategy is to think about the papers or projects you want to accomplish and tackle analyses for each one at a time

Rule 3: Break analysis into single papers

This paper was about triadic types of brokerage described by Gould & Fernandez.

We did not attempt to explore:
- Other types of brokerage
- Knowledge utilization
- Attitudes about research
- Student outcomes
- Etc
Rule 4: Code for one paper at a time

Don’t try to code everything at once

Start with one paper or project

Code only the data and themes relevant to that project

Bottom Line – Coding for a single paper or project helps make the task easier & less monumental
Rule 4: Code for one paper at a time

We only coded interviews for mentions of information transfer where we could identify both the source and recipient of the information.

We did not code for:
– Attitudes about brokerage
– Attitudes about research
– Satisfaction with the process
– Etc.
Rule 5: Code w/ consistent unit of analysis

Determine a **unit of analysis** before you start coding

**Chunk** data eligible for by the unit of analysis

Bottom Line – Coding with a consistent unit of analysis makes it easier to determine coder reliability and to describe your coding process to others.
Rule 5: Code w/ consistent unit of analysis

Before coding, we partition all our interview transcripts into roughly 100-word chunks.

All coders are looking at the same chunks when deciding whether to apply a given code.

This makes it easy to
- Describe the method
- Refer back to chunks
- Compute reliability
Rule 6: Write for a specific audience

Pick a venue before you start writing

Understand (the audience for) that venue
- How much do they know about your topic?
- How much do they know about the methods in your study?
- How long can the paper be? How should it be formatted?

Tailor your writing based on audience knowledge

Bottom Line – Writing to a specific audience will increase the chances that your work is understood

Rule 6: Write for a specific audience

We selected the *American Journal of Community Psychology* as our target venue.

Based on this venue/audience, we:
- Carefully described triadic networks
- Used selected, abridged quotes
- Cited other work published in AJCP

We did not:
- Use network jargon
- Write more than 40 pages
Questions for Discussion –

What challenges do you think you or your research team might encounter trying to implement these rules?

What other challenges have you encountered that have left you “drowning in data”?

What other rules of thumb have worked for you?

Questions later?
Email us at mispi@msu.edu