Published on *data.nicar.org/uplink* (http://data.nicar.org/uplink)

Risk tool helps show schools' toxic threats

By *Patrick*
Created *04/21/2009 - 11:39*

headlines

Set of fields that compose a story headline
slugdoc:
filename

author details

Fields containing byline, author affiliation, email and bio.
Main author:
Blake Morrison
Main Authors Email:
bmorrison@usatoday.com
USA Today
and Brad Heath
Contributing Author's Email:
bhheath@usatoday.com
bio:

Blake Morrison is an investigative reporter and the deputy enterprise editor at *USA Today*.
Brad Heath is a national reporter at *USA Today*.

Teaser

Teaser:

Outside hundreds of schools across the country, children are exposed to air that appears to be rife with chemicals that can exacerbate asthma or cause cancer.

Outside hundreds of schools across the country, children are exposed to air that appears to be rife with chemicals that can exacerbate asthma or cause cancer.

The reporting that led us to that astounding conclusion began with a relatively straightforward question: What's in the air outside the nation's schools?

To find the answer, we turned to a variety of state and federal databases that pinpoint schools, detail the chemicals released from industrial facilities and estimate the potential severity of toxic air pollution across the country.

Many newspapers have dealt extensively with the government's most basic pollution information that is contained in the Toxics Release Inventory [1], a database published each year by the U.S.

Environmental Protection Agency. The TRI lists the tons of chemicals sent into the air by major industrial plants and government installations. (IRE members may purchase a ready-to-use version of the database [2] from the IRE and NICAR Database Library). Journalists have used it to identify major polluters and to look at local trends. We were able to take those reports much further by turning to an EPA computer model called Risk-Screening Environmental Indicators, or RSEI [3]. It was an essential tool in our research.

The model, which is essentially a computer simulation, is built on the reports companies file with the TRI, but it also weights the chemicals based on the potential harm they can cause. For example, exposure to small amounts of chromium, which is a carcinogen, can have far more severe effects than exposure to similar levels of sulfuric acid. The model also uses EPA air dispersion models to estimate how those chemicals will travel through the air once they're released from the smokestack. The result is an assessment not just of what chemicals are belched into the air each year, but also where they end up and their potential impact.

We combined those data with information about hazardous waste sites, air quality monitors, major polluters and a variety of other records. We found that the air at 435 schools across the nation appeared to be more toxic than the air at a suburban Cincinnati school that was shut down in 2005 because the air was laced with carcinogens at levels 50 times higher than what Ohio deems acceptable. Our experience offers a few key lessons for anyone undertaking a project that relies on rooting through complex — and often highly technical — databases:

1. Outsource what you can.

In our case, we had no choice but to get outside help. The EPA offers a version of its model for free online [4]. However, to do what we wanted, we needed access to a far more detailed set of data produced by the RSEI model. We were able to get them from professors at the Political Economy Research Institute [5] at the University of Massachusetts Amherst, who had spent months and thousands of dollars wringing the data from an EPA contractor.

RSEI works by dividing the nation into a grid, with each cell measuring one square kilometer. For each square kilometer of the country, it produces an estimate of which chemicals were in the air, where those chemicals came from and how harmful they might be. The databases that store those results are huge. To make the job easier, we sent the UMass researchers coordinates for the nation's schools; they gave us the data for the cells that matched those locations.

The experience made converts of us on the value of outsourcing data work. There are tradeoffs, of course. University researchers don't always operate on the same schedule we do. But having their help kept us from having to root through hundreds of gigabytes of technical data. They'd worked with the data before, so they were able to tutor and guide us around landmines. The researchers' participation lent their own credibility to our work.

We relied on that kind of expert guidance - and credibility - again when we monitored the air outside 95 schools around the country. We partnered with scientists from Johns Hopkins University and the University of Maryland, who guided our research and analyzed our results [6]. We couldn't have done it without their help. Even if we could have, no one would have believed us.

2: Share all the data with your team.

Assembling our piles of government data was just the beginning. Next, we had to root through

it. That meant everyone needed to have access.

To make that happen, we built an application on our intranet that the reporters and editors working on the project — and our bosses — could use to query the data, see the details on any school or just rummage around. That let us ask the database hundreds of detailed questions without having to reprogram anything. (The setup wasn't hard, and it didn't cost anything. We built our Web site using free software (MySQL database manager, Apache Web server and PHP) on a spare desktop computer in the newsroom. We combined that with a wiki where we kept track of key documents and findings.)

Sharing the data helped focus our reporting. It gave all of us a better understanding of what the environmental threats were and let us find problems in the data. Also, it led to some of the conclusions that ultimately drove our series of reports.

Another big advantage was that we were able to use our internal site as a prototype and test bed for the database we ultimately published online.

3. Learn the nuances.

All databases have quirks, but models are especially tricky.

Most databases we work with list facts (or at least purport to), such as the details of a particular crime or the cost of a particular mortgage. Models are simulations, and the data you get back show what might have happened or what could have happened. It's easy — maybe even tempting — to take the results and treat them like so many new facts. It's also a good way to get in trouble and reach conclusions you can't support.

We spent months talking to academics, regulators, contractors and others. We learned about the limitations on RSEI; it makes assumptions about how tall smokestacks are and which way the wind blows. The version we used — the latest available at the time — was based on 2005 emission reports, which meant circumstances on the ground could have changed. We learned how academics had used the model in their research and how the EPA uses it.

At first, we thought about doing more. For example, the model's output is so specific that it's mathematically possible to calculate the additional risk a person in any given area would face of contracting cancer based on exposure to levels of toxic chemicals. However, based on everything we learned, we chose a more conservative approach. We ranked the nation's schools based on the levels of toxic chemicals in the air and their potential harm, which is an approach consistent with how the EPA uses the model.

4. Publish your data.

One of the most important steps we took was putting data online. The Web site, www.smokestack.usatoday.com [7], went online Dec. 8, which was the same day our report began appearing in print. The site lets our readers search for information about pollution around 127,800 public, private and parochial schools.

The database made local news out of what would otherwise have been an abstract national story. It gave readers a way to find out how industrial pollution touched their lives and their schools. That had a huge impact. It led to hundreds of stories in newspapers and on television across the country and became fodder for bloggers and activists. Most importantly, it became a vital tool for parents and educators who wanted to know more.

We didn't share everything. Some of the numbers produced by RSEI are too complex to put in front of a general audience, for example. Instead, we tried to simplify what we found; we showed readers how their schools ranked nationally and what chemicals the model said would have been in the air in 2005. A model can never answer the question of whether the air is dangerous, but the data we provided at least gave parents and regulators a starting point for finding out more. Visit our Web site for more about our methods [8].

database education environment open source Web

Links:
[1] http://www.epa.gov/tri/
[2] http://data.nicar.org/node/46
[3] http://www.epa.gov/oppt/rsei/
[4] http://www.epa.gov/oppt/rsei/pubs/get_rsei.html
[5] http://www.peri.umass.edu/
[6] http://www.usatoday.com/news/nation/environment/school-air-snapshotchart.htm
[7] http://www.smokestack.usatoday.com
[8] http://content.usatoday.com/news/nation/environment/smokestack/methodology