Challenges in the Effective Use of Master Data Management Techniques
Table of Contents

Introduction ................................................................. 1
Consolidation: The Typical Approach to Master Data Management . 2
Why Consolidation Fails to Meet Data Consumption Needs .......... 3
Master Data Management Considerations ........................... 5
Rethinking the State of MDM ........................................... 6

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 mdmbbook.com. He can be reached at loshin@knowledge-integrity.com.
Introduction

Master data management is a maturing technology with expectations for continued growth in products and services. Early in 2012, Gartner Inc. predicted that "worldwide master data management (MDM) software revenue will reach $1.9 billion in 2012, a 21 percent increase from 2011," with a market "forecast to reach $3.2 billion by 2015."\(^1\) Apparently the industry is still targeted for a comprehensive boom.

However, the evidence seems to indicate that the traditional approaches taken from implementing a master data management program do not succeed as planned. Even with the expected growth in the software side of the industry, Gartner also suggests that "when IT managers start an MDM program, they often struggle to get business stakeholders on board," and "fail to demonstrate the business value of MDM."\(^2\) And analyst Andy Hayler, of The Information Difference, also notes that among MDM project owners, only "24 percent rated their own projects as ‘successful’ or better."\(^3\) Our experience suggests that a lack of clear goals and a developed information strategy are fundamental barriers to MDM success.

Early MDM efforts cited lack of a business sponsor as a primary culprit, along with the absence of data governance or data quality. Yet these pronouncements are indicative of more fundamental challenges as the root causes of disappointment for MDM success, especially when the intent is to create a “single source of truth.” This is largely due to a combination of factors, including:

- The overwhelming suggestion that MDM can deliver a golden record for the enterprise.
- The pervasive marketing messaging for MDM that overpromises the benefits of the master data repository in a vacuum, without focusing on an overall data management or data governance strategy.
- The focus on technology assessment, acquisition and deployment over the satisfaction of business needs.
- Simultaneous, yet uncoordinated MDM activities initiated independently by siloed departments and divisions.
- Issues in proper project/program scoping and resourcing for management, maintenance and business process migration.
- Limited success in developing processes for soliciting business data consumer requirements.

Delivering on the promise of MDM requires some further consideration to differentiate between two separate efforts: the development and population of a master data repository versus establishing comprehensive data visibility that adds value to a variety of business processes. The former is the “how” behind MDM, while the latter is the “why.”

---

MDM is an enterprisewide activity that must go beyond the needs of any single business function, so it is important to finesse any recognized barriers to success. In this paper, we look at how the focus on data consolidation (targeting the elusive single source of truth) has created a scenario in which MDM programs can lead to increased data management complexity, more inconsistent data and an inability to meet business needs. We then take a step back to see what the impacts are to an MDM activity, and then summarize by suggesting that the industry rethink data management strategies that expect to take advantage of shared repurposed enterprise information.

Consolidation: The Typical Approach to Master Data Management

To understand the root causes of some of these failure patterns for master data management, we must first consider the typical methodology for motivation, design and implementation of an MDM project. Generally, the catalyst for considering MDM as a solution is the confusion stemming from the proliferation of data sets carrying similar information about the same core entities such as “customer” or “product.” These multiple views are used in different business functions, processes and applications across the enterprise.

There is a presumption that because data models, data attributes and attribute values vary across these different data subsystems, no single business application can possibly maintain a record with the most accurate or current data representing each entity. Given that premise, the logic is that merging data from this collection of data sets will surely provide a consolidated record with the most accurate values, the proverbial “single source of truth.”

In turn, the MDM activity focuses on the consolidation tasks:

- Identifying data sources containing information about the entity selected to be mastered.
- Creating a target master data model for the selected entity.
- Developing extraction scripts to pull data from the sources.
- Cleansing extracted data.
- Developing business rules for “survivorship,” or arbitrating among the source values by data attribute to select the one that will be loaded into the target master model.
- Selecting the identifying attributes used to uniquely identify a specific entity.
- Developing merging rules for linking records by identifying attributes.
- Merging data from different sources into the target model.
- Creating a master index through which entity identification and resolution can be performed on records that are candidates for linkage to the master repository.
- Loading the consolidated record into a master repository.
- Developing the synchronization processes to repeat the consolidation on a periodic basis.
The objective of the task is the creation of the consolidated target data set that contains what is expected to be “master” data from a variety of sources. Rules of data element-value survivorship are applied as part of the consolidation and merging process, resulting in the selection of only one of the source values to populate the target master record. In essence, though, what ends up happening is that data is dumped into a repository, which is then used to merge similar records together to deliver a single representation or record for each entity.

The completion of the repository and its installation as the “golden record” are a significant milestone, and often signify the end of the project. There are a number of tasks related to configuration and maintenance, but often the creation of the master data silo is seen as the result of the process.

Why Consolidation Fails to Meet Data Consumption Needs

Needless to say, the creation of the master repository cannot be seen as the end of the MDM process, because the master repository is purposeless without any applications to use it within defined business processes. Regardless, one could say that the consolidation approach to master data management has a number of weak points that lead to a more general failure to create any business advantage.

In retrospect, the surprising number of unsuccessful or underperforming MDM projects is a byproduct of numerous potential failure points that contribute to the inability to meet enterprise needs or to satisfy downstream consumer expectations. Some examples include:

- **YADS or “Yet Another Data Silo.”** Without proper planning for functional integration and use of the data in a master repository, the IT department essentially has created another data silo that requires oversight, synchronization and maintenance. But without planning for actual business users, the expense for ongoing maintenance will be seen as a waste of resources, leading business sponsors to stop funding the project.

- **Absence of user input.** The consolidation approach is largely a technical one, and typically does not take (individual or business process) consumer data requirements into account. The result is that potential users are not properly served by the consolidated repository and will shy away from expending their own resources to adapt their applications.

- **No planning for migration.** Many of the resources allocated for master data management are used to build the repository, often leaving little or none for any business process and application transitions or functional migrations. After the fact, there is a realization that additional resources are needed to review existing applications; identify touch points for the master data entities; and redesign, develop and deploy changes to existing applications. But at that point, it may be too late to go back for the additional funding necessary to activate the master repository.
• **Loss of data.** The algorithms used for linkage, merging and survivorship to consolidate multiple records into a single “golden record” are, by their very nature, lossy. Survivorship implies that some data values are used while others are discarded, usually based on arbitrary rules defined by IT staff members. In some usage scenarios, this data loss negatively affects the downstream business processes.

• **Loss of meaning.** As suggested in the previous item, the objective criteria for merging and consolidation are defined by the IT staff members, and the defined rules may ignore semantics implied by the originating business process context. The perennial question “What is the definition of customer?” is indicative of the variance in meanings associated with commonly used business terms. But when multiple records representing a customer are merged into a single record, the change of context eliminates any meaning implied by the original contexts. This introduces conflicts with concept semantics from the original sources.

• **Misalignment with enterprise data initiatives.** Master data management is an enterprise initiative, as are other data management activities such as data quality management and data governance. All of these organizational data management initiatives must be coordinated so they can make use of their synergy. Often, however, MDM activities are isolated within individual divisions or groups, leading to a failure to capitalize on existing data management tasks and data governance programs.

• **Absence of process governance.** The absence of governance integrated into existing business processes allows inconsistencies to be introduced into application data subsystems that bypass the master repository. For example, a business process that requires looking up customer records would need to be adjusted to ensure that data-entry personnel properly search for matching records, even in the presence of errors, prior to creating new (yet inadvertently) duplicative customer records.

• **Extra enterprise data.** Very often an organization uses entity data that originates outside of the enterprise, is managed by third parties or is managed within proprietary environments. This trend may grow as more organizations use cloud-based applications. It is difficult, if not impossible, to impose a constraint to only use entity data taken from the master repository when the consuming applications are outside of corporate administrative control.

• **Shadow IT.** Desktop productivity tools such as spreadsheets and personal databases are often used in ways that ignore defined data policies as well as the master repository. As these personal data artifacts eventually make their way into production information flows, information that should be characterized as potential master data bypasses the master data repository and its associated services.
Figure 1: Questions arise when consolidating data into a master repository without considering how the master data is used.

The upshot is that although a consolidation-oriented project plan for MDM will result in a delivered master data repository, much of that investment and effort will be wasted when trying to integrate that repository in a meaningful way. The effort necessary to retool existing production systems only results in the same business applications providing the same functionality, with only the auxiliary benefit of improved data quality. Integrating master data into renovated or newly acquired applications does suggest potential benefits for enterprise coordination, but the value is not delivered until after the new applications have been deployed.

Funding decisions often map to the perceived value of a program, and when it seems that there will be a delay in value delivery, activities are often dampened or postponed. This is why many organizations attempt MDM multiple times before eventually developing a shared data asset that enables value creation.

Master Data Management Considerations

The fundamental flaw in the consolidation approach to MDM is the concept of a “master” repository. The presumption is that an organization consisting of disparate business functions, managed within siloed budgets and governed by divisions and groups, can combine its data sets to create a single repository for all the information about any specific entity (customer, product, agreement, etc.). This presumption is predicated on faulty assumptions regarding organizational preparedness and alignment, coupled with inflated expectations about technical silver bullets, such as:
1. **Organizational semantic consistency.** The creation of a consolidated single record for any entity concept requires consistent definitions and semantics for all instances of similar data from numerous sources. But the structure and representation varies, and independent data models and usage scenarios are engineered with a bias toward business function needs. For example, there are certain to be differences in customer names, locations and contact information, depending on the application. Variant definitions of a customer name, location, product description, etc. imply differences in the ways those data values are used, and merging and eliminating those variations without assessing the impact of doing so will lead to process flaws downstream.

2. **Organized cross-functional collaboration.** The corporate application environment evolves organically depending on the immediate needs of each business function. A retrospective assessment of application architecture that reviews the environment and attempts to overlay a conceptual architecture is, of course, an illusion. That means that unless an organization already has succeeded in executing enterprise programs using cross-functional collaboration, it probably does not have the level of maturity and preparedness to deploy an enterprise data asset management program such as MDM.

3. **Reliance on IT to solve the problem.** The IT department is often relied on to make decisions regarding the governance and oversight of data even though the responsibility should be shared among the business process owners who use that data. That is coupled with the technologists’ promotion of MDM as a solution for collaboration regarding data policies, data standards and agreement for definitions of commonly used business terms and other metadata artifacts. The result is a rush to build a repository for data consolidation in the absence of a blueprint and road map for what data consumers really will need now and in the future.

In other words, MDM is seen as the technical silver bullet that will solve the problems introduced by variance in definition and perceived semantics. However, the realization that the consolidated “single source” is insufficient to meet business needs only arrives after it has been built.

**Rethinking the State of MDM**

The conclusion is that the consolidation approach to master data management creates barriers to acceptance and integration. Yet organizations often attempt to restart MDM programs in exactly the same way: arguing for the benefits of the “golden record,” selecting a single domain to master, and yet again front-loading the effort in extraction and consolidation. In each of these cases, because the process has not changed, the results will not change either, leading to repeated stalls and minimal added value.
Instead of rebooting the MDM activity in the same old ways, it would be worthwhile to rethink both the intent and potentially recognized value of the master data concept. The alternative to driving the activity as a consolidation effort is to not expect the creation of a single master repository. That means changing the implementation approach from being technology-based and consolidation-focused to being value-based and consumption-focused. This creates a more reasonable approach where the idea of master data is not providing a single source of truth but providing access to a consistent representation of shared information.

The value of master data is then based on enabling each business application to access all of the information about the involved entities (such as “all products purchased by all members of a customer’s household”) without the issue of interim consolidation decisions that eliminate or change any of that data visibility. In turn, that relies on the assurance of accessibility, consistency, currency and quality of data coming from the original sources, allowing business process users to interpret and use data sets in ways that are consistent with their original creation.

When the MDM program is based on satisfying the data availability requirements of downstream consumers, your organization can map out how those requirements are solicited, captured, documented and addressed within a longer-term road map and plan that can deliver incremental value at predictable milestones.

This phased approach begins with the identification of the key data consumers in the organization, and incorporates foundational facets of developing the road map:

- **Enterprise consumer engagement.** Prior to designing and building a repository, you must understand the user expectations and requirements for that shared information. The first phase involves engaging these consumers so their specific needs, requirements and expectations (for both current business processes as well as future needs) can be solicited and incorporated into the strategic enterprise data management plan.

- **Data governance.** Putting the right amount of policy definition and management in place is critical to ensure consistency of use for shared information and data visibility. The governance practice encompasses the operating model for oversight (including any hierarchical organization for a data governance council or committee) as well as the operational aspects that direct how data stewards address and resolve data issues resulting from the broad-based data visibility implied by master data management.

- **Metadata collaboration.** Prior to exposing a shared view of data to the community of data consumers, one must provide a common view of the business terms, definitions and uses that are pervasive across the application landscape. This means selectively collecting and documenting corporate metadata for critical shared-data concepts. It also demands harmonization of the definitions and contextual semantics to ensure that data users are comfortable relying on the shared data. Harmonization also ensures that relevant information is not eliminated as part of the consolidation process.
• **Data quality management.** Solicitation of enterprisewide data requirements must be balanced with an established enterprise framework for data quality assurance and management. This must include incident management, data quality standards, data controls, measurement and monitoring, cleansing when necessary, and defined processes for investigation and remediation of data issues and flaws.

• **Integrated identity resolution.** As more business-process consumers recognize the need for comprehensive data visibility, their applications must be retooled to search for and use shared information about core master entities. Identity resolution services are typically seen as part of an MDM tool suite, and all consuming business-process models must be augmented with integrated use of identity resolution at each entity data touch point.

Interestingly, in addition to supporting the master data strategy, each of these facets adds value to the organization independently, either in terms of synchronizing data usability, reducing confusion about data semantics or improving enterprise data quality. That suggests that the best way to rethink MDM is to develop an incremental implementation plan that delivers the benefits of consistent shared data while setting the stage for future efforts. These efforts could include indexed access into a federated view of entity data that originates and lives in different enterprise (or potentially even extra enterprise) data resources. This alternative approach uses the existing information landscape, but does not introduce the inconsistencies or engineering challenges of connecting existing or future applications to an amorphous consolidated data dump.

Master data management is not just something you install and fill with data. By transitioning from a consolidation approach to a data utilization approach, you will see how the disciplines and capabilities of master data management contribute to a long-term information strategy that uses best practices to enable comprehensive information visibility.
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®.