Supporting Your Data Management Strategy with a Phased Approach to Master Data Management
Table of Contents

Changing the Way We Think About Master Data ......................... 1
Master Data Consumers, the Information Vision and Working
  Backward ........................................................................ 2
Oversight: How Data Governance Supports the Data Strategy .... 3
Understanding: Establishing the Foundation with
  Shared Metadata ................................................................ 4
Control: Data Quality Management ........................................ 5
Internalization: Integrating Identity Management into the Business
  Process Model ................................................................... 6
Changing Behavior with Master Data ........................................ 7
The Phased Approach to MDM: Developing the Plan ............... 8

Content for this paper was provided by David Loshin, President of Knowledge Integrity Inc. Loshin is a recognized thought leader and expert consultant in the areas of data quality, master data management and business intelligence. He has written numerous books, white papers and Web seminars on a variety of data management best practices. His book, Master Data Management, has been endorsed by data management industry leaders, and his MDM insights can be reviewed at mdmbook.com. He can be reached at loshin@knowledge-integrity.com.
Changing the Way We Think About Master Data

As master data management (MDM) tools and techniques continue to mature, many organizations that embark on their MDM journey find that the typical approach focusing on data consolidation creates barriers to acceptance and integration. However, those same companies often attempt to reboot their master data program in the same way by promoting the benefits of a “golden record” providing a “single source of truth.” As a result, they front-load their efforts to extract, integrate and consolidate data. But as organizations recognize that the resulting consolidated sets of records have inherent inconsistencies and are difficult to incorporate into user applications, the concept of a single “master” repository appears to be somewhat flawed.

The presumption that one can create a single repository that includes all of the information about a specific entity (customer, product, agreement, etc.) is predicated on a faulty premise of consistent definitions and semantics for similar data. Yet different definitions of fundamental business data concepts – such as customer name, location, product description, etc. – imply differences in the ways data values are used. Unfortunately, the general approach to MDM seeks to merge records by eliminating variations without assessing the downstream impacts, potentially leading to process flaws.

If the goal of master data management is to integrate business value dependencies into a long-term information strategy, it is worthwhile to rethink both the intent – and potentially the value – of the concept of “master data.” Therefore, the focus must shift away from delivering master repositories. Organizations must transition the implementation from being technology-based and consolidation-focused to being value-based and consumption-focused. This more reasonable approach to the idea of master data is not creating a single source of truth but providing unobstructed access to a consistent representation of shared information.

In this paper, we examine how this approach can influence a phased organizational plan for fully embracing master data management. This plan incorporates five key fundamental techniques to enable master data management success, including:

- Business data consumer engagement.
- Data governance.
- Collaborative semantic metadata management.
- Data quality management.
- Identity resolution and management.

Rethinking MDM enables the development of a phased approach, with an incremental implementation plan. Taking such an approach delivers the benefits of sharing consistent, high-quality data while aligning the milestones and deliverables of a medium- and long-term information strategy. Transitioning from a consolidation approach to one that focuses on addressing business user information expectations and demands reframes the disciplines and capabilities of master data management. This approach can contribute to the long-term information strategy that uses best practices to enable comprehensive information visibility.
Master Data Consumers, the Information Vision and Working Backward

For practical purposes, master data management is a set of disciplines and tools enabling business data consumers to access a unified view of shared data about one or more specific master data domains such as customer or product. Yet, while the technical processes for extraction and consolidation drive the activity in most MDM programs, the actual intent of MDM is to satisfy business consumer needs for access and delivery of consistent, shared data.

Satisfying user needs means “working backward” by understanding how master data is employed within critical business functions and processes – and how that data is to be consumed by the users. This means identifying the key data consumers in the organization and soliciting their needs and expectations (both now and in the future). Information architects must work with the business teams to understand how the organizational mission, policies and strategic performance objectives are related to the use of master data. Finally, it’s necessary to understand how improvements in information sharing will maximize corporate value.

A gap analysis is performed to determine what must be done to the current data environment in order to satisfy future data consumption objectives. Performing this gap analysis helps in two ways. First, it isolates key strategic goals for data sharing that must be put into place before any technical MDM approach can add value. Second, and more importantly, it establishes the value of fundamental data management best practices that benefit the organization beyond the needs of a specific master data project. Effective data management practices penetrate all aspects of information use across the enterprise, such as:

- **Data governance**, which formulates the policies, processes and procedures to ensure that data use complies with explicit business policies; engages business data owners; identifies key data assets to be managed and shared; and delineates specific data requirements and quantifiable measures.
- **Metadata collaboration**, which defines standards for business glossary terms and definitions, representations for conceptual data elements, and alignment of models in ways that will reduce conflicts when data sets are merged into a unified master view.
- **Data quality management**, especially when deploying inspection and monitoring compliance with defined data standards and rules; and integrating services implementing data controls directly into the application infrastructure.
- **Integration of identity resolution and management** within business process model, which best satisfies the ongoing need for maintaining unified views for specific entities such as customer or product.
Use of these best practices does more than lay the foundation for improved information. These practices highlight the relationship among business processes, information and use. They also emphasize the need to adjust staff member behaviors as access to master data provides greater customer and product data visibility. And while the success of your strategic management plan for MDM must have milestones and deliverables aligned with these disciplines, the organization will directly benefit from each practice area independently.

**Oversight: How Data Governance Supports the Data Strategy**

Because enabling comprehensive visibility into a composed view of uniquely identifiable entities will continue to be part of the information strategy, there must be a reliable approach for:

- Ensuring that proper validation and verification is performed and approved as new data is created or acquired.
- Confirming that enterprise requirements for the quality of shared information are satisfied.
- Accessing and managing the composed view of shared information within defined security controls.
- Guaranteeing the consistency, coherence and synchrony of data views composed from a variety of sources.

Data governance provides the foundation for mapping operational needs to the framework of a sound data strategy designed around unified master data domains. There are numerous aspects of instituting a data governance program, including these key practices:

- **Data governance program management** – Developing an ongoing program management plan that identifies roles, defines responsibilities, provides templates for key artifacts (such as data policies, data quality requirements and policy deployment plans) and specifies the types of tools to support the data governance program.
- **Data governance operating model** – Specifying the organizational structure for operationalizing data governance, the interaction and engagement model to create a data governance council, the development of ongoing agendas, meeting schedules, and the establishment of measures to ensure that progress continues.
- **Definition and deployment of data policies** – Developing a framework for the process and workflows related to drafting, reviewing, approving and deploying data policies, as well as integrating business rule validation within the application infrastructure.
- **Data stewardship** – Defining the operational procedures for data controls, inspection, monitoring and issue remediation related to data policy violation.
While data governance is a prerequisite for rolling out access to a master data environment, there are some interim benefits as data policies are defined and deployed. The collaborative discussion that maps business policies into information policies helps highlight the interdependences of information use across business functions – paving the way for harmonized definitions that reduce complexity and inconsistency. As a result, there are fewer situations that require reconciliations, a bypass of organized data provisioning methods and the need for continuous reinterpretation of semantics.

Understanding: Establishing the Foundation with Shared Metadata

One of the most challenging aspects of MDM is that the consolidation process generally seeks to create one “perfect” record that contains the best values drawn from a set of source records. The determination of the golden record is often based on rules defined by the technical data practitioners, who are not always aware of the consequences of those rules.

There are many organizations where the same business term is used with different meanings, where different terms are used that share the same meaning or there are slight variations in definition. These slight variations are not relevant when the data sets are used for their original purpose. But, these differences can lead to discrepancies and inconsistencies when the data sets are combined.

For example, the business term “customer” can have widely different meanings depending on whether the source system is for sales, marketing, finance or customer support. Merging those data sets with the assumption that the term has a consistent meaning will lead to reports with questionable customer counts. If the root cause is the absence of standards that guide shared data use, the remedy is establishing a metadata management framework supporting best practices for shared definitions and semantics, including defining and observing policies and processes for:

- Managing the metadata management life cycle. This begins with identifying key business terms and phrases, and moves to agreed-upon definitions by collaborating with the subject-matter experts to refine an acceptable common definition. Alternately, this phrase recognizes when overloaded business terms have different meanings and clarifies them with differentiating names and definitions.

- Data standards. While de facto standards may already be used, any variation creates additional work for any data consolidation or aggregation effort, affecting operational master data management, business intelligence and analytical applications. Defining and enforcing data standards reduces variance, simplifying data integration while retaining consistency in semantics.
• **Collaborative agreements.** Introducing data governance and data stewardship opens the door for agreeing to existing or newly defined data standards, business glossary terms and definitions, data element concepts, and corresponding data types and sizes. A governance program facilitates these agreements, and collaboration tools can supplement the tasks associated with the metadata management life cycle.

• **Data lineage.** Data lineage needs to be mapped from creation (or acquisition points) to the various modification or usage touch points across the application landscape. An accurate lineage map helps in understanding the business application and process impacts of modifications to definitions or underlying data structures. It also enables the insertion of data validity controls for inspecting and monitoring data quality and usability.

A reasonable investment in metadata management can add value to the organization by facilitating communication about shared concepts across business functions, while reducing variance and complexity. Metadata management also adds value by smoothing the differences between data sourced from functional silos – and paves the way for an effective MDM effort.

### Control: Data Quality Management

Data governance encompasses the program management required to manage data consumer expectations and requirements, along with collaborative semantic metadata management. However, the operational nexus is the integration of data quality rules into the business process and application development life cycle. Directly embedding data quality controls into the data production workflows reduces the continual chore of downstream parsing, standardization and cleansing. These controls also alert data stewards to potential issues long before they lead to irreversible business impacts.

Engaging business data consumers and soliciting their requirements allows data practitioners to translate requirements into specific data quality rules. Data controls can be configured with rules and fully incorporated into business applications. Data governance procedures guide data stewards through the workflow tasks for addressing emerging data quality issues. Eliminating the root causes for introducing flawed data not only supports the master data management initiative, it also improves the overall quality of enterprise data. Data quality management incorporates tools and techniques for:

• **Data quality rules and standards.** Providing templates for capturing, managing and deploying data quality rules – and the standards to which the data sets and applications must conform – establishes quantifiable measures for reporting quality levels. Since the rules are derived from data consumer expectations, the measures provide relevant feedback as to data usability.

• **Data quality controls.** Directly integrating data quality controls as part of the application development process means that data quality is “baked in” to the application infrastructure. Enabling rule-based data validation ratchets data quality out of downstream reactive mode and helps data practitioners address issues within the context of the business application.
• **Monitoring, measurement and reporting.** A direct benefit of data quality rules, standards and controls is the ability to continuously inspect and monitor data sets and data streams for any recognizable issues, and to alert the right set of people when a flaw is detected.

• **Data quality incident management and remediation.** One of the most effective techniques for improving data quality is instituting a framework for reporting, logging and tracking the status of data quality issues within the organization. Providing a centrally managed repository with integrated workflow processes and escalation means that issues are not ignored. Instead, issues are evaluated, investigated and resolved either by addressing the cause or determining other changes to obviate the issue. The visibility into the point of failure (or introduction of a data error) coupled with the details of the data quality rules that were violated help the data steward research the root cause and develop a strategy for remediation.

While one of the proposed benefits of MDM is improved data quality, in reality it’s the other way around: To ensure a quality MDM deployment, establish best practices for proactive data quality assurance.

**Internalization: Integrating Identity Management into the Business Process Model**

The previous phases – oversight, understanding and control – lay the groundwork of a necessary capability for MDM: entity identification and identity resolution. The increased inclusion of data sets from a variety of internal and external sources implies the increased variation of representations of master data entities such as customer, product, vendor or employee. As a result, organizations need high-quality, precise and accurate methods for parsing entity data and linking similar entity instances together.

Similarity scoring, algorithms for identity resolution and record linkage are mature techniques that have been refined over the years and are necessary for any MDM implementation. But the matching and linking techniques for identity resolution are just one part of the solution. When unique identification becomes part and parcel of the business process, team members become aware of how their commitment to maintaining high-quality master data adds value across the organization. Identity resolution methods need to be fully incorporated into the business processes that touch master entity data, implying the need for:

• **Enumerating the master data domains.** It may seem obvious that customer and product are master data domains, but each organization – even within the same industry – may have numerous data domains that could be presumed to be “mastered.” Entity concepts that are used and shared by numerous organizations are candidate master domains. Use the data governance framework to work with representatives from across the corporation to agree on the master data domains.
• **Documenting business process models and workflows.** Every business process must touch at least one master data entity. For an MDM program, it’s critical to understand the flow of business processes – and how those processes are mapped to specific applications. The organization must also know how to determine which applications touch master data entities.

• **“CRUD” (create, read, update, delete) characteristics and process touch points.** Effective use of master data cuts horizontally across different business functions. Understanding how business processes create, read or update master data entity instances helps the data practitioner delineate expectations for key criteria for managing master data (such as consistency, currency and synchronization).

• **Data access services.** Facilitating the delivery of unobstructed access to a consistent representation of shared information means standardizing the methods for access. Standard access methods are especially important when master data repositories are used as transaction hubs requiring the corresponding synchronization and transaction semantics. This suggests the need to develop a layer of master data services that can be coupled with existing strategies for enterprise data buses or data federation and virtualization fabrics.

• **“Master entity-aware” system development.** If one of the root causes for the inadvertent replication of master data stems from siloed application development, the remedy is to ensure that developers use master data services as part of the system development life cycle. Couple the delivery of master data services with the proper training and oversight of application design and development.

The methods used for unique identification are necessary but not sufficient for MDM success. Having identified the business applications that touch master data entities is a prelude to exploring how the related business processes can be improved through greater visibility into the master data domains.

### Changing Behavior with Master Data

As organizations seek to unite their information assets using MDM, it’s become apparent that getting the most from master data is dependent upon properly trained users. Without training, the value of even the most accurate and current information about customers and products will be limited. This is illustrated using three different scenarios.

• **As a first scenario,** many MDM projects have been justified in terms of improving revenue generation through increased cross-selling and up-selling. Yet many companies have sales teams organized around specific products. Having visibility into all of a customer’s purchases yields limited benefits if salespeople are not retrained (and compensated) to sell items from other product lines and recognize the opportunities for cross-selling and up-selling.
• In another scenario, companies may create master inventory data sets for the purposes of reducing costs via spending analysis and vendor management. However, to properly analyze aggregate spending by product category, the people logging equipment and supply purchases must perform data entry within the confines of strict data standards for naming and product categorization. For example, paper clips might be referred to as “clips,” “paper clips,” “paperclips,” “trombones,” etc. But all would be categorized as “paper fasteners” for the purposes of analyzing costs. When items that are not tagged within the proper product hierarchies are entered, the spending analysis application will not properly roll up the aggregate expenses by product category.

• Finally, regulatory compliance may be affected when users are not aware of master data availability. For example, pharmaceutical companies are obliged to report the total amount of various payments to physicians and other health care providers. However, large companies with many subsidiaries may not share the same metadata for categorization of payment type, which will prevent the proper aggregations and potentially result in penalties due to noncompliance.

A phased MDM approach not only plans for the technology integration, it also lays out a strategy for the appropriate change management to adjust business processes, workflows and behaviors. The phased approach engages users to document how they interact with the existing business processes and coordinate progress in relation to shared master data domains. Then, enterprise architects can use the business process models and data lineage to:

• Enumerate the business processes that touch candidate master data domains.

• Consider how processes within individual business functions rely on specific master data attributes.

• Examine how the business user interacts with master data domains as part of the workflows.

• Identify opportunities for streamlining the business process and improving value creation as a result of a unified view of shared master data.

These opportunities can be prioritized based on their potential lift. Process owners can work with their teams to develop educational materials, schedule training sessions and employ ongoing mentoring to incrementally change behaviors in a beneficial way.

The Phased Approach to MDM: Developing the Plan

This paper has suggested that establishing fundamental practices and capabilities as a prelude to launching a master data management program will increase the probability of MDM success.

The paper also suggests that these capabilities provide more general benefits beyond paving the way for MDM. The next step involves some internal reflection and assessment about levels of maturity as a planning exercise. Ask these types of questions:

The Phased Approach to MDM: Developing the Plan

This paper has suggested that establishing fundamental practices and capabilities as a prelude to launching a master data management program will increase the probability of MDM success.

The paper also suggests that these capabilities provide more general benefits beyond paving the way for MDM. The next step involves some internal reflection and assessment about levels of maturity as a planning exercise. Ask these types of questions:
• Which of these capabilities already exist within the organization?
• For each capability, what is the level of maturity?
• What are the expectations for these capabilities to support master data management?
• What are the gaps?
• What are the corporate value drivers that would be positively affected by each capability?
• What is the relative priority?
• What needs to be done to plan, socialize and execute their development?

Use the template in Table 1 as guidance for prioritizing tasks, and capturing and reviewing the answers to these questions. The highest priority tasks are those that support the long-term MDM program and provide near-time value. From that, it’s possible to assemble a more immediate set of project plans to launch the design, development and deployment of the selected capability.

And this is just the start. Don’t limit yourself to what is described in this paper. As you go through the process, you will uncover more data management best practices that can be deployed across the enterprise. As long as those practices are aligned with the organizational objectives for the master data management program, they can be dovetailed into the ongoing program plan.

<table>
<thead>
<tr>
<th>Foundation Area</th>
<th>Capability</th>
<th>Exists?</th>
<th>Current Level of Maturity</th>
<th>Required Level of Maturity</th>
<th>Value Drivers</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Governance</td>
<td>Program management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>Operating model, roles and responsibilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>Templates for data policies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>Process for defining and approving data policies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>Data policy deployment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>Requirements assessment for data governance tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>Acquire data governance tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td>Deploy data governance tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foundation Area</td>
<td>Capability</td>
<td>Exists?</td>
<td>Current Level of Maturity</td>
<td>Required Level of Maturity</td>
<td>Value Drivers</td>
<td>Priority</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>--------</td>
<td>---------------------------</td>
<td>----------------------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>Metadata</td>
<td>Process for metadata definitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Business term glossary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Data standards process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Collaborative agreement process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Manage data lineage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Metadata tools requirements assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Acquire metadata tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Deploy metadata tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Define data quality processes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Process for data controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Data quality incident tracking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Data quality assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Evaluation and remediation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Requirements for inspection and monitoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Data quality reporting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Data quality tools requirements analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Quality</td>
<td>Deploy data quality tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity Management</td>
<td>Identify master data domains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foundation Area</td>
<td>Capability</td>
<td>Exists?</td>
<td>Current Level of Maturity</td>
<td>Required Level of Maturity</td>
<td>Value Drivers</td>
<td>Priority</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------</td>
<td>---------</td>
<td>---------------------------</td>
<td>---------------------------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>Identity Management</td>
<td>Business process models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity Management</td>
<td>Business process workflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity Management</td>
<td>Analysis of master data touch points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity Management</td>
<td>Design data access services</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity Management</td>
<td>Implement data access services</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity Management</td>
<td>Modify system development life cycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change Management</td>
<td>Business process assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change Management</td>
<td>Identify opportunities for improvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change Management</td>
<td>Improvement project materials and training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Data management practices: self-assessment and prioritization.
About SAS

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 65,000 sites improve performance and deliver value by making better decisions faster. Since 1976 SAS has been giving customers around the world THE POWER TO KNOW®.