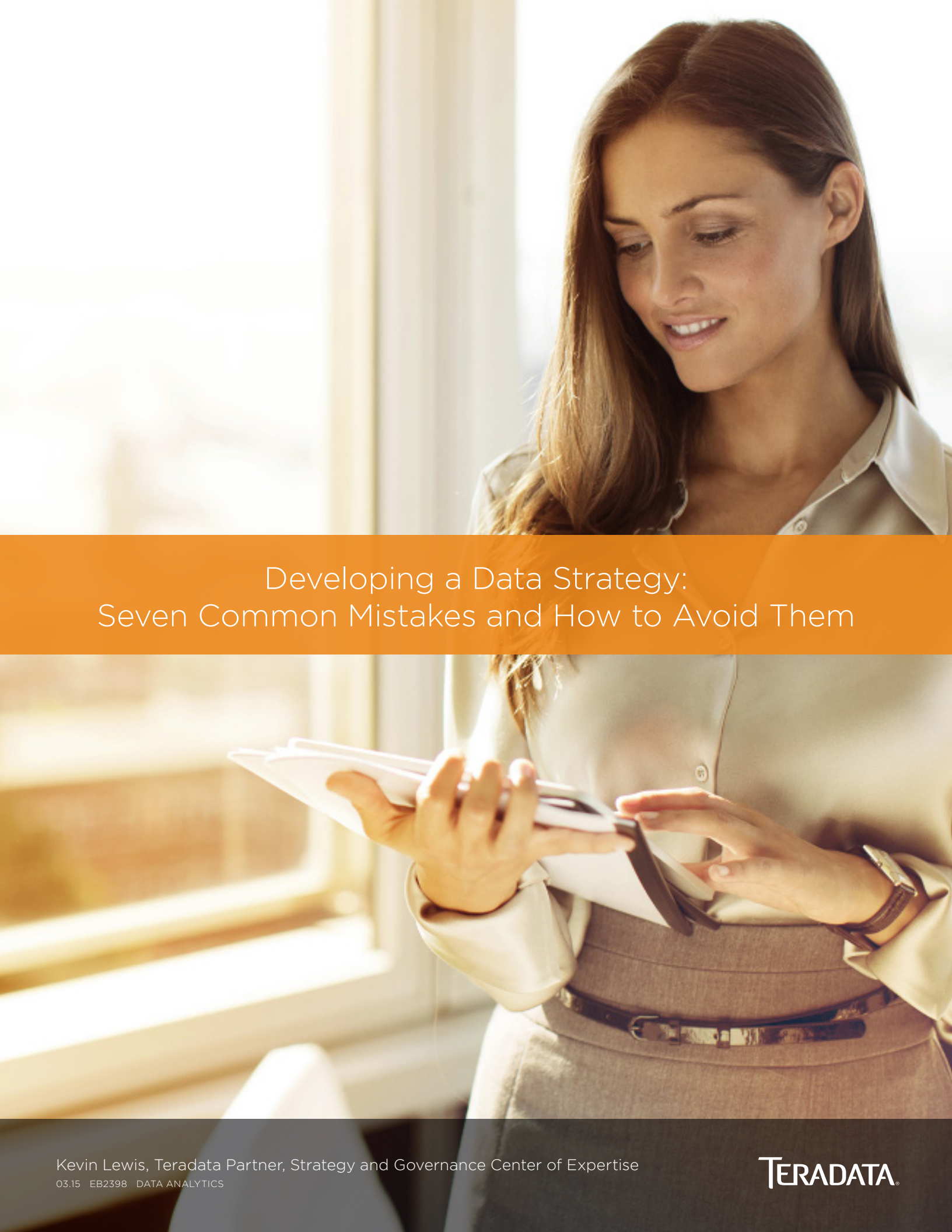

Developing a Data Strategy: Seven Common Mistakes and How to Avoid Them

Kevin Lewis, Teradata Partner, Strategy and Governance Center of Expertise

03.15 EB2398 DATA ANALYTICS

TERADATA.

Table of Contents

Introduction

Many companies have serious challenges with their data management programs, but are looking in the wrong places for the solutions. For example, I'm often asked by an IT manager, "How do I get the business excited about data management?" My answer is simple. You can't. Don't even try. You can only get them excited about their own goals and how you can help them achieve those goals.

Some of these companies have a general sense that things are just not moving in the right direction—there may be morale issues, cost overruns, a poor relationship between IT and the business, and other issues. Sometimes the situation is more serious. For example, in some cases, data and analytics programs of various kinds (data warehouse implementations, data governance programs, etc.) have spent millions, only to be suddenly cancelled—without much argument from the people in the business who were supposed to obtain value from these programs.

Having worked over the past several years with dozens of companies across multiple industries, I've observed a handful of problems over and over again that—in one form or another—are almost always at the root of these challenges.

In this white paper, I'll describe seven common strategic mistakes organizations make when creating a data and analytics program . Then I'll prescribe four steps to get the program on track either from the outset or when the symptoms are serious enough to reconsider the overall approach. The key is to stay connected to important business initiatives. That advice may seem obvious, but you will see in the following mistakes that there are ways that data and analytics leaders can be convinced that they are adequately connected to the business when in fact they are not, and this makes all the difference.

Strategic Mistake #1 – Having No Plan for Data Management Across Business Initiatives

This problem is the most familiar so we will address it first. It is the problem we in the data management industry have been focusing on for years. Organizations initiate individual projects that address specific business objectives, such as improving targeted marketing, optimizing inventory levels, or reducing customer churn. These projects may have a fair amount of success, but over time, they collectively produce an environment in which data is scattered everywhere. Different initiatives have different requirements, so separate teams build entire solutions independently, from the ground up. For example, the Sales data requirements for a marketing initiative are slightly different than the Sales data requirements for an inventory replenishment project. In response to these requirements, the organization builds two (and eventually more) independent feeds of Sales data to separate databases and applications, along with all the necessary data modeling, data quality checks, technology infrastructure, and so on. The result is excessive technology costs, the inability to analyze data across data domains, slow delivery times, and bewildering complexity. In this situation, organizations simply don't want to go through the trouble of sharing data, or they feel data sharing is unrealistic because of project schedule pressures and the complexities of considering enterprise needs.

Sharing data is like sharing a house. If my wife and I were building a new house and couldn't agree on what the kitchen should look like, we wouldn't build two kitchens! We would figure out how to meet both of our needs with the same kitchen. It would be ridiculous and irresponsible to do otherwise. But companies build two, three, or twenty kitchens in the same house all the time.

Unfortunately, as organizations attempt to recover from Strategic Mistake #1, they often end up creating new problems. As soon as the organization begins to think in terms of "enterprise," they usually over-correct by distancing themselves from the business initiatives that require the data. After all, too much attention to individual projects caused this problem in the first place, right? As a result, the following mistakes are now becoming just as common, and are even more insidious; the people involved realize something is not quite right, but can't articulate exactly what it is.

Not having a data management plan leads to:

- Increased costs
- Ineffective ability to analyze data across systems
- Slow data delivery
- Increased complexity

Strategic Mistake #2 – Implementing Buzzwords Without Linking to Business Initiatives

When I visit an organization, a data and analytics leader will often tell me their organization already has a plan in place to implement data management capabilities. The plan will show all the right functions: data quality management to be implemented by February, metadata management by June, data stewardship in August, etc. However, when I ask several different ways what the drivers are for these capabilities, there's a struggle for a good answer. Although masked as real business needs, in reality, the drivers are the capabilities themselves. For example, it may be that there was a previous assessment or audit of some sort that indicated (correctly) that data quality management is inadequate. But what specific business issues were resulting from the data issues? And more to the point, what important projects were at risk if data quality problems are not addressed? Without clearly identifying how data management capabilities enable important business initiatives, then these proposed practices are the proverbial solutions in search of a problem.

In some cases, the data management program (or data warehousing, business intelligence or big data) is itself a major initiative and identified as strategic. But what is driving that initiative? It should be something related to what the company does for a living, such as selling shirts or reducing the cost of selling shirts. Sometimes when I look at the internal slide ware being used to "sell" data and analytics, if it weren't for the company logo, I wouldn't be able to tell whether I'm working with a biotechnology research organization, a railroad, or a jelly bean wholesaler. That's a problem.

> "When a project is focused on implementing a data domain from A to Z without any specific business uses to drive the scope... it is a lumbering giant, headed resolutely toward a sheer cliff."

Strategic Mistake #3 – Implementing Data Domains Without Linking to Business Initiatives

When I see this mistake, organizations have usually started to develop their strategy based on real business needs. By gaining an understanding of these needs, they settle on one or more data domains that are important to the business. So far so good. But the problem results when those business needs are either not the right ones (see Strategic Mistake #4) or they are not used to drive the scoping and implementation of the selected data domains. Once the first project begins, all attention is concentrated on the data itself, not on the planned uses of the data. For example, a company may accurately identify Customer data as very important. But which transactions are the most important to link to the customer first, second, and third? And there may be a thousand or more attributes of a customer; which ones should be implemented now and which can be deferred? When a project is focused on implementing a data domain from A to Z without any specific business uses to drive the scope, then the project takes much longer, costs much more, and has a very good chance of being cancelled before it is completed. It is a lumbering giant, headed resolutely toward a sheer cliff.

In this situation, I (carefully) ask this question: If this data and analytics initiative were to be cancelled right now, who would be upset? Will a sponsor of an important business initiative say something like, "We are counting on the implementation of that data! If we don't get it, the entire supply chain program is at risk! You can't cancel it!" In many cases, executives would instead say, "Well, it seemed like a pretty good idea, but it doesn't really affect our major programs. We can live without it."

Strategic Mistake #4 – Looking in the Wrong Places for Business Value

Not all business drivers are created equal. Strategic business initiatives are by definition more important than the wish list of business analysts. It is admirable that data professionals interview business analysts to understand their need for information and the challenges they have to overcome on a daily basis. In these conversations, they will hear about various challenges of assembling and integrating data, managing data with suspect quality, and

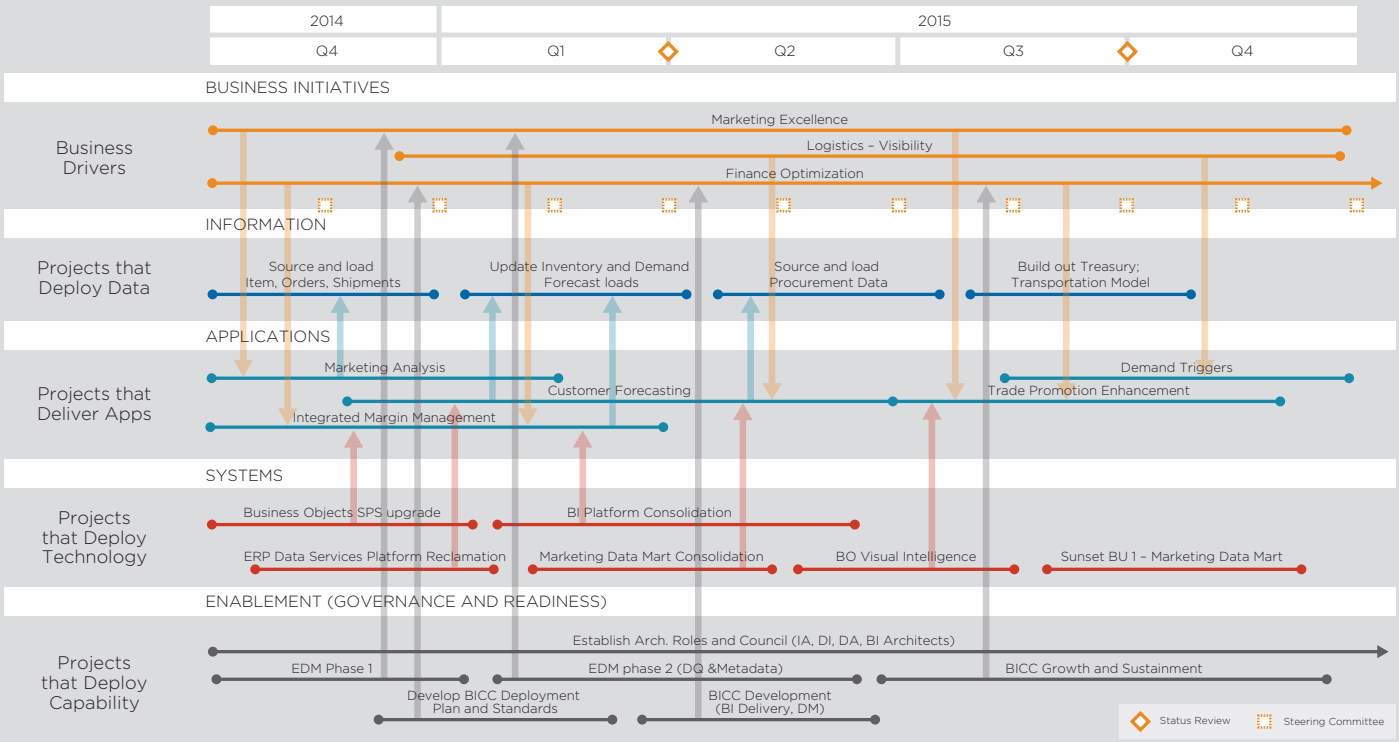

Figure 1. Analytic Roadmap Example.

organizing and presenting the data to decision makers. They may even take the step to document a respectable return on investment (ROI) for new analytical requirements and then make a proposal for funding. The problem is that these projects just add to the stack of funding requests—most of which have more than adequate ROI identified or have other strategic drivers. When data deployment projects are proposed as separate and independent—though important—initiatives, they are less likely to be given priority.

I am not saying that understanding the needs of business analysts is bad. Instead, I'm saying that you should favor the business analyst needs that are in support of the already-planned (and funded) business initiatives. The question to ask business leaders is not, "What data do you need and what is the value?" The appropriate question is "What are your planned and funded business initiatives and what data and analytics will be needed for their success?"

Strategic Mistake #5 – Using Current Systems as the Basis for a Roadmap

Often, a motivation for building a data and analytics program is to either consolidate existing data marts or to integrate data from a specific set of core transactional applications. There is nothing wrong with this goal, per se, but the problem occurs when these current systems become the primary basis for the data and analytics program roadmap, and the desire to eliminate data marts or integrate source data becomes a goal in itself, with little direct relationship to business initiatives and objectives.

With a plan like this in place, the data and analytics team usually straddles the fence between genuinely integrating data and just porting or copying the data from existing systems. When the metric of success is simply hitting the "done" button for as many sources and data marts as possible, then, predictably, genuine integration is usually sacrificed in order to claim the metric is moving in the right direction. Without real business drivers (i.e. business initiatives that will make use of the data for real business purposes) there is no rational basis for prioritizing the

projects or the detailed work within projects. There is usually a thin veneer of a business case—teams will check with some end users to see what they "want" in terms of data attributes and data quality as well as what they might need at some unidentified point in the future. But that is very different from meeting the needs of high priority business initiatives.

There are, however, exceptions. For example, if performance or availability of existing data marts is having an identifiable and direct impact on business goals, or if the current hardware and software costs are excessive, then consolidating data marts might make sense. But if these are truly the drivers, then they should be used to prioritize the roadmap and the approach. Which data marts have availability issues that are having the largest business impact? And with these drivers alone, there is really no business case for data integration, so it would be better to simply "forklift" the data marts to a new environment with better availability. That is, there should not be a mismatch between the actual, identified business goals and the roadmap, however well intentioned.

Strategic Mistake #6 – Being Content with Executive Sponsorship

Many books and articles (on most any business or technology topic) will tell you that executive sponsorship is the most important success factor. And it certainly is important. But executive sponsorship can be a mixed blessing. If you are suffering from one of the other indicated mistakes, then executive sponsorship may not help. In fact, it may even hurt. An executive sponsor can keep a doomed program going forward in spite of strategic issues. For example, a politically powerful CFO who has a number of analysts in his or her area may understand the challenges of disparate data and be very much in favor of a remedy. He or she may promote the (important but vague) idea of "consistent definitions" or "integrated data" and may even get the data management program itself elevated to the level of a top strategic initiative. But if that initiative is still focused on the buzzwords, the data itself, current systems, or a value proposition that is not directly tied to core strategic initiatives related to the real business, then the program will likely live well beyond its natural expiration date exclusively because of who is behind it. And what happens when this executive leaves the company, changes jobs, or becomes enamored with some other initiative? At least the end will come quickly and mercifully.

Strategic Mistake #7 – Confusing Analytic R&D with IT Projects and Operations

When I visit client organizations, a frightening number will have at least one serious person saying something like, "I really don't think 'data warehousing' makes sense any more. It takes too long. We should put all our data in Hadoop and let our end users access whatever they want." It is indeed a great idea to establish an environment that enables exploration and quick-turnaround analysis against raw data and production data. But to position this approach as the primary data and analytics strategy of the organization is nothing short of professional malpractice.

The problem is that people are confusing experimentation with IT projects. There is a place for both, and there always has been. Experimentation (or discovery, research, ad-hoc analysis, or whatever term you wish to use) should have lightweight processes and data management practices—it requires prioritization of analysis activity, security and privacy policies and implementation, some understanding of available data, and so on, but it should not be overburdened with the typical rigor required of projects that are building solutions destined for production. But once a prototype is ready to be used on a regular basis for important business functions, that solution should be built through a rigorous IT project leveraging an appropriate solution development life cycle (SDLC), along with a comprehensive enterprise architecture plan including, yes, a data warehouse that provides integrated, shared, and trusted production data. Organizations often abandon the rigor of IT projects stating that it is no longer needed, when in reality it is simply a way to avoid the hard work of improving scoping and implementation practices in need of repair since long before the advent of "big data".

> Organizations that confuse analytic R&D and IT Operations risk a disconnect among IT and business expectations, and even worse, failed business results.

The Solution

There are four steps that can be taken by any organization to get the data and analytics program on the right track. The details may differ by company, but at the heart of all this is one idea: examine the strategies and initiatives that are already important to the business and show how your program will help them—not how it will theoretically, potentially, maybe one day help them, but how it—if approved—will help them.

Step 1 – Gain the *Right Kind* of Sponsorship

For this first step, consider the past, present, and future. What real issues and opportunities can you point out? Looking at the past,reflect on the business application projects (within important business initiatives) that have been done independently and explain how this has resulted in disparate data and why that matters from a business perspective. Point out the real problems that have resulted. Are there issues identifying all the places where personally identifiable information (PII) needs to be protected? How many different extract, transform, and load (ETL) processes have been built to collect the same data from the same sources to different targets in slightly different form? How long did these projects take? How much did they cost? What if these projects had planned to share data—what savings could have been obtained? How much faster would projects (within real business initiatives) have been after the initial implementation of shared data?

Then look at the present. What projects (any project, not just data and analytics projects) are happening right now that are duplicating effort and exacerbating the problem of disparate data? Are there projects that are currently building processes to collect the same old data from the same old sources yet again? (Are you building your twentieth slightly different kitchen in the same house?)

Now consider the future. Look for the corporate strategies and planned enterprise, departmental, or business unit initiatives. Are your business executives and their partners in IT asked to produce solution roadmaps of some kind? I have seen these go by many names, such as strategic plans, functional roadmaps, long range plans, etc. Also examine the project portfolio, which is usually re-evaluated in an annual funding process. (Incidentally, this is the time to influence these business initiatives by advocating for the benefits of new technologies and techniques where you believe they can have an impact on objectives. But if these ideas don't get incorporated into funded business initiatives, put them to the side and focus on what does make the cut.)

While examining these initiatives, find which initiatives, if left on their own, will deploy the same data in different form, along with the same data management functions, such as data quality management and master data management (whether these terms are used or not). Don't wait for the requests for projects to come to you. It is your job to get involved at this level of planning and assert the role of shared data and data management practices.

The need for data could come in the form of enabling an application (e.g., feeding manufacturing data to a labor planning application) or analysis to support an initiative (e.g., analyzing the impact of a new manufacturing process on productivity). At this point, you don't need to drive out all the details. You are only looking for enough examples to gain sponsorship and to develop the discipline of linking to business initiatives. In this way, you are primarily "sponsored" by the funded and important business initiatives. With these benefits identified, a named sponsor of the data management program itself is a very nice bonus, but it's a lousy substitute for being needed by a top initiative.

Step 2 – Develop an Analytic Roadmap Linked to Business Initiatives

With sponsorship and some specific initiatives to consider, it is now time to develop a detailed program plan. This is where you thoroughly examine the current and future business initiatives and determine which data is most important to deploy and in what order. Consider which applications will leverage the data, what systems (infrastructure) is needed, and what enabling data management capabilities will be required. The criteria do not have to be rigorously scientific, but should include the timing and value of the initiatives, the quality and integrity of source data, and the number of initiatives requiring the same data.

When the analytic roadmap is created, it is crucial to keep the link to the business initiatives. Remember it is not enough to simply use the information collected here to get funding to deploy key data domains. The initiatives must be used to scope the projects, so that you are not just delivering Customer data, for example, simply because you have shown that it is really important. Instead, you are delivering Customer data to support the One-to-One Marketing, Call Center Optimization, and the Social Media Sentiment Analysis initiatives, assuming these initiatives are linked to direct business value and are already approved and funded.

You may be concerned that if you take this approach, you will be tied too closely to these initiatives, and you will find yourself right back at Strategic Mistake #1. That is, the design of the data will not be suitable for an enterprise approach. This is a fair concern and a common reason why people tend to over-correct. But there is an appropriate middle ground. By focusing on multiple initiatives, funding the data deployment through a corporate budget, and following proven architecture and design principles (e.g., modeling at the lowest level of detail, obtaining data from original source, building flexible "right-time" ETL processes, etc.) you can deliver only the data needed in the near term while enabling extensibility and scalability for the medium and long term. You should not fear additional work later. You should only fear excessive re-work later.

Step 3 – Align Enabling Data Management Capabilities with the Analytic Roadmap

In this step, you determine which data management capabilities are needed and for what purpose. Some data management capabilities—at least at a minimal level of maturity—will be needed regardless of the initiatives and business goals. For example, data profiling will be needed to evaluate and analyze the quality and demographics of data from identified sources. However, now that you know your business goals, you can determine which data quality issues are most important so the data profiling can have some direction. You will not solve all data quality problems. Because of the focus on business initiatives, you now know that good quality data means the data is suitable for the targeted initiatives. A data quality issue is defined as one that will have an adverse effect on the initiatives, not just that the data isn't as it should be. A reporting application that produces financial reports will have a different set of data quality requirements than a sales forecasting application for automated inventory replenishment. Focus on the quality you need, when you need it.

Enable long term extensibility and scalability by:

- Focusing on multiple initiatives
- Funding data deployment through a corporate budget
- Following a set of design principles such as modeling at the lowest level of detail

Other specific data management capabilities will be required depending on the initiatives being served and the state of the data in the organization. For example, if you are deploying Customer data and it is found that to meet the needs of the near term initiatives you will need to reconcile data from multiple sources, then you may want to propose a master data management (MDM) initiative, starting with Customer data. In this way, the capability (MDM in this case) has meaning and value and is not done for its own sake with only a vague sense of potential value. It is done to make already-planned initiatives successful, with Customer attributes appropriately scoped and prioritized.

Now you have an integrated plan that includes business initiatives, the shared data needed to support those initiatives, and the data management capabilities to manage the data effectively—all linked to real business value that would be hard to dispute, because you are supporting the initiatives that have already run the gauntlet required for funding approval. You also have a basis for making good

decisions on the appropriate systems or infrastructure projects to include on the roadmap, based on planned applications, data types, and basic business volumetrics—not just guesses and assumptions.

Step 4 – Implement the Roadmap

Here, the analytic roadmap is implemented step by step. The roadmap is a series of projects with interdependencies and should be managed as such. If you have done your job right, then other programs will see your data and analytics program as a dependency. That's a good thing because it means your program is needed. A delay in your data and analytics program should concern someone in the business. Cancelling the program should be unthinkable.

As new enabling capabilities are incrementally developed, they should be embedded within the processes of the organization so effective data management becomes a natural part of the way business is done. For example, a regular review of common data needs should be aligned with functions such as enterprise architecture, strategic planning, and the annual funding process. If there is an architecture review board (ARB) then the goal of shared data and data management capabilities should be enforced in this function. Also, data management functions should be added to the standard development SDLC so that, for example, data profiling is just another task expected on every project that integrates data from one or more sources.

Conclusion

In my experience, people involved in the data management profession are genuinely trying hard to remedy the challenge of disparate data that has proliferated over many years. People in our profession tend to have a certain level of dissatisfaction with the status quo. We don't like to see the same solutions created over and over again with limited, if any, new value. We like to think of the big picture; the long term. While appropriately pushing for an enterprise focus, we just have to stay grounded and not forget our real business.

Kevin Lewis, Teradata
Partner, Strategy and Governance Center of Expertise
Kevin.Lewis2@teradata.com

10000 Innovation Drive, Dayton, OH 45342 **Teradata.com**

03.15 EB2398