



April / May 2012

zData Perspectives

Can Data Quality Be Forced?

by Craig S. Mullins

amazon



DB2 Developer's Guide: A...

\$53.99

Shop now

How many times have you been surfing the web only to encounter a form that requests a slew of personal information before you can continue on? You know what I'm talking about. A company markets a white paper or poll results or something else that intrigues you, so you click on the link, and bang, there you are. You don't have the information you wanted yet, but if you just fill out this form then you'll be redirected to the information.

trying to target prospective customers? After all, that is their job. And the information was evidently interesting enough to get you to click to it, right?

So what is my point? Well, I have a couple of them. The first point is that these web forms need to be more stringently developed. For example, you should never be able to type characters into a phone number field. I'm talking about basic edit checks that every programmer should have been taught to do in Coding 101.

Makes you want to scream, doesn't it? Some folks just shut down their browser or move on to something else. Some folks enter partially accurate information to see how little they need to provide without getting rejected. And some folks just provide bogus information.

You also can check for and reject commonly submitted bogus items. For example, Mickey Mouse will never be your customer. And an address of 1313 Mockingbird Lane may be good for The Munsters, but not your customers. And while you're at it, any phone number with a 555 prefix can be summarily rejected, too.

Now sometimes completely bogus information won't work. Maybe the form requires an e-mail address to which the information will be sent. But hev. that is what Gmail and Yahoo

If you are really interested in accurate data, take the time to do some more robust edit checking. Do the area code and zip

Mail were made for, right? Just create a new address, fill in the form using it, collect the information, then shut down or ignore that e-mail account for the rest of your life.

Of course, if you are trying to sign up for a webinar this might not work because many companies remove generic e-mail addresses. Who can blame them? They are conducting a webinar to drum up business and gather leads. If you provide a generic e-mail address the organizer will assume that you are not a good lead or maybe even a competitor trying to gather intelligence.

Then there is the phone number. I almost never supply an accurate phone number. If the form allows, I type in "do not call me" as my phone number. If not, then I use the information number, 555-1212 (with my area code). I get more than enough cold calls for things I don't need already, thank you.

The point I'm trying to make is that these marketing tactics are responsible for the creation of a lot of bad quality data. But at least some of the data must be useful or the marketers would not use these tactics. And who can fault marketers for actually

code entered actually exist? Do they match the City and State that was entered? For example, if someone enters the 512 area code (Austin, TX) but enters Pittsburgh, PA for the city and state, you know the data is bogus. Or at least suspect... after all, people do move and take their mobile phone number with them. I have a friend who has moved from Chicago to Florida to New York to Texas and he still has a mobile phone with Chicago's 630 area code.

And if you want to go even further you can match up company names to known addresses for that company to verify that an actual, accurate company name is being provided. Of course, there are exceptions here, too. Maybe you work from a home office and you've provided a legitimate address.

The bottom line is that organizations can do better at verifying data in their customer-facing web applications. But even then, you just can't force data quality. There will still be people "out there" (like me) who find ways to enter good enough data that will not have someone e-mailing them or calling them up trying to sell them something all the time. And the data quality fight continues on...

From [zJournal](#), April / May 2012.

© 2012 Craig S. Mullins,

DB2PORTAL.com