Data State of Mind

BY MARYJO WEBSTER
DIGITAL FIRST MEDIA

After nearly 15 years of computer-assisted reporting and countless hours of teaching data analysis, I've come to believe that the key to successfully using data to produce better journalism is what I call the "data state of mind."

I picked up this term from someone else, and it's most certainly the digital-age equivalent of the "documents state of mind" that investigative reporters Donald Barlett and James Steele ingrained in the rest of us long ago.

Both phrases boil down to the same idea: data and documents are sources, just like people, but a little more reliable. Querying data is just like asking someone a question. You must decide how to phrase the right question to get the right answer. And each question usually leads to more questions. Just like an interview.

A data state of mind, though, tends to focus more on measuring or quantifying something. The best journalists start out by asking questions, such as, "How many deficient bridges are in my state and how has that changed over time?" or "What percentage of low-income students were grade-level proficient in math and reading last year?"

Having specific questions helps reporters find the focus of stories and, more importantly, helps them figure out what information they need to collect. This is a much better starting point for a story idea than setting out to write about "bad bridges" or "students falling behind."

A reporter with a data state of mind would first set out to find a spreadsheet or database that answers not only the specific question and many others but also provides examples and/or people needed to bring the story to life.

Yes, a reporter could probably call one or two human sources and get an answer to the first question, but is that really enough?

Not in my book. There are some stories that simply cannot be done without this kind of data analysis. And they always end up being really good stories.

Here's a very simple example from my newsroom at the St. Paul Pioneer Press: A business reporter wanted to bust the myth that the December holiday season was the busiest time of year at the Minneapolis-St. Paul International Airport. Many sources had told him this myth was untrue. But he didn't want to write a story only quoting people. He knew that wouldn't be sufficient.

So he went hunting for data and quickly found a Federal Aviation Administration database that listed all the flights in and out of each airport. He downloaded a full year's worth of data, and I helped him analyze it. We totaled the number of flights by day and by week and sorted the weeks by the number of flights. The busiest week? The first week of June, just after Minnesota schools let out for the summer.

I think teaching all reporters how to develop this data state of mind — even before setting them down with a spreadsheet — should be the first step in teaching data journalism.

The business reporter at the Pioneer Press had learned computer-assisted reporting skills many years before, but he felt really rusty. He needed help with the specifics of writing queries in Microsoft Access.

But the key thing is that story never would've come about if he didn't have that data state of mind.

I've met reporters who have a data state of mind (or something similar) ingrained in them even before they ever learn anything about data journalism. I've found that these individuals tend to grasp CAR skills really quickly.

So how do you teach the data state of mind? Here are a few ideas that reporters and/or editors could implement on their own:

- Encourage reporters to hunt down the data that exists on their beat. Gather key bits of information about what data is kept, how it's collected, who the "data keeper" is, how you could get a copy of this data, etc. Maybe even try to take the data keeper to lunch sometime to chat.

- Insist reporters frame their stories as questions, not statements or nouns. Rather than saying the story is about the city's "street improvement spending," have them ask "How much money did the city spend on street improvements last year?"

- Both reporters and editors can tune their radars for data opportunities, such as trend stories that vaguely say something has changed over time or an interview in which a source cites numbers from an analysis.

- When you find yourself asking a government agency for summary data — e.g., total number of crimes last year — stop and ask yourself if there might be details in the underlying data that would be useful either for this story or a follow-up.

- Breaking news events are a great time to practice. There's a fatal car accident, and you follow up a day or two later with the question, "How often has this happened at this location?"

MaryJo Webster started last month as the Senior Data Reporter for Digital First Media. Previously, she spent nine years as the computer-assisted reporting editor for the St. Paul Pioneer Press. She also teaches data journalism at the University of Minnesota. Previously, she worked at USA Today, the Center for Public Integrity and IRE.