

# How reliable is the data in your data warehouse?

Sarah Taghavi

In recent years, there has been a great deal of interest and investment in [data warehouses](#), [data mining](#), and various tools to capture corporate [business intelligence](#). More often than not, the priority is on "out-of-the-box" solutions and rapid development technology.

While it is possible to implement a corporate data warehouse relatively quickly, it is not necessarily true that you have [clean data](#) to extract and use. In fact, there is a greater probability that the data in your warehouse is not reliable or properly integrated. This is most likely due to the fact that the operational data was not reliable in the first place.

I strongly recommend that organizations contemplating building a corporate data warehouse--and those that have experienced failed data warehouse projects--keenly focus on the quality (integrity) of data in their existing operational systems.

What is vital to remember is that practical and robust data warehouses are not about the technology itself. They do not implement the analytical and reporting requirements of your company by installing database engines and polished applications. The tools are certainly worth serious consideration, but they are the lesser part of any workable solution. The workable solution must essentially be about your data and the corporate business rules relating the data elements.

If you find yourself in charge of building the corporate data warehouse, your very first and essential step should be to engage the experienced data architects, business analysts, and subject matter experts. By starting there and by keeping the primary goal in mind (reliable data in your data warehouse), you have started on the right track. As always, you should expect many nay-sayers and those who impatiently wait for the data warehouse to magically and rapidly materialize.

For the majority of companies building data warehouses, the basic motivation is to integrate unreliable operational data created by their poorly designed databases and poorly interfaced application systems. Most data warehouse builders, however, usually start with a different intention: to obtain business intelligence and enable corporate online analytical analysis or reporting. This is not possible without dependable data. When it comes to defining and invoking proper data cleansing and transformation functions, they must ultimately also be experienced in resolving the issue of unreliable corporate data.

If your systems consistently annul and fragment the business rules or if the business rules are not well defined, then no amount of work performed in the staging area (preparing, formatting and cleansing data for data warehouse usage) will successfully result in clean data.

You will soon find that even the cleaned data in the data warehouse cannot be directly and simply fed back into your operational systems. By the time you realize the number of data integrity problems, it will already have cost the corporation a lot of money.

You may think that corporate business intelligence cannot be derivable from fragmented and often ill-defined data structures or application systems. In the absence of proper attention to enterprise data model and data integrity issues, many data warehouse builders are in effect, operating by such a principal. Undertaking one of the most valuable and resource intensive corporate entities--the data warehouse--you have a perfect opportunity to introduce and establish corporate data standards, incorporate enterprise data architecture, clean the data, and thus maximize return on investment.

### **About the Author**

Sarah Taghavi is CEO of [Business Information Engineering Corporation](#).