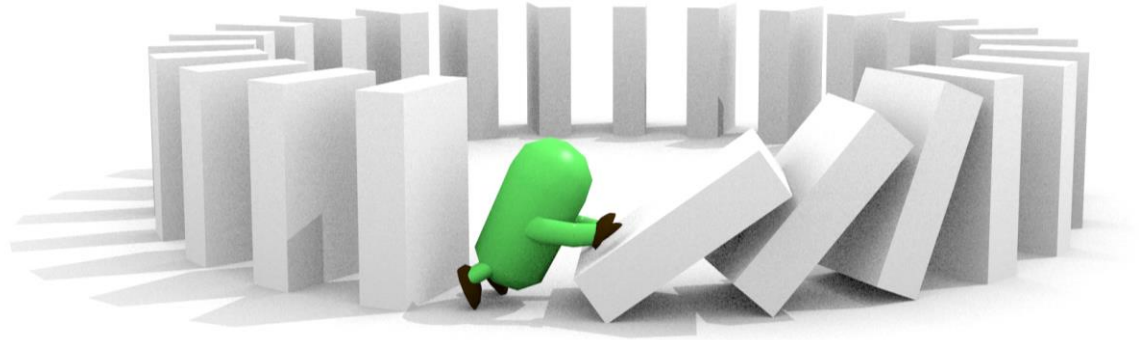# I only changed one line…

When managing data, any data, everyone would, I'm sure, exercise caution about arbitrarily changing values, but a few simple fixes: correcting spellings; removing duplicates; filling in obvious values and so on, that is always a good idea, isn't it? Good data managers always want to be responsive to their users, they are inclined to push through situations, identify the best fix, apply it and move on. Sometimes, however, this type of proactive approach can have unintended consequences. I've recently come across a number of cases where experienced people have allowed themselves to be lulled into making some really stupid mistakes by exactly this type of "just do it" thinking.

The issue is that we are not experts in how the data will eventually be used, the result of apparently minor actions can potentially have surprising consequences. Anyone who's worked with subsurface data has seen quite a few examples, such as where a mistake in spelling a datum caused a seismic survey to be positioned 20km away from its real location, or where mistaking units caused a Wyoming well to show up off the coast of Ghana. Real situations have a tendency to behave in ways that will always come back to bite you.

Passwords on corporate systems are a good example. Naïve IT staff see that short easy to remember passwords are easier to crack so they do the simple thing, they refuse to accept short passwords or those without numbers, mixed case letters and strange symbols. This was taken to the extremes in the case of the Russian spies Richard and Cynthia Murphy who were discovered in 2005. In order to secure one of their systems they had to use a 27 character random password, of course no normal person could ever remember such a long sequence, so they wrote it down on a note which they placed conveniently close to the computer. Well the FBI at least found it convenient when they raided the house and found the piece of paper, as a result the system was easier for US authorities to access than if they had used the string "p4ssw0rd" or their own date of birth.

Experienced data manipulators know that there are going to be times when it is easiest to just fix the value and get on. Most of the time that is exactly the right thing to do, but those of us who've been subjected to Murphy's Law will try and inject a bit of paranoia into the process: always keep a copy of the original data; as far as possible keep note of **all** the minor changes you've made; try and review what you've done with an experienced colleague (it's amazing how often their first question will reveal a better approach or something dubious about a simple step); finally get someone who's an expert to look over at least a sample of your results. It's impossible to do anything worthwhile without making the occasional mistake. The best practitioners know that they will do something stupid eventually, so they make sure they can detect their own idiocy as quickly as possible and then recover from it with grace.