CrossTalk would like to thank 309 SMXG for sponsoring this issue.

Data Mining and Measurements

It was not too long ago that I asked one of our data analysts to see what "gold nuggets" he could find in a metrics data set that had more than 20 years' worth of accumulated software development data that had never been fully analyzed.

I told him that I was uncertain what he would find, but that I bet there was some very valuable information hiding in that data. I even went so far as to tell him that the data was like a big drum of dirt that contained one or more gold nuggets. After some false starts and some frustration, it became apparent that the first problem was to identify what exactly we are looking for in the data. Most people would think the answer to that question would be obvious but quite the contrary was true.

There were in fact some obvious things to look for, but it was not until we started brain storming and coming up with some more obscure ideas and ways to look at the data that we started to learn what we needed to learn from the data set. The better we can define what we are looking for in our data the easier the task of data mining becomes. In addition, the more we understand about what we are looking for the better we can define the metrics that need to be collected for the future.

We have spent a lot of time collecting and analyzing data over the years and it has become clear that we continue to learn about data analysis, data collection and metrics identification but it seems like in many cases it is a slow and difficult process.

The information we mine from our metrics drives improvements in our software development processes but it seems like we can never get enough data to answer all of the questions. This issue of CrossTalk has compiled articles concerning Data Mining and Measurements. I believe these articles can help each of our organizations to make better decisions about data collection, metrics and data analysis.

The articles included hopefully can help each of our organizations learn how to optimize data mining and metrics identification. I hope you enjoy this issue of CrossTalk and I hope it helps your organization to move make significant strides in Data Mining and Measurements.

Karl G. Rogers
Director
309th Software Maintenance Group