

Implementing a Data Management Methodology:

A Do-It-Yourself Guide for High-Value Data Across the Enterprise

A DataFlux White Paper



YOUR DATA.
YOUR BUSINESS.
ONE SOLUTION.



Since the economic downturn of 2008-2009, organizations worldwide have scrambled to reorient and reconfigure their operations to maximize existing processes in the face of slowing sales. To do this, businesses turned to every asset – including production equipment, personnel and facilities – to find ways to maximize revenue, reduce costs and mitigate risks.

For many businesses, another corporate asset – data – has become a target in the search for a more profitable organization. But, the dynamics of data management are more chaotic than ever. Data is now collected and saved from every conceivable source – internet applications, front-office and back-office systems, trading networks, social media – and this complexity requires companies to have a sophisticated, deliberate process for managing this vital information. After all, data holds the key to sales, marketing, customer support, production and other initiatives. Without an accurate view of customers, products, materials, locations and assets, how can a company compete in today's marketplace?

Because of these factors, the need for data management has never been higher. This paper will explore a new methodology for integrating data management principles into the organization. Through this proven lifecycle for data management, companies have the ability to create more accurate, integrated and controlled data to support every part of the business.

The Business Case for Data Management

In a 2009 Gartner survey, the research firm found that the average organization loses \$8.2 million annually from poor-quality data. Further, of the 140 companies surveyed, 22% estimated their annual losses resulting from bad data at \$20 million – and 4% put that figure at a staggering \$100+ million.¹

For companies who are facing data management problems, the primary question seems to be, "Where do we start?" As with any corporate initiative, the best starting point is to set measurable business goals. After all, no one has ever managed, moved or standardized data for the sheer thrill of the task. There must be a business reason to conduct any data management initiative.

At the highest level, there are three primary reasons why organizations perform any business function: to stay out of trouble, to make money or to spend less money. To translate that into more business-friendly terminology, these three reasons are:

- **Governance, Risk and Compliance** – Better data creates a more accurate view of the organization to help understand when the company is at risk from failing to meet regulatory requirements, complying with industry standards and other external and internal pressures.

¹ SearchDataManagement.com, Aug. 25, 2009. "Poor data quality costing companies millions of dollars annually" by Jeff Kelly.

- **Business and revenue optimization** – Better data makes the company more profitable by improving business processes that support every phase of the business.
- **Cost control** – Better data supports increased efficiency throughout the organization, driving down the costs of both direct (related to production) and indirect (personnel and other administrative) costs.

The key for data management professionals is to tie the data management efforts to each of these business objectives. The next few sections will discuss how data management technologies and processes map to these broader business objectives.

Governance, Risk and Compliance (GRC)

The need for a GRC strategy is critical for every industry. And often, data management is done in a very piecemeal way. For example, many retail banks have product-oriented risk management systems, such as a risk management system for loans, one for credit cards, one for banking (savings and accounts) and still another for mortgages. There is often no single, integrated, customer-oriented risk management system to calculate the true risk exposure of every customer for all credit-risk products that a customer holds.

In this scenario, if a customer fails to make a loan payment, it can often take days – or perhaps several weeks – to lower the credit limits on credit cards held by the same customer. Because there is no single, integrated, operational view of customer and product data, the bank risks an increase in bad debt. However, with a single view of a customer, banks can easily determine the total exposure of each customer with visibility into all transactions and how the transactions are related. This is just one example of how the data management function can impact risk management – and vice versa.

The first principle in a data-oriented GRC strategy is the establishment of **data governance**. Data governance is a corporate initiative that ensures that important data assets are formally managed throughout the enterprise. By implementing sound data governance practices, a company can extend a single set of business rules across all applications, platforms and business units – and apply corporate standards to all data sources. Enterprise data quality and business rules monitoring technologies are critical data management tools that enforce data governance standards in today's organizations.

Another risk management strategy focuses on **compliance** with both external regulations (e.g., Sarbanes-Oxley and Basel II) and internal policy requirements (e.g., corporate standards for mergers and acquisitions). This form of risk management seeks to keep the organization within the guidelines of these requirements.

For data management professionals, the tenets of compliance are nothing new. However, the breadth and depth of compliance regulations are on the rise. Organizations can deploy entity resolution technology, for example, to alert the company if it is doing business with a known or suspected terrorist, criminal or fraudster. In other cases, business rules management can ensure that a company has an accurate and timely view of data across sources. This is vital for fulfilling financial reporting regulations such as Sarbanes-Oxley, Basel II, Solvency Act and other oversight requirements.

Risk mitigation is another key piece of a GRC effort. This effort can overlap with the compliance initiatives (am I doing business with someone who is also on a watch list of fraudsters?), but it can take other shapes as well. Companies worldwide must find ways to uncover duplicate or spurious transactions to eliminate fraudulent spending. Organizations also need to examine M&A activity to make sure the company is making the right decisions about new customer channels. Each of these activities exposes the company to risk, whether from regulatory fines or a poor-performing business.

Entity resolution is often a critical component of the data management strategy for risk management. With this technology, organizations can match data from multiple systems to automatically flag transactions that require manual intervention. Many organizations – particularly in insurance, healthcare and other highly regulated industries – have a complex reporting matrix to rationalize revenue, and the complexity can be exploited in the form of fraudulent transactions. Entity resolution, data quality and data integration tools are also used to help better understand all facets of the organization, leading to a better picture to mitigate risks.

Business and Revenue Optimization

Every company wants to find ways to do business more effectively and efficiently. There are several business initiatives that fall under the goal of business and revenue optimization, which can align the company for success with a sound data management strategy.

The first area of focus is **customer satisfaction and loyalty**. In most organizations, every interaction with the customer – from store purchases to website visits and support calls – creates a data trail that can increase the company's knowledge of the customer. Unfortunately, this data often ends up in disparate, disconnected systems. Sales and marketing may use one system, while billing and logistics stores information in another. This creates multiple, incomplete views of the customer that can foil retention efforts, cross-sell/up-sell campaigns and other customer-facing initiatives.

For data management professionals, the key is to build information that can support enhanced customer service as well as create a single customer view to facilitate more profitable relationships. Master data management (MDM) is one preferred data management strategy that can be used to create that single, unified customer view. Similarly, data warehousing, data migration and data consolidation efforts can help

intelligently integrate data within a single application to provide a more coherent view of the customer.

The next area that can improve business effectiveness is **business agility**. This is a broad segment that encompasses methods to understand market dynamics, customer buying patterns and competitive pressures. Through more consistent and accurate data, organizations can explore new markets, better understand market dynamics and uncover customer needs.

In the data management arena, efforts that further the goals of business agility include synchronizing data from different applications as well as physically migrating data from multiple sources. Additionally, companies can benefit from creating business rules that can enforce standards for new and existing data—enabling better business decisions in both operational and analytical environments.

The final area in this category is aimed squarely at the practice of **revenue enhancement**. Most organizations struggle with inaccurate billing or invoicing. Disparate and inaccurate data often hinders the efforts of accounts receivable, billing and logistics. The existence of multiple customer identities, in different business systems, is a persistent problem that can foil customer acquisition effort. By reconciling those views through entity resolution, the organization can have one view of the customer – and support better accounting and billing processes. The company can also turn to MDM technologies to improve the accuracy and consistency of customer data across sources, helping streamline back-office processes and applications to improve performance of capital assets.

Cost Control

Organizations can increase profitability either by increasing revenue or by decreasing the cost of doing business. To control costs, businesses must identify and reduce wasteful spending – or find ways to “do more with less.” These goals are very hard to accomplish without good data across the organization.

For example, manufacturing operations are highly dependent on shared information. The customer data created in sales is needed in customer service, marketing, finance and distribution. Product data is needed in manufacturing, development, planning and stores. Order data is needed throughout the enterprise. Maintaining consistency and preventing conflicts to this data, across operational applications and processes, is critical to business operations. Basic errors in the data chain can lead to manufacturing errors, additional distribution costs, oversupply in distribution centers and, ultimately, customer defection.

There are a variety of cost containment objectives that better data can directly support. The first is **modernization**, which aims to find ways to optimize existing processes and procedures to reduce costs on a cross-functional basis. This initiative can lead to a host of tasks, but companies often focus on centralizing applications from a multitude of business units. Modernization can also involve building more

cross-functional teams and strategies to reduce wasted spending and improve resource allocation.

Data management technologies and practices can support modernization by facilitating the migration of data sources during the consolidation of multiple applications to a single enterprise source. There is also a need to extend these capabilities to real-time views via data synchronization, which enables integrated, high-quality data to more effectively support business functions.

Companies looking for cost containment also focus on **productivity**. Here, the organization looks to increase efficiency and accuracy of business processes and personnel. For large organizations with global activities underway, there are often multiple, overlapping business functions going on in parallel. By focusing on the data that drives these business functions – and ways to share and synchronize information across these sources – an organization can become more productive and drive down costs.

Data management professionals have the expertise and tools to provide more immediate, accurate views of corporate information to support business activities. An enterprise data quality or data governance program can put data standards in place, regardless of the originating source application. For example, if an insurance firm has a business unit for claims and another for underwriting, both teams may need to use a separate application to conduct business. However, by writing data quality or data governance rules to monitor the data in those applications, a corporate-level group can see when new data does not conform to established standards, allowing the firm to identify potentially spurious data before it inhibits workflows and time horizons.

The final elements for cost containment are **spend management** and **supply chain optimization**. The processes that support production – from buying materials and spare parts through the delivery of the finished product – is a critical element of any cost control effort. Companies need to find ways to reduce spend and expand negotiating power with suppliers through intelligent vendor consolidation. They must also seek methods to support just-in-time manufacturing by streamlining information on parts, inventory and supplies.

Various data management strategies can help improve cost control initiatives. Data consolidation and data synchronization efforts help create a more cross-functional view of information. MDM efforts take that a step further, providing a single source for customer, product, materials or asset data. The principles of data governance and data quality covered earlier are vital, as these topics have typically received little-to-no attention from an enterprise-level data management strategy. However, by managing this data more effectively, the organization can achieve tangible cost savings.

The Data Management Methodology

Back to the primary questions that companies have: “Where do we start?” Organizations must approach data management in the same fashion that they manage any process – with a well defined, repeatable methodology. In the past, data was considered to be a byproduct of IT systems. Data management elevates the role of data to a key business asset that can help achieve significant business goals. To accomplish this, you need a data management lifecycle methodology to manage, monitor and maintain data to benefit every phase of the business. This six-phase process is:

- Define
- Discover
- Design
- Execute
- Evaluate
- Control

The next few sections will examine how this process organizes the work of people and leverages technologies throughout the cycle.

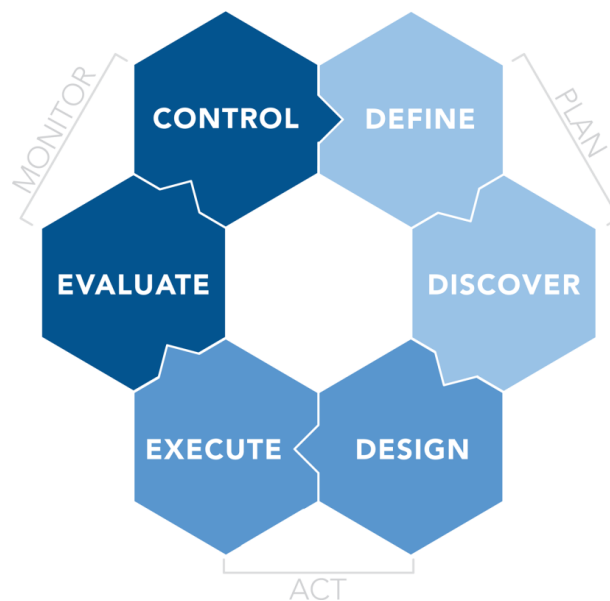


Figure 1. The DataFlux Data Management Methodology is a step-by-step process for planning, implementing and controlling data management projects.

Define

Imagine deciding to drive 500 miles to an area that is new and unfamiliar. The first instinct would be to grab a GPS or a map to help plan the trip. After plotting every turn and getting some guidance, only then would you set out on the drive, reasonably sure of how to get to the destination.

The Define phase of the data management methodology is just as important as mapping out a journey. The decisions made at this phase will guide the collection, organization, enhancement, monitoring and retirement of your data assets throughout the process. While you don't need all the answers at the beginning, you need a solid plan on how to proceed – and what the ultimate success indicators will be. Also, these success indicators have to map to the business problem identified earlier; the reason for the project (cut costs, mitigate risks, enhance revenue, etc.) provide a crucial guide for the Define phase.

During the definition phase, an organization should first answer a series of questions about:

- **People** – *Who's involved? And for what purpose?* This step outlines everyone involved in the data management process, including executive support, manager/director sponsorship and business and IT involvement. Organizations also set up steering committees and/or stewardship teams to facilitate collaboration on cross-functional issues.
- **Roadmap** – *Where are we now? Where do we want to go? What obstacles are in our way?* Often, the first task after selecting the right people is to determine a path to a successful outcome –including the definition of a successful outcome.
- **Source systems** – *What data will we need? Where is that data coming from?* The roadmap tells the story of where the project is intended to go. This part of the Define phase will inform the team on the source systems and which data will play a role in the data management project.
- **Business processes** – *Which business processes will be affected? How will better data enhance the way the organization operates?* This part of the Define process maps the data management strategy to existing business processes. Better data can ultimately streamline business processes, as less time is spent reconciling confusing views or managing poor-quality data.
- **Business rules and data definitions** – *How do we define "customer?" How do we want to optimize procurement and spend?* This phase seems simple enough, but it can be deceptively difficult. . Billing might define "customer" as anyone that receives an invoice, while customer support may only want to know who the end user is. These decisions will form the basis of business rules and data definitions that will guide later phases.

Discover

A quick inspection of your corporate data would probably find that it resides in many different databases, managed by many different systems, with many different formats and representations of the same data. Data can only be useful if you understand where it is, what it means to your organization and how it relates to other data in your organization. The Discover phase is designed to do just that.

Estimates for ERP and data warehouse implementation failures often run as high as 65 to 75%. In almost every instance, this is due to the same problem – a fundamental misunderstanding about the quality, meaning or completeness of the data. These are problems that should be identified and corrected prior to beginning the project.

Every new application implementation, data warehouse development, data migration or consolidation initiative should start with data discovery. Additionally, any time that new data sources enter your organization, start with data discovery. Data discovery has several components to it, and each prepares you for your data initiatives:

- **Data exploration** – This diagnostic phase is concerned with documenting the data in your organization and the characteristics of that data. Data discovery arms you with information about the accuracy, consistency and reliability of your data.
- **Data profiling and auditing** – Data profiling alerts you to data that does not match the characteristics defined in the metadata compiled during data exploration. But, more importantly, data profiling can also tell you if the data meets the business rules and definitions established in the Define phase. In addition, data profiling can help you determine the relationships across your data sources – where you have similar data, where data is in conflict, where data is duplicated and where data may be dormant.
- **Data cataloging and business vocabulary** – You need a development environment where data sources can be combined and rationalized. A place where you can group data sources into projects to allow you to work across your data sources and develop a consistent environment for managing your data. Data cataloging lays the groundwork for all data management tasks to follow. Data catalogs must be augmented with business definitions and vocabularies, allowing the business user to comfortably navigate the landscape.

Design

After completing the first two steps of the data management methodology, you will be able map your strategy, identify sources, understand the underlying formats and structures, as well as assess the relationships and uses of data. Now you face another challenge – taking all of these different structures, formats, data sources and data feeds, and creating an environment that accommodates the needs of your business.

The Design phase requires consolidation and coordination, all the while concentrating on three major imperatives:

1. **Consistency of rules** – Ultimately, an organization needs one set of business rules that can be stored centrally but deployed across all data sources, applications and lines of business.
2. **Consistency of the data model** – The data model is the single, definitive source for how your data maps to your business. Through the process of creating a well-structured data model, you identify the appropriate source systems and begin to reconcile multiple views, if required.
3. **Consistency of business processes** – During the Define and Discover phases, you will identify processes that are potentially impacted. Now, the task is to provide consistency across these processes. When creating business rules, you have to know how to reconcile questions like “Is this a new customer or an existing customer?” or “Is this a customer in good standing?” By understanding the processes impacted, you can design more effective rules to automate business processes.

Execute

Now that the business users have established how the data and rules should be defined, it is up to the IT staff to ensure that databases and applications adhere to the definitions. There are many types of architectures involved in this phase: enabling ERP and CRM applications via proprietary interfaces, enabling data marts and data warehouses via extraction, transformation and loading (ETL) flows, enabling MDM systems via service-oriented architecture (SOA)/ETL or other technologies. The method and management of enabling the data in any of these environments is a decision that IT has to make in order to ensure the integrity and integration into the various systems.

One potential pitfall in the Execute phase is to duplicate the rules and standards from the Design phase for each application or data source. When duplicating the rules and definitions across siloed, unrelated systems, multiple, point-to-point interfaces are inadvertently created. These rules definitions must then be updated, separately, each time a rule or business initiative changes.

Naturally, this approach is highly impractical for the IT team to manage. A better solution is to build the definitions once and ensure that you have the ability to collectively apply those definitions across your organization. As one IT director put it: “We want to build our standards and rules once and then have the ability to use them repeatedly and propagate to the entire organization seamlessly.”

For each data source, each business process and each application that is modified to the new data definitions, you need to:

- Understand the requirements
- Validate that the new integration meets the requirements
- Deploy the interface into production

By repeating this process during the execution phase, you can create the data management rules to guide the collection and organization of data, test its integrity, and move to the next phase of the process.

Evaluate

A healthy data lifecycle requires a robust monitoring and reporting system. The data needs to be consistently monitored so it remains fit-for-purpose for your organization. Why is this so critically important? After all, you just spent lots of time, energy and resources to get your systems to a point where the business users have a consistent and validated view of your organization. Isn't it time to just enjoy the success of all this effort?

Actually, the opposite is true. Very few organizations are static – they are forever growing and evolving. For example, you add new partners that bring new data to the table. Your business changes, sales regions are created or modified, you take on new initiatives and you develop new products. All of these changes must be reflected in your data, which makes the Evaluate phase so important.

Your mantra for success at this point needs to be: 1. Monitor; 2. Review; and 3. Optimize. Data should be monitored and validated as it enters your organization to verify it is meeting your rules. Those rules need to be constantly monitored to ensure they are still meeting the needs of your business. Efforts in discovery, design and execution will allow you to consolidate the rules and requirements into a single environment. With the ability to centralize the required data management rules, the changes can be immediately propagated across the organization, without duplication of effort.

Monitoring is a joint activity between IT and business users. IT monitors and validates that systems are running within their required service level needs. Business users also benefit from the monitoring reports – constantly reviewing the reports and validating that business needs are being met while making changes when the business needs change.

Control

One thing is certain in today's information age: a wide variety of data will continue to quickly pour into your organization. It is easy to see why data is a key asset. However, it is also important to recognize when data needs to be retired. The Control phase is about reassessing data. If data is no longer useful to your organization, you must be able to retire the data appropriately. This allows you to free up resources that are being expended maintaining the data environment.

For example, let's look at a common data problem facing financial services firms. When mergers, acquisitions and divestitures occur, you need the ability to purge or re-categorize data. You don't want to spend resources managing the data of a company that no longer exists.

Lastly, it is important to promote your successes across your organization. When you began your lifecycle, you were solving a business problem. By the time you have reached this phase in the lifecycle, you should have improved your business. Communicate and evangelize these messages to help everyone from senior management on down recognize that the efforts were successful and the business is improved. This demonstrates the business benefits of a sound data management methodology across the organization, and it paves the way for support of future initiatives.

Summary

The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second rule is that automation applied to an inefficient operation will magnify the inefficiency.

Bill Gates, Founder of Microsoft

For years, companies have poured time and resources into implementing complex and encompassing business applications to manage processes and make operations more efficient. While this tactic worked sporadically, Bill Gates' quote points out a very important fact. Better applications cannot fix a bad process. And, similarly, bad data can't support an effective business system.

The increased focus on data management within organizations around the world comes at a unique time. The amount and complexity of data is increasing, but there is a rising need to get more value from data. The good news: the data is already there. The bad news: the people, processes and technology required to manage this data is rarely in place.

The topic is an important corporate initiative for the next few years. After all, information on customers, products, assets, finances and employees may be the "last frontier" in the search for reduced costs, improved revenue and managed risk. Without a sound strategy for data management, organizations risk disaffected customers, a bloated supply chain and a heightened exposure to risk. Data can be the foundation for the next great success story – or the reason for the next round of business failures.



YOUR DATA.
YOUR BUSINESS.
ONE SOLUTION.

www.dataflux.com

Corporate Headquarters

DataFlux Corporation
940 NW Cary Parkway
Suite 201
Cary, NC 27513-2792
USA
877 846 3589 (USA & Canada)
919 447 3000 (Direct)

DataFlux United Kingdom

Enterprise House
1-2 Hatfields
London
SE1 9PG
+44 (0)20 3176 0025

DataFlux Germany

In der Neckarhelle 162
69118 Heidelberg
Germany
+49 (0) 6221 4150

DataFlux France

Immeuble Danica B
21, avenue Georges Pompidou
Lyon Cedex 03
69486 Lyon
France
+33 (0) 4 72 91 31 42