

10 Characteristics of Successful BI Data Projects

News Flash: Hardware, Tools, Technologies and Platforms don't determine the success of data projects. Every major company, platform and tool set in the data space has both successes and failures on their resume. This means that there must be other factors that influence the success rate. So, what are the characteristics of successful data projects?

We've reviewed hundreds of projects over the past 20+ years, and we've found that all of the successful ones look alike. They have the same characteristics, regardless of industry, technology, tools, or delivery partners. Here are 10 characteristics we always see in the most successful data projects:

Successful BI projects have a clear purpose and decision flow. The broad purpose of the data system is always clearly defined (analytical/operational/both, internal/external/both, new/existing with changes) and it doesn't change. This is different from the business case (see below, ROI). The purpose of the project is set up-front and other decisions (technology stack, tools, partners, etc.) follow later.

They have a data model that models the data and supports the purpose of the project. The data model models your data, not some other company's data and certainly not some amalgamation of concepts of data. Successful data projects always begin with a robust data model based on the data to be loaded, not some pre-set, industry standard abstract data model. The key elements in great data systems are simple: (a) tables and attributes, (b) a list of data sources, (c) a mapping from the data sources to the tables/attributes and (d) a list of processes that create non-fact data. The data models always satisfy the broad purpose of the data system and resolve any challenges with legacy data environments.

At every point, from business users to designers to developers, everyone speaks the same language. The data environment uses the same terms that the business team uses. So, in telephone terms, a rate plan is called a rate plan (not a subscription) and data allowance is called data allowance (not bonus units).

Every data attribute has one, and only one, parent. The parent can be a source system or a process, but there can never be two parents. If two attributes look alike (i.e., billing address) but they came from different sources, they are given different names (i.e., "billing address from web selfservice" vs. "billing address from retail"). Arbitration rules determine which attribute to use. As business rules change over time, the underlining data still exists, providing flexibility over time.

Quality, context and appropriateness are up-front tests of each and every data attribute. Rigorous governance and review processes exist to make sure we are certain that we know what we know. In great projects, one or more team members are tasked with challenging the veracity of every data source, feed, attribute and process. These teams always include a healthy dose of "*Does This Make Sense?*" thinking.

Great data projects use the right amount of data for each task. Many analytical tasks can (and should) be run against samples of the data. A proper sampling technique allows analysts to run many more queries per day, leaving full validation for only those learned insights that suggest operational possibilities. Operational programs, on the other hand, must use much larger data sets including the full data set.

Great data projects include rigorous data ingestion controls as an essential part of built-in quality. Data is either perfect, or it's bad. These controls prevent the loading of bad data. A combination of expected results, field level validations, trending analyses and other techniques are always present in great data projects. In the best of projects, there isn't a place called "suspense".

Great data projects incorporate data vigilance at every stage. Workflow and data audit processes exist to ensure that the basics of data are done right, every hour, every day, every week, every month, every year. Great data projects begin with the understanding that every data environment, left unchecked, will degrade over time. Vigilance is an additional (and mandatory) 'V' word in great data systems. Successfully building to accommodate the other 'V's (volume, velocity, variety, veracity, and value) is important, but the most successful projects put vigilance at the head of the list.

Great BI data projects are based on the understanding that the customer isn't the most important thing...it's the only thing. The system truly knows each and every customer. This one might seem obvious, but one of the leading sources of data system failures is the lack of coherent customer understanding.

Great data projects always show a significant return on investment (ROI). From a business perspective, the ROI of a system is the most important metric that determines the success of a data system. Successful BI projects always deliver measurable operational cost savings and/or provable revenue increases.