The MIT Information Quality Industry Symposium

Phil Teplitzky

Independent Consultant

Title:
Data Quality - When is Good Enough, Enough

Abstract:
One of the major problem-facing practitioners of Data Quality is to know:

1. the Cost Benefit Ratio of Quality, when the knee of the curve has been reached and the cost of a minor increase in Quality becomes excessively expensive and no longer worth the effort

2. What level of Quality is actually needed for the Business Functions at hand? Different levels of materiality exist in the commercial world and the cost to exceed that level of quality may not in fact be worth the effort.

3. That absolute Data Quality, while the Holy Grail of the field, is like the Holy Grail unachievable.

There are several factors at work that, make achieving 100% Data Quality a laudable but unachievable goal, they are:

- Law of Data Uncertainty - this is similar to the Heisenberg Uncertainty, the act of measuring something changes it. We may know the correctness of a data element at a specific time, for a specific domain but not over time for all domains. Mainly because we:
  - Cannot know with certainty that the data has not changed at some time in the future from the time we certified its correctness. For example, we can say that on 10/1/xx the closing price of IBM on the NY Exchange is $50 but we do not know if on 10/2/xx the price will change. Which it often does, corrections are common.
  - That the meaning of a data element changes by the perspective of the person defining it. Closing Price means any of the following depending on who you are:
    - Bid
    - Asked

        - Settled
        - after market trade as of time
    - Perspective is everything - the meaning of data changes based on your perspective and who and what you are doing. In fact, there are many Data Stewards for the same data element and they are all providing definitions of the element using a language that is by its nature ambiguous. There are many definitions and nuances for the same word and we are using words to describe the meaning. Accuracy and precision of meaning decreases as the exponential sum of the ambiguity of the language being used to define it. Even if one takes the alternative of mathematical definition, it is a length between 2 and 4 inches, there is an inherent ambiguity since our ability to measure something is dictated by what level of accuracy and precision we wish to include. It is not 3 inches; it is 3.0001 inches plus or minus .001 inch.

In this presentation, we will address this issue and discuss practical and cost beneficial ways to address them.