1.1 DRIVING THE NEED FOR MASTER DATA

Our world is increasingly becoming more connected in many different ways, and as the degree of separation shrinks, the amount of information we have at our fingertips seems to have exploded. Agile organizations that can adapt to the flood of largely redundant, yet sometimes conflicting pieces of data find that a combination of information sharing and operational collaboration is a differentiating factor for organizational success. In particular, as the speed and efficiency of the underlying infrastructure improves, so does the organization’s ability to control and take advantage of information assets through information sharing and collaboration. In turn, as our partners exchange and share more information, connectivity is enhanced as well—it is not just systems that work better together, but the people managing those systems forge better working relationships, leading to more effective management of the business and, ultimately, to competitive advantage.

However, the more we share information, the more we realize that years of distribution of computing power and business applications across the different lines of business has led to “islands of information coherence.” Historically, business applications were designed to meet operational business needs for specific areas of focus; resources have been aligned for vertical success, and to that end, the de facto application architecture evolves organically to support the operations of each line of business, with potential repercussions at the enterprise level. In general, application architectures designed to support operational processes within each business application area required their own information technology support and its accoutrements:
data definitions, data dictionaries, table structures, and application functionality, among others, all defined from the aspect of that business application.

As a result, what we would like to refer to as the “enterprise” is often composed of many applications referring to multiple, sometimes disparate sets of data that are intended to represent the same, or similar, business concepts. Alternatively, the same or similar names are used to refer to concepts that are completely distinct. Administrative control of data still largely resides with the line-of-business management, and as long as the data are used only to fuel operations, there are probably no issues. However, the recent trends in information use tend toward projects for consolidating and transforming data into information and then into actionable knowledge, and that requires collecting the instances of the data representing the critical business concepts from across business boundaries for the purpose of the common good. In other words, a transition toward applications intended to meet enterprise-level business needs must rely on a consistent view of data drawn from many data assets.

To exploit that information for both operational and analytical processes, however, an organization must be able to clearly define those business concepts, identify the different ways that data sets represent commonly understood business concepts, integrate those data into a consistent view, and then make the consistent view available across the organization. This need has introduced a significant opportunity for organizational information integration, management, and sharing, thereby creating a truly integrated enterprise. This is the essence of the master data management (MDM) challenge—envisioning how to organize an enterprise view of the organization’s key business information objects and to govern their quality, use, and synchronization to optimize the use of information to achieve the organization’s operational and strategic business objectives.

This chapter explores the history of enterprise master data, describes master data and master data management, and highlights the benefits of instituting a master data management program to support enterprise information management. The chapter also looks at the characteristics of master data and provides a high-level overview of what goes into an MDM program. We will also consider some of the organizational challenges of instituting an MDM program. Finally, the chapter provides an overview of the rest of the chapters of the book.
1.2 ORIGINS OF MASTER DATA

Originally, there was not a master data management challenge. In the early days of computing, the groups within an organization typically relied on a single computing resource that housed all the applications and all associated data sets. Before the revolution of structured data architecture inspired relational databases, flat data files ruled the world. Of course, the absence of a concept of data normalization led to significant data redundancy, sometimes addressed via code tables, yet the business impacts related to data redundancy were largely ignored. But because applications were developed to automate the straightforward processing that could be performed in batch, data duplication within one application or synchronization of data across different applications were essentially minor issues.

The introduction of workgroup computing in the 1980s coupled with the explosion of desktop computing systems with access to their own applications ushered in an era of information management distribution. At the same time, data architecture evolved to incorporate the relational view, leading to a reengineering of many existing applications to exploit the sleekness and performance of new Relational Database Management Systems (RDBMS). Administrative control of a business application along with its required resources gave business managers a degree of freedom and agility. However, by virtue of that distribution, the managers within one line of business could dictate the processes and constraints associated with the development of their own vertical applications to run their own lines of business. The distribution of administrative control, of course, led to variance in the ways that commonly used business concepts and objects were defined.

At the same time, the increase in both power and functionality at the desktop has engendered an even finer granularity of data distribution. This allows greater freedom in describing and modeling business information. Whether it is in the mainframe files, the database server, or in the desktop spreadsheets, we start to see a confusing jumble of concepts, along with creative ways of implementing those concepts. This is the first driver for master data management—the ability to rationalize the definitions and meanings of commonly used business concepts when possible and the ability to differentiate different business terms when they do not refer to the same concept.

Interestingly, since the mid-1990s, the pendulum has swung back to centralized computing. Organizational initiatives such as those
for deploying Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), or Supply Chain Management (SCM) drive the development of technical infrastructure such as data warehouses and the institution of applications that help improve the business. These initiatives are launched with the intention of consolidating the organization’s data into an information asset that can be mined for actionable knowledge to improve business productivity. Although the centralization of information for analysis and reporting has great promise, it has introduced a different challenge: as data sets were integrated and transformed for analysis and reporting, cleansing and corrections applied at the warehouse imply that the analysis and reports may no longer be synchronized with the data at their corresponding original sources. The differences in these reports alerted business users to qualifying the trustworthiness of the data, leading to suspicion in the correctness of the centralized system as opposed to the reporting systems on which the managers had always relied. This suggests more drivers for master data management—ensuring the quality of data that is used for analytics and maintaining consistency of the data across analytical systems as well as operational systems.

1.2.1 Example: Customer Data

Every company deals with customers, and within the company, each customer may be touched by many different business operations related to the business life cycle: marketing, sales, fulfillment, support, maintenance, customer satisfaction, billing, or service. In each of these contexts, the customer may be cast in different roles, with different aspects of the customer’s attributes being particularly relevant. Consequently, every business application may value some attributes over others in relation to the business context. The telemarketer may insist on accuracy of the telephone number to avoid calling the same prospect twice; the sales representative is concerned about duplication; the shipping department craves high-quality location information.

Within each business application, there may be a customer database, each with its own critical data attributes, and all of these records may be subjected to aggregation when consolidated into a data warehouse. But you clearly want to ensure that your business processes don’t fail because the customer appears multiple times in different data sets. In addition, you want to be confident that the customer’s activities are accurately portrayed in management reports.
This example demonstrates a critical aspect of the apparent distribution of information: data sets maintain representations of business concepts (such as “customers” or “products”) in different ways, even if they refer to the same real-world entities. And although different business applications record transactions or analysis regarding entities and their activities, for integration purposes it is desirable for all the business applications to agree on what those entities are, what the activities are, and whether the business terms used to describe these entities and activities truly refer to the same real-world ideas. We can summarize a number of objectives:

- When it can be determined that the entities referred to within the data sets refer to the same business concepts, integrate the multiple variations of those business entities into a unified view of uniquely identifiable data objects.
- Qualify the distinction between sets of data entities when it can be determined that they do not refer to the same business concepts.
- Provide a set of processes that enable enterprise applications to share that unified view of the business objects within the enterprise.

### 1.3 WHAT IS MASTER DATA?

In any organization, there are going to be commonly recognized concepts that are the focus of business processes, such as customers,
products, suppliers, vendors, employees, finances, and policies. We have established that as line-of-business silos rely on workgroup application frameworks, disparity has crept into organizational systems, introducing duplication and distribution of variant representations of the same “things.” There has been a growing desire for enterprise integration projects, such as those driven by expectations for a unified view of data as expressed by systems designed for customer relationship management (CRM), using a cleansed “product master file” and providing that all-elusive 360° view of the customer or other qualified enterprise reference repositories. In each of these instances, the underlying objective is to create a synchronized, consistent view of an organization’s core business entities.

So far we have used terms such as “critical business objects,” “business concepts,” or “business entities” when referring to the common data themes that exist across any business, and the underlying information objects represent the organization’s master data. But what characteristics define master data? Master data objects are those core business objects used in the different applications across the organization, along with their associated metadata, attributes, definitions, roles, connections, and taxonomies. Master data objects are those key “things” that matter the most—the things that are logged in our transaction systems, measured and reported on in our reporting systems, and analyzed in our analytical systems. Common examples of master data include the following:

- Customers
- Employees
- Vendors
- Suppliers
- Parts
- Products
- Locations
- Contact mechanisms
- Profiles
- Accounting items
- Contracts
- Policies

Within a master data object category there may be implicit or explicit hierarchical relationships. For example, there may be individual customer contacts within each client organization, or a product may be a shrink-wrapped collection of smaller product items. Classifications and reference tables are likely to be included as well.
Consider the following transaction: “David Loshin purchased seat 15B on flight 238 from Baltimore (BWI) to San Francisco (SFO) on July 20, 2008.” Some of the master data elements in this example and their types are shown in Table 1.1.

In fact, reviewing the master data objects presented in Table 1.1 poses some new questions—for example, we have a concept of a product ("seat 15B") that in its own right has a number of attributes associated with it:

- It is nonrefundable.
- Standby is permitted, but the customer must pay the price difference.
- The fare is subject to change until the ticket is purchased.
- Limited seating, advance purchase, and other conditions may apply.

Next, consider the difference between the seat and the flight. The flight is essentially a container for seats, and the flight also has attributes associated with it, such as a departure gate, an arrival gate, an on-time departure percentage, and an on-time arrival percentage.

Also, recognize that we are using the name “Baltimore” and “BWI” interchangeably, with “Baltimore” referring to the airport closest to and serving the location called Baltimore. In other instances, though, “Baltimore” refers to a different concept (such as a geopolitical region located in the state of Maryland) and may (or may not) include the more specific location referred to as “BWI” (which can be determined via a lookup table to indicate Baltimore Washington International Airport). Clearly, the more one drills down into the specific meanings or what we will call “semantics” of data values—data elements, records representing data

<table>
<thead>
<tr>
<th>Table 1.1 Master Data Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Master Data Object</strong></td>
</tr>
<tr>
<td>Customer</td>
</tr>
<tr>
<td>Product</td>
</tr>
<tr>
<td>Product container</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td>Location</td>
</tr>
</tbody>
</table>
entities, data sets that are collections of data entity representations, and so on—the more data about the data (or “metadata”) we discover.

Master data tend to exist in more than one business area within the organization, so the same customer may show up in the sales system as well as the billing system. Master data objects are used within transactions, but the underlying object sets tend to be static in comparison to transaction systems and do not change as frequently. Master data objects may be classified within a hierarchy. For example, we may have a master data category of “party,” which in turn comprises “individuals” or “organizations.” Those parties may also be classified based on their roles, such as “prospect,” “customer,” “supplier,” “vendor,” or “employee.” Although we may see one natural hierarchy across one dimension, the nested ways in which we describe, characterize, and classify business concepts (commonly referred to as “taxonomies”) may actually cross multiple hierarchies in different ways. For example, when we use the concept of a “party,” we may simultaneously be referring to an individual, a customer, and an employee. Alternatively, a “customer” may be either an individual party or an organization.

In turn, the same master data categories and their related taxonomies are used for analysis and reporting. For example, the headers in a monthly sales report may be derived from the master data categories (e.g., sales by customer by region by time period). Enabling the transactional systems to refer to the same data objects as the subsequent reporting systems ensures that the analysis reports are consistent with the transaction systems.

A master data system comprising a master data set is a (potentially virtual) registry or index of uniquely identified entities with their critical data attributes synchronized from the contributing original data sources and made available for enterprise use. With the proper governance and oversight, the data in the master data system (or repository, or registry) can be qualified as a unified and coherent data asset that all applications can rely on for consistent, high-quality information.

1.4 WHAT IS MASTER DATA MANAGEMENT?

Master data management is a collection of best data management practices that orchestrate key stakeholders, participants, and business clients in incorporating the business applications, information management methods, and data management tools to implement the policies, procedures, services, and infrastructure to support the capture, integration, and subsequent shared use of accurate, timely, consistent, and complete master data.
In other words, an MDM program is intended to support an organization’s business needs by providing access to consistent views of the uniquely identifiable master data entities across the operational application infrastructure. Master data management governs the methods, tools, information, and services for doing the items listed in the sidebar.

**USING MDM TO SUPPORT BUSINESS NEEDS**

- Assess the use of commonly used information objects, collections of valid data values, and explicit and implicit business rules in the range of applications across the enterprise
- Identify core information objects relevant to business success that are used in different application data sets that would benefit from centralization
- Instantiate a standardized model for integrating and managing key information objects
- Manage collected and discovered metadata as an accessible, browsable resource, and use the metadata to facilitate consolidation
- Collect data from candidate data sources, evaluate how different data instances refer to the same real-world entities, and create a unique, consolidated view of each one
- Provide methods for transparent access to the unified view of real-world data objects for both existing and newly developed business applications
- Institute the proper data stewardship and management policies and procedures at corporate and line-of-business levels to ensure a high-quality master data asset
1.5 **BENEFITS OF MASTER DATA MANAGEMENT**

Understanding the scale and scope of justifying, designing, planning, implementing, and maintaining an MDM program is itself a significant investment. Seeking executive-level championship for the tasks associated with deploying an MDM solution suggests a need for justifying the business value of master data management. However, recognize that regardless of how good the data management practices are that support MDM, master data integration and master data management are not *truly* the end objectives; rather they are the means by which other strategic and operational objectives are successfully accomplished, such as those promised by customer relationship management systems, product information management systems, enterprise resource planning systems, and the like. Although the principles that drive and support MDM reflect sound data management practices, it would be unusual (although not unheard of) that senior management would embrace and fund the MDM effort solely for the sake of establishing good practices. That being said, the business justification for MDM should be coupled with programs that will benefit from the availability of the unified view of master data. Providing a master unified view of key business information objects enables significant benefits in business productivity improvement, risk management, and cost reduction. More concretely, master data management can be justified in support of a business initiative that relies on any of the following MDM benefits:

**Comprehensive customer knowledge.** Organically developed application infrastructures often support the same type of customer data functionality in different ways. For example, a bank may have multiple customer banking interfaces: branch banking, automatic teller machines, bank by mail, Internet banking, telephone transactions, and even text-messaging. Customer records may be created, updated, or retired through any of these applications, but in an uncoordinated environment, there would be no way to know the number of distinct customers, how they like to bank, or how many different ways they have tried to execute transactions. A master data repository for customer data provides the single source for consolidating all customer activity, which can then be used to support both operational and analytical applications in a consistent manner.

**Improved customer service.** A unified view of master data objects enables improvements in meeting customer expectations with respect to availability, accuracy, and responsiveness to their orders, inquiries, as well
as providing transparency and governance in managing and protecting customer data.

**Consistent reporting.** Absent governance, duplication, and complex transformations applied differently at different places in the information processing flow introduce differences into the resulting generated reports. Reliance on end-user applications to digest intermittent system data extracts leads to questions regarding inconsistency from one report to another. Reliance on the reports generated from governed processes using master data reduces the inconsistencies experienced.

**Improved competitiveness.** The need to react to new business opportunities rapidly with limited resources challenges organizations to produce new business capabilities faster. MDM reduces the complexity of integrating new data and systems into the organization, thereby increasing agility and improving competitiveness.

**Improved risk management.** At a low degree of granularity, the more data points that applications must touch, the greater the potential for duplication, inconsistencies, or missing information to skew calculated risk assessment, either at the customer level or across lines of business. More trustworthy and consistent financial information improves the business’s ability to manage enterprise risk.

**Improved operational efficiency and reduced costs.** Replication of the same data often is linked to replication of activities associated with managing those data sets, ranging from typical data management routines (backups, maintenance), to licensing costs for infrastructure (such as RDBMS or Extract/Transform/Load (ETL) product license and maintenance costs), to specialized applicationware or services (such as data enhancement or geo/demographic appends). Formulating a unified view of the data enables the organization to reduce operating costs and tasks.

**Improved decision making.** Inconsistency across business intelligence activities often occur because of replication or duplication in the underlying data used to drive the decision-making process. Questions regarding the consistency of reports can stymie management decision making, leading to missed business opportunities. The information consistency provided by MDM across applications reduces data variability, which in turn minimizes organizational data mistrust and allows for clearer (and faster) business decisions.

**Better spend analysis and planning.** Master data associated with product, supplier, and vendor data improve the ability to aggregate purchasing activities, coordinate competitive sourcing, be more predictable about future spend, and generally improve vendor and supplier management.
Regulatory compliance. As one of the major enterprise areas of risk, compliance drives the need for quality data and data governance, and MDM addresses both of these needs. Information auditing is simplified across a consistent master view of enterprise data, enabling more effective information controls that facilitate compliance with regulations such as Sarbanes-Oxley, as well as other more industry-focused regulations such as these that apply in the United States: 21 CFR Part 11, the U.S. Patriot Act, Bank Secrecy Act, Basel II, antikickback statutes, and Graham-Leach-Bliley, among others.

Increased information quality. Collecting metadata made up of standardized models, value domains, and business rules enables organizations to more effectively monitor conformance to information quality expectations across vertical applications, which reduces information scrap and rework.

Quicker results. A standardized view of the information asset reduces the delays associated with extraction and transformation of data, speeding the implementation of application migrations, modernization projects, and data warehouse/data mart construction.

Improved business productivity. Master data help organizations understand how the same data objects are represented, manipulated, or exchanged across applications within the enterprise and how those objects relate to business process workflows. This understanding gives enterprise architects the opportunity to explore how effective the organization is in automating its business processes by exploiting the information asset.

Simplified application development. The consolidation activities of MDM are not limited to the data; when the multiple master data objects are consolidated into a master repository, there is a corresponding opportunity to consolidate the application functionality associated with the data life cycle. For example, there may be many applications that create new product entries into various product databases. Consolidating those product systems into a single resource allows us to provide a single functional service for the creation of a new product entry to which the different applications can subscribe. Introducing a technical service layer for data life cycle functionality provides the type of abstraction necessary for deploying a service-oriented architecture (SOA).

1.6 ALPHABET SOUP: WHAT ABOUT CRM/SCM/ERP/BI (AND OTHERS)?

Since the mid-1990s, data quality and data integration tools have matured in concert with a recognized desire on behalf of senior
managers to aggregate and consolidate replicated or duplicated copies of common or shared data objects that are peppered across disparate or distributed enterprise systems. In fact, MDM initiatives are largely aligned with the need to establish governance and quality across enterprise systems. So what differentiates today’s master data management initiatives from previous attempts at enterprise data consolidation? One might question whether there is any significant difference at all, which suggests that MDM may be destined for the same fate the earlier projects faced.

On the surface, MDM appears to be yet another attempt at consolidating data into a single “system of record” or “single version of the truth.” But consider customer data integration, product data integration, and enterprise dimension analysis. These ideas have been introduced within the context of data warehousing, business intelligence, sales force automation, customer relationship management, and so on. Yet, to some extent, the promise of many technical applications (such as CRM) has not been realized, and over time there has been growing skepticism as to their success. For example, many have been critical of the inability to effectively exploit a CRM system for its intended benefits. The reasons for this may not lie in the technologies per se but perhaps are influenced by four factors:

1. Despite the recognition of their expected business value, to some extent many of the aspects of these earlier projects were technology driven, and the technical challenges often eclipsed the original business need, creating an environment that was information technology centric. IT-driven projects had characteristics that suggest impending doom: large budgets, little oversight, long schedules, and few early business deliverables.

2. MDM’s focus is not necessarily to create yet another silo consisting of copies of enterprise data (which would then itself be subject to inconsistency) but rather to integrate methods for managed access to a consistent, unified view of enterprise data objects.

3. These systems are seen as independent applications that address a particular stand-alone solution, with limited ability to embed the technologies within a set of business processes guided by policies for data governance, data quality, and information sharing.

4. An analytical application’s results are only as good as the organization’s ability both to take action on discovered knowledge and to measure performance improvements attributable to those decisions. Most of these early projects did not properly prepare the organization along these lines.
These kinds of projects focus on integrating data from multiple sources into a single core repository, and each reflects some aspects of a master data management project. However, what differentiates MDM from previous enterprise integration efforts is that rather than having primarily a technology focus, MDM initiatives typically have a business focus, concentrating on the process of entity identification and validation with the business clients. Whereas MDM uses tools and technology, the combination of that technology with sound business and data management practices is what provides hope for a resounding success.

The intention of an MDM program is to create a single repository of high-quality master data that subsequently feeds applications across the organization with a synchronized, consistent view of enterprise data. The most critical aspects of a successful MDM solution require high-quality integration of master data instances from across the enterprise, and this relies heavily on the following:

- Inventory of data objects used throughout the enterprise
- Methods for identification of key data objects that are candidates for integration into a master data asset
- Resolution of the definitions, usage scenarios and intentions, and the meanings and semantics for these entities, as well as hierarchies and object relationships
- The ability to seamlessly facilitate standardized information extraction, sharing, and delivery
- A quality-directed migration process coupled with data survivorship rules for consolidating the “best records” for the master data asset
- An approach to transparently expose services to enterprise clients for accessing and managing the master data asset
- A governance framework for managing continued integration of enterprise data into the master data environment

Not surprisingly, the technical aspects of the tasks enumerated here depend on the traditional data quality, data integration, and data management tools and methods that most likely are already in place: database management; data profiling for discovery and analysis; parsing; standardization for data cleansing; duplicate analysis/householding and matching for identity resolution; data integration for information sharing; and data governance, stewardship, and standards oversight to ensure ongoing consistency.

In fact, many organizations that are implementing MDM programs have solutions that have evolved out of these traditional techniques.
It is common for master data integration programs to have evolved from customer data quality, product data quality, data assessment and validation, and data integration activities. Alternatively, although targeted solutions have been developed to support MDM functionality, they are often triggered by the introduction of data quality activities to support the technical infrastructure acquired for a specific purpose (e.g., enterprise resource planning or customer relationship management). Lastly, those successful evolutions occur when the ownership of the business benefits of the traditional approaches is transferred to the business side of the company, which emerges as data governance is introduced across the information sharing and collaboration context.

1.7 ORGANIZATIONAL CHALLENGES AND MASTER DATA MANAGEMENT

Numerous technologies have, in the past, been expected to address parts of this problem, providing customer master tables, industry-specific consolidated product management systems, and the like. But these applications have been criticized (perhaps unfairly) as a result of the organizational management approach to their implementation: largely IT-driven, presumed to be usable out of the box, lack of enterprise integration, and limited business acceptance. Resolving the issues pointed out by that criticism is what defines some of the considerations for implementing a successful master data management program:

- Effective technical infrastructure for collaboration
- Organizational preparedness
- “Round-trip” enterprise acceptance and integration
- Measurably high data quality
- Overseeing these (and other) processes via policies and procedures for data governance

That last point is one of the more important ones: what has emerged as a critical factor for success in growing a master data management program is establishing enterprise-wide data governance. It is important to understand how the migration of oversight and governance responsibilities from the lines of business to a centrally managed center of accountability will have impacts on the organization. Realize that years of distributing business applications into vertical lines of business has led to discrete islands of information, and the vertical alignment of IT and data management structures associated with those lines of business have erected barriers to what might be called
“horizontal collaboration.” This vertical alignment makes it difficult to resolve the differences in roles and responsibilities in relation to the isolated data sets as they are integrated into a master view. In addition, the politics of information ownership and management have created artificial fiefdoms overseen by individuals for whom centralization holds no incentive. Lastly, consolidating master data into a centrally managed data asset transfers the responsibility and accountability for information management from the lines of business to the organization. Therefore, some of the greatest challenges to success are not technical—they are organizational. It is largely because of these issues that MDM should be considered as a “program” and not as a project or an application. For example, these directives offer good suggestions for distinguishing a successful MDM program from one destined for failure.

**Organizational preparedness.** Anticipate that a rapid transition from a loosely coupled confederation of vertical silos to a more tightly coupled collaborative framework will ruffle a number of feathers. Assess the kinds of training sessions and individual incentives that must be established in order to create a smooth transition. These issues will be addressed in greater detail in Chapter 2.

**Data governance.** As management responsibility and accountability transition to the enterprise team, it is important to define the policies and procedures governing the oversight of master data. By distributing these policies and seeking proactive comments from across the different application teams, you have an opportunity to create the stewardship framework through consensus while preparing the organization for the transition. A comprehensive look at data governance is provided in Chapter 4.

**Metadata management.** All aspects of determining need, planning, migration strategy, and future state require a clarified view of the information about the data that is used within the organization—its metadata. The metadata associated with an enterprise master data set do more than just describe the size and types of each data element; rather these metadata include business definitions, valid value sets, accompanying business rules, and usage mappings, among other relevant data. A metadata registry provides a control mechanism or perhaps even a “clearing house” for unifying a master data view when possible, as well as helping to determine when that unification is not possible. The criticality of metadata to support MDM is discussed at great length in Chapter 6.
**Technology integration.** As is often the case, new technologies that are dependent on application packages are like the tail wagging the dog. Recognize the need to integrate technology to support the process instead of developing the process around the technology. Technology for data quality and data integration is covered in Chapters 5 and 10, with a focus on MDM solution architectures in Chapter 9.

**Anticipating change.** As with any paradigm shift, proper preparation and organization will subtly introduce change to the way that people think and act. The nature of a master data management program involves integrating data for competitive advantage, and this provides an opportunity for business managers to encourage individuals to begin to understand the differences between existing data replicas and a master data environment and to explore new ways to improve business productivity through the use of a master data asset.

### 1.8 MDM AND DATA QUALITY

Master data management success depends on high-quality data. Data quality can be improved through a master data management program. This codependence drives a large part of the activities that underlie the ability for the business to make use of a master data object framework:

**Assessment.** The ability to identify core data objects that should be incorporated into a master data repository depends on a structured process of assessment that relies on automated tools for analysis. This analysis comprises tools and techniques for empirically identifying candidate master data sets, primary keys, foreign keys, implicit relational structure, and embedded business rules before any data integration can begin.

**Integration.** The nature of variation associated with master data (person names, addresses, telephone numbers, product descriptions, etc.) demands that tools be used to help resolve the variation in representation of specific entities from disparate data sources. Standardization, parsing, and matching/linkage techniques have been available as part of any data cleansing tool kit, and the value of using those technologies to support new methods is abundantly clear.

**Assurance.** MDM is not going to be a case of “build it and they will come.” Organizational stakeholders will participate in the integration and consolidation process only as long as they are able to benefit from the process, implying the need for a high degree of confidence in the high
quality of master data moving forward. Auditing and monitoring compliance with defined data quality expectations coupled with effective issue response and tracking, along with strong data stewardship within a consensus-based governance model, will ensure ongoing compliance with application quality objectives.

Without a focus on managing the quality of master data, we run a risk of degenerating from an enterprise information management program to just another unsynchronized data silo. Data quality is critical to MDM success, and we will see how this theme penetrates almost all aspects of the MDM program.

1.9 TECHNOLOGY AND MASTER DATA MANAGEMENT

Although MDM should not be considered a technology project, it is clear that developing an MDM program cannot be done without leveraging tools and technology. The technical aspects of master data management rely on tools supporting data profiling, complex data analysis, metadata management, data modeling, data integration, parsing, standardization, record linkage and matching, cleansing, services-oriented architecture, access control, data federation, and data delivery, among others. However, this book is not about the tools but rather how those tools are employed to achieve the consolidated view of uniquely identifiable master data objects to support business strategy and objectives. Therefore, the descriptions of the tools and technology are presented in support of the MDM processes.

1.10 OVERVIEW OF THE BOOK

The objective of this book is to raise awareness among those tasked with developing MDM solutions of both the organizational and technical challenges and to help them develop a road map for success. To that end, the book concentrates on identifying the issues of critical importance, raises some of the questions that need to be asked and answered, and provides guidance on how to jumpstart that process within your organization. The book has 13 chapters.

- Chapter 1, Master Data and Master Data Management, introduces the historical issues that have created the need for master data management, describes both “master data” and “master data management,” begins
to explore what goes into an MDM program, and reviews the business value of instituting an MDM program, along with this overview of the rest of the book.

- Chapter 2, Coordination: Stakeholders, Requirements, and Planning, describes who the MDM stakeholders are, why they are relevant to the success of an MDM program, and what their expected participation should be over the course of the program’s development.

- Every organization exhibits different levels of maturity when it comes to sharing consolidated information, and Chapter 3, MDM Components and the Maturity Model, provides a capability model against which an organization’s maturity can be measured. By assessing the organization’s current state, considering the level of maturity necessary to achieve the organization’s objectives, and determining where the organization needs to be, one can assemble an implementation road map that enables action.

- Master data management is an enterprise initiative, and that means an enterprise data governance program must be in place to oversee it. Governance is a critical issue for deploying MDM. In Chapter 4, Data Governance for Master Data Management, we discuss how business policies are composed of information directives, and how data rules contribute to conformance to those information directives. We’ll look at what data governance is, introduce data stewardship roles and responsibilities, and propose a collaborative enterprise data governance framework for data sharing.

- No book on MDM would be complete without a discussion of the value of data quality, and Chapter 5, Data Quality and MDM, examines the historical evolution of MDM from data quality to its reliance on high-quality information. This chapter provides a high-level view of the data quality components and methods that are used for the purposes of master data integration.

- The key to information sharing through an MDM repository is a solid set of data standards for defining and managing enterprise data and a comprehensive business metadata management scheme for controlling the use of enterprise data. Chapter 6, Metadata Management for MDM, discusses data standards and metadata management and explores how master metadata is managed.

- As part of the process, it is necessary to identify the master data object types and determine the data assets that make up those object types across the enterprise. In Chapter 7, Identifying Master Metadata and Master Data, we look at the process of identifying and finding the data
sets that are candidates as sources for master data and how to qualify them in terms of usability.

- A core issue for MDM is creating the consolidation models to collect and aggregate master data. Chapter 8, Data Modeling for MDM, is where we will look at some of the issues associated with different source models and how to address data modeling issues for MDM.

- There are different architectural paradigms for master data management, and in Chapter 9, MDM Paradigms and Architectures, we look at existing application and information architectures and different architectural styles for MDM, how they all reflect a spectrum of implementations, and the pros and cons of each of those styles.

- Given a model and understanding the sources of master data, the next step is the actual process of data consolidation and integration, and Chapter 10, Data Consolidation and Integration, looks at collecting information from across the organization and formulating that into the integrated master data asset.

- The power of MDM increases greatly when the master data can be integrated back into the existing application environment. Chapter 11, Master Data Synchronization, discusses the needs and approaches for synchronizing data back to the existing applications.

- The value of MDM does not lie solely with the integration of data. The ability to consolidate application functionality (e.g., new customer creation) using a services layer that supplements multiple application approaches will provide additional value across the existing and future applications. The topic of a functional application services layer is covered in Chapter 12, Master Data Management and the Functional Services Layer.

- The book concludes with Chapter 13, Management Guidelines for MDM, a summary of the guidance provided throughout the preceding chapters to inform management decisions. To address the ongoing management issues, we offer some management guidelines for transitioning the project from developing the business case to maintaining a successful program.

1.11 SUMMARY

Master data management is more than just an application—it is a composition of people, tools, methods, and policies that will mold the future as organizations seek to exploit the value of the corporate information asset. The secrets to success lie in understanding how
MDM will transition your organization into one with a strong data governance framework, articulating the roles and responsibilities for data stewardship and accountability and creating a culture of proactive data quality assurance. Consider how transitioning to the different target architectures will impact the way you do business and prepare your organization for the rapid change. A successful master data management implementation will lead to a more effective integration of business and technology, as well as better organizational collaboration and productivity, and will ultimately increase your competitive advantage.