FIND THE STORY

Powering your environmental reporting with data

Thoughts to keep in mind on a story

Question: Is there an on-topic dataset?

Yes. Excellent. Consider whether you can mesh it with a related dataset or add other useful information.

No. Don't despair. Consider building your own.

First example:

Making datasets play nicely with each other

America's super polluters

Data details

- Two EPA datasets: The Toxics Release Inventory (TRI) and the Greenhouse Gas Reporting Program.
- Each has information for thousands of power plants, factories and other industrial sites. (Not *all*, but a lot.) Companies self-report.
- Not all the facilities on one are also on the other, but there's a lot of overlap. You can see which facilities have reportable greenhouse-gas emissions *and* air emissions they're required to report to the TRI.
- HOWEVER: The two datasets don't use the same IDs. You need a shared ID system to pop the two files into Microsoft Access, SQL Server or another database manager and easily connect them.

Finally … the breakthrough

It turns out the EPA has *another* unique ID system for facilities, one that reaches across programs. Which makes perfect sense, right? But it took a while to find the EPA staffer who knew this and could explain where to find an additional dataset that would allow us to crosswalk from the IDs we had to the ones we wanted.

The EPA has since included these Facility Registration Service ID numbers in the TRI dataset itself. A nice change.

Quick pause for a public service announcement

Q: What is your data-driven story about?

A. Not data. People. (Or, if you're reporting on, say, endangered species, animals *and* people.)

As amazing as your findings might be, you'll have a far better story if you use them as a jumping-off point.

How are people affected? Who are some of those people?

Lessons learned from this story

- Always ask, "Is there a way to …?" Keep asking additional people, if necessary.
- Build external data checks into the process, early on if it makes sense. For this story, we gave site operators plenty of time to sift through their records and see if there were any problems with our analysis or any context we should have upfront.
- Newsrooms working together can accomplish more.

Consider newsroom-to-newsroom partnerships

Why?

- Amplify your work. Get the story in different places, or get more stories than you could produce on your own.
- Share resources (reporting, photography, graphics, editing) to get more bang for the buck.

Some fruits of the partnership with USA TODAY Network:

Monmouth landfill tops list of worst Shore air polluters

Russ Zimmer, @RussZimmer Published 9:55 p.m. MT Sept. 28, 2016 | Updated 5:40 a.m. MT Sept. 29, 2016
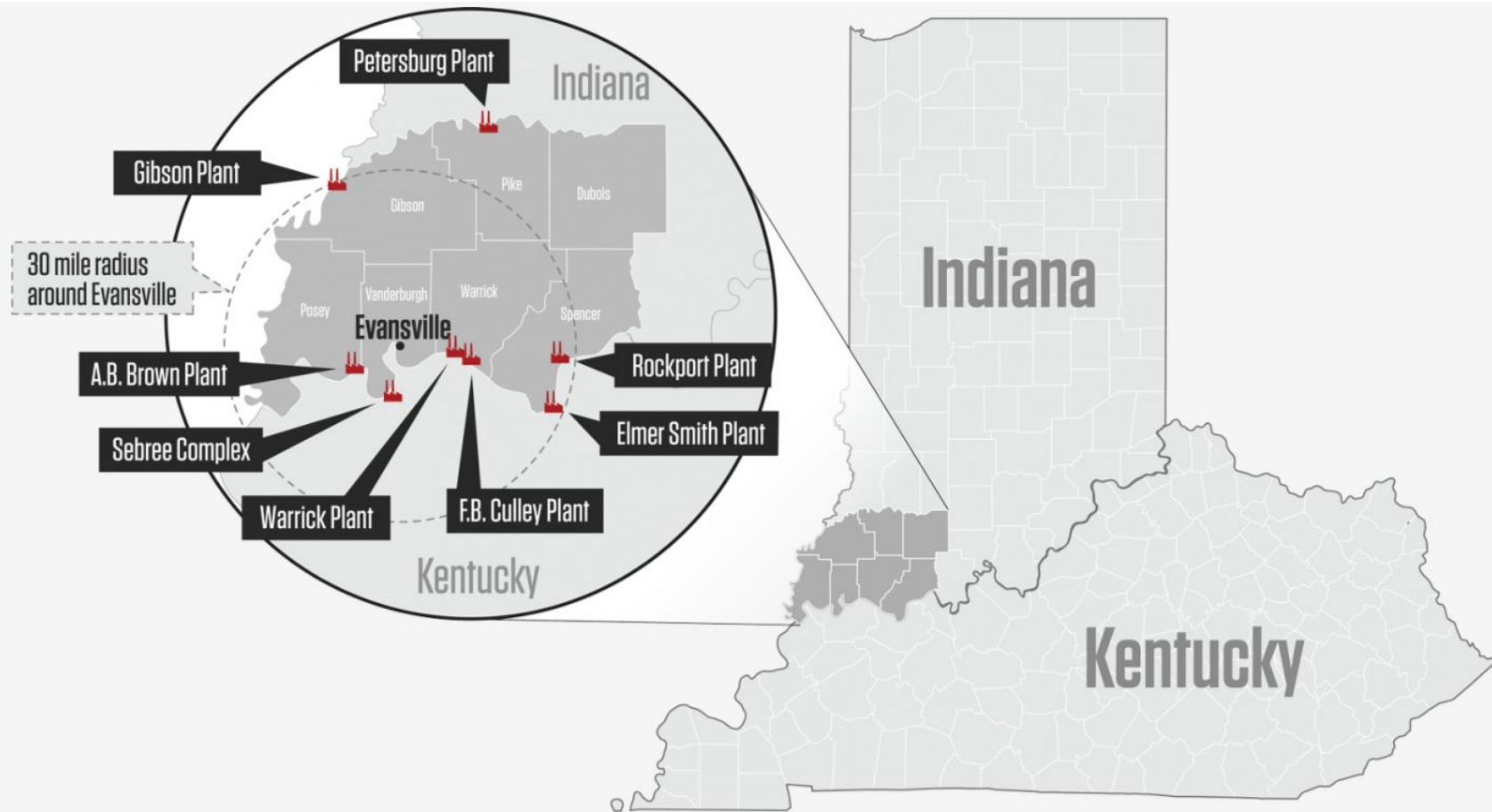
Iowa ranks in top 20 for toxic air releases

Donnelle Eller and Jeffrey C. Kummer, Des Moines Register Published 11:09 p.m. CT Sept. 28, 2016

SRP's northern Arizona coal plant one of the biggest carbon emitters in the country

Ryan Randazzo, The Republic | azcentral.comcPublished 6:06 a.m. MT Sept. 29, 2016

Weather Channel map for the win

That's it for Story No. 1.

Any questions before we move on?

OK, fine. But what if there's no data?

Well, that's annoying. But don't assume it's the end of the road for the story you hoped to tell.

Time for example No. 2: The case of the nonexistent national dataset on lead exposure.

Reuters' lead investigation

Unsafe at Any Level: The thousands of U.S. locales where lead poisoning is worse than in Flint

Problem #1: National, state and even county-level data on lead levels in blood tests obscure local hotspots -- the places most in need of help.

Solution: Get local data!

Which brings us to problem #2: There's no national dataset with local-level blood-test results.

So Reuters reporters went the DIY route.

They called state after state, asking for local figures to create their own dataset.

How Reuters handled the problem of non-responsive states:

"The nationwide map constructed through this analysis has empty spaces: The available data includes 21 states, home to around 61 percent of the U.S. population. Health departments in some states didn't possess the data or respond to records requests. Others wouldn't share it, saying they weren't required to, or citing patient privacy laws."

This is what the reporters discovered with Census tract and ZIP code data in hand:

"Reuters found nearly 3,000 areas with recently recorded lead poisoning rates at least double those in Flint during the peak of that city's contamination crisis. And more than 1,100 of these communities had a rate of elevated blood tests at least four times higher."

Lessons learned

- Localized data tells local as well as national stories. National-level data, by contrast, cannot accurately tell local stories. State and even county data aren't fine-grained enough, either, in a variety of cases.

- Get what you can, and start publishing. Lagging agencies that haven't coughed up data may do so later. Keep writing.

- Ask a lot of questions about the data. Assume it's "laden with caveats and sometimes even errors," as co-author Mike Pell puts it. Look for problems and limitations.

That's it for Story No. 2.

Any questions before we move on?

*How can data that's not about the environment help with environment stories?*

That brings us to the final example: How the Department of Education and the Department of Transportation told us something striking about air pollution.

The invisible hazard afflicting thousands of schools

Across the country, in big cities and small towns, kids attend schools so close to busy roads that traffic exhaust poses a health risk

By Jamie Smith Hopkins 🐦 🔗 email   5:00 am, February 17, 2017 Updated: 2:23 pm, February 20, 2017
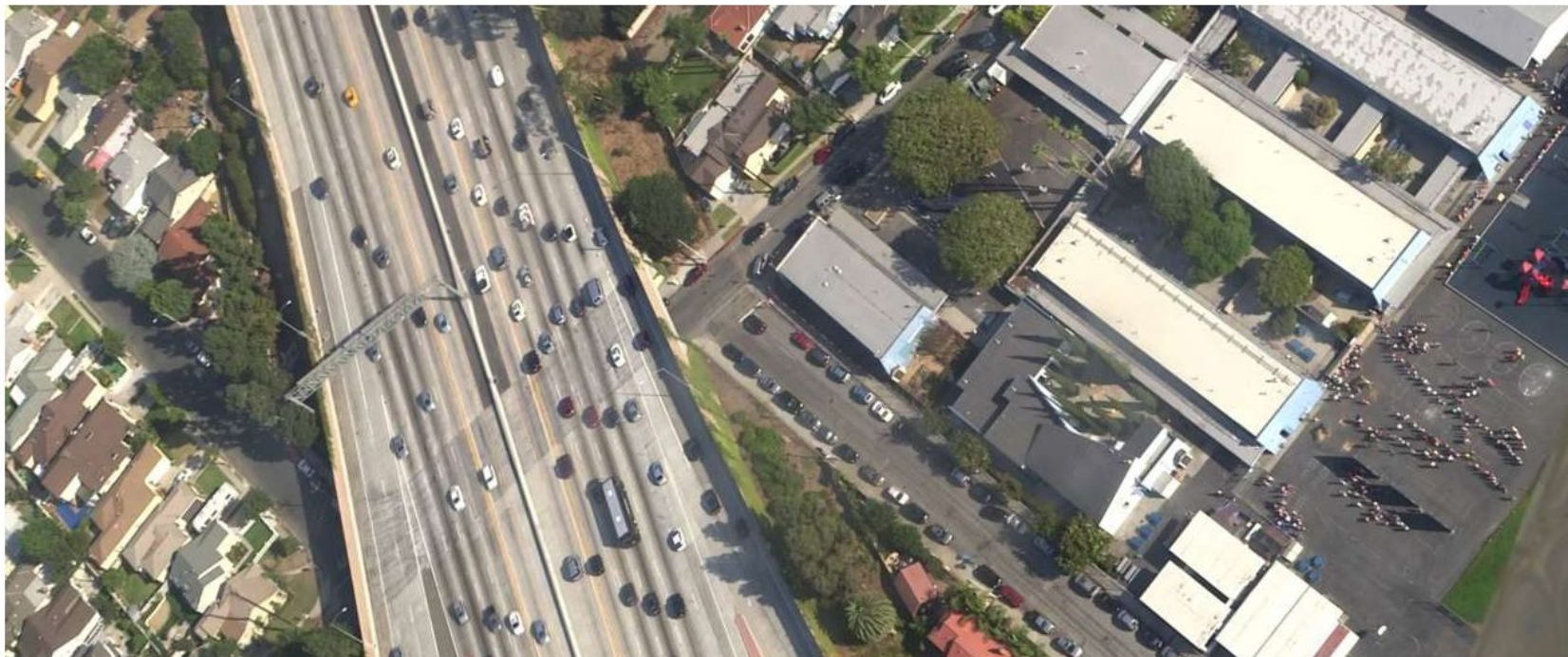
Students line up outside El Marino Language School as vehicles zoom by on Interstate 405 in Culver City, California. Photo courtesy of Stephon Litwinczuk

Data details

- Two datasets, one from DOE's National Center for Education Statistics and the other from DOT's Federal Highway Administration.
- No unique ID to join on this time. Mapping was the key. Which schools are close to busy roads?
- Key finding: "Nearly 8,000 U.S. public schools lie within 500 feet of highways, truck routes and other roads with significant traffic, according to a joint investigation by the Center for Public Integrity and Reveal from The Center for Investigative Reporting."

Lessons learned

- Complex subject? Talk to experts about your data analysis before you get far into it. Ask for help identifying the best data, finding pitfalls and defining the parameters.
- Seemingly simple questions might be hard to answer. How close does a school need to be to a busy road to risk air contamination? How much traffic constitutes a "busy" road with unhealthy air?
- Focusing on two communities in different stages of wrangling with the problem -- particularly if one has found some type of solution -- can help people just starting the same journey.

More lessons learned

If you're in a newsroom, team up with awesome colleagues. This story would not have been possible without the data chops of CPI's Chris Zubak-Skees. (He deserved a byline, but he didn't want to take it.)

He also made the cool and useful interactive you'll see on the next slide.

Search for a school or location

within 500 feet of a road with 30,000-plus vehicles on an average day

within 500 feet of a road with 10,000-plus vehicles and 500-plus trucks on an average day

Click to interact

Montgomery Blair High School
Silver Spring, MD
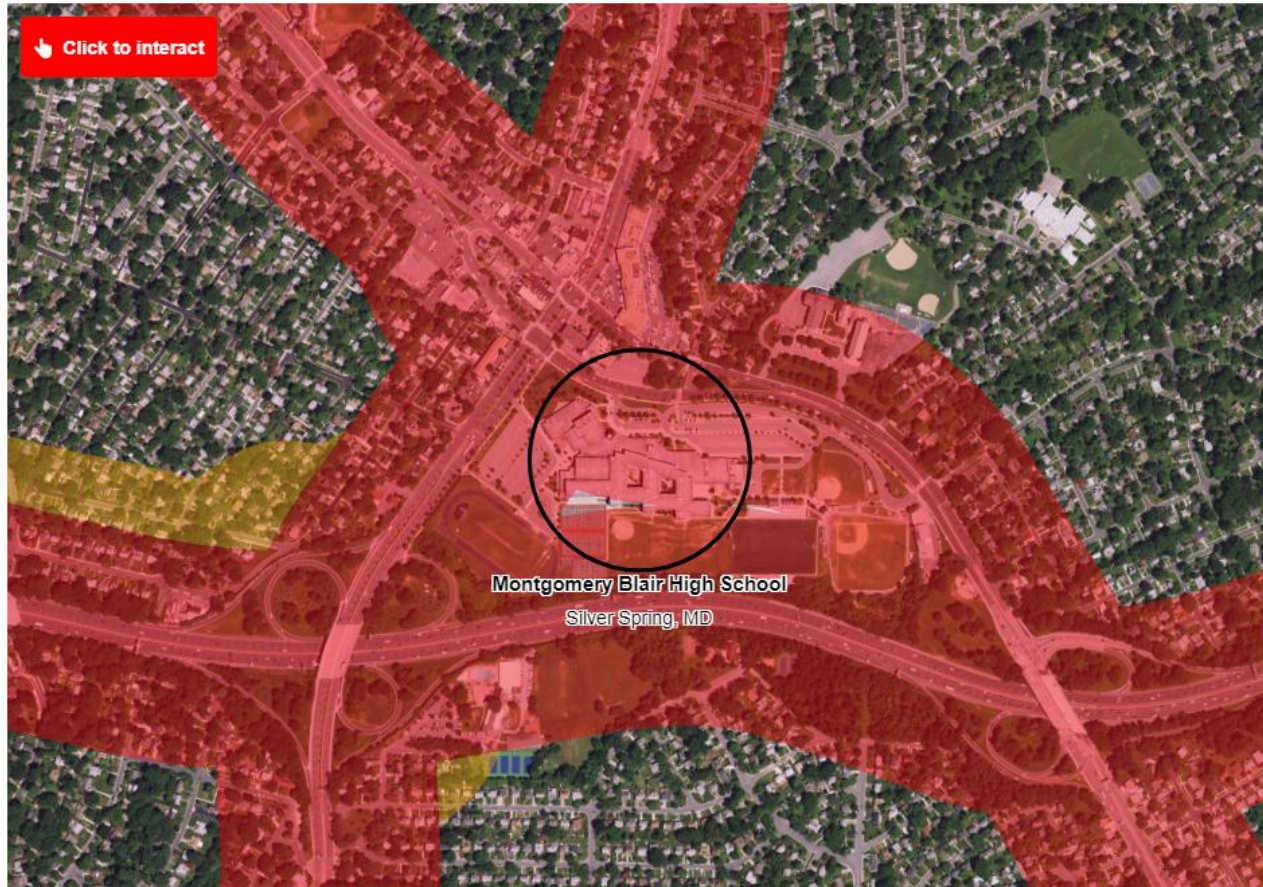
Why to team up with another newsroom, Part II

Reveal, our partner on this story, was very involved in the data-verification effort (cleaning, checking, tossing out problematic datapoints).

Reveal's Eric Sagara conducted the regression analysis that allowed us to see whether race or income affects the likelihood of attending a school beside a busy road.

And the resulting radio piece by Reveal, "School Haze," is a great deep dive. Two stories are better than one.

Q&A time

Questions? Ask away.