

BIG BAD DATA

Move over Smart Grid,- a new buzz word has bumped you out of first place in the utility business - it's Big Data. I'm not sure the term is even grammatically correct. Does it mean more data than you have ever had? Like the term Smart Grid, no one really knows for sure what it means. What we do know is that it's big in the IT business.

I recently interviewed a candidate for a job and asked him what he knew about GIS. In his response he mentioned the term Big Data maybe 10 times and Smart Grid only 7. He didn't say much about GIS. I concluded that Big Data was now bigger than Smart Grid.

So in the utility business, the biggest type of Big Data is all of the interval meter reads they get from the new smart meters, which is really not smart grid, but advanced metering infrastructure (AMI), which is not automatic metering anymore (AMR), which the utility business bumped from being a buzz word several years ago. As you can see it's not easy keeping up with these trends.

The first use of the term Big Data came from Doug Laney, an IT industry analyst. He defined big data as the three "V's". Volume – lots of data. Velocity – data coming in at high speed and Variety – different kinds of data – maybe even unstructured.

Certainly, Advanced Metering Infrastructure (AMI) data begins to qualify as big data. There is a ton of it, it is coming pretty fast and there is a fair amount of variety to it.

I believe the buzz around Big Data has more to do with what you do with the data than the actual volume, velocity or variety of data. If a customer wants to reduce consumption and save money, it's more important to know what's running during a high consumption interval than to have 5 minute granularity of their consumption data.

It seems to me it's all about context.

Big Data plays a strong role in politics. An incumbent Democratic president, running for a second term during a time of huge economic turmoil and high unemployment (inherited from a Republican president) facing a wealthy Republican governor serves as a reminder that Big Data is not about volume, but about context.

Of course that was the presidential election of 1936.

The publication, Literary Digest sent a survey to 10 million voters asking who they were going to vote for. Nearly a quarter responded. The results strongly indicated a win for Alf Landon, the Republican. You may recall from your high school history, that Franklin Roosevelt won the election of 1936 in the biggest landslide in history up to that point. Since then pollsters such as Gallup and others learned that sheer volume of data doesn't predict winners and losers. Context matters. If Literary Digest had asked a few context questions, they would have discovered that the people who returned the survey tended to vote Republican anyway.

This was Big Bad Data.

What does all this have to do with spatial information and GIS? Even in politics, location matters. In fact it matters a lot. If a utility wants to know how to use their AMI data to drive better business decisions, it needs to look at the context of all this consumption data. What's the relationship of the consumption to demographics, of shifting weather patterns, not just temperatures, but shifting populations? What will the impact of certain patterns of consumption be on the low voltage wires and transformers feeding neighborhoods? Where will the adoption of electric vehicles have the most negative impact? How will utilities assess changes in pricing policies? Who will be the winners and losers and more importantly, where will they be and what will they be thinking and doing. Big Data will be a part of that. That context is about location.

Really Big Data is probably coming from social networking. But even that Big Data can be misleading if not put in context. If a utility just looked at tweets from customers for negative comments, they might get just one impression. Not every segment of the population tweets. Utilities can mine that data to figure out just what people and what kinds of people are thinking about and then integrate that data with their GIS. They can associate the negative tweets with spatially enabled customer satisfaction surveys, with places where outages have been frequent, with places where construction activity has inconvenienced customers or where the utility has cut down their favorite trees.

Big Data without context may lead to placing bad bets.

There is no doubt that AMI, SCADA, customer payment data, real time weather data is pretty big. Each has its own focus. The challenge for utilities will be to avoid using Big Data in isolation, but to combine it with all kinds of data to put it all in context.

One way to tame big bad data is to put it in the context of location and that means GIS.