PREPARING YOUR ORGANIZATION FOR MASTER DATA MANAGEMENT

By David Loshin

Sponsored by

tdwi.org
Preventing Your Organization for Master Data Management

By David Loshin

2 FOREWORD
Eliminate the barriers to success for master data management.

2 NUMBER ONE
Identify consumers of master data and document their expectations.

3 NUMBER TWO
Understand how repurposing data enables ungoverned proliferation of data.

3 NUMBER THREE
Develop master data services for application migration.

4 NUMBER FOUR
Plan processes for synchronization with proprietary systems.

4 NUMBER FIVE
Align success criteria for master data management across the organizational chart.

5 NUMBER SIX
Harmonize structural inconsistencies and standardize data elements.

5 NUMBER SEVEN
Unify master concept semantics when designing master models.

6 NUMBER EIGHT
Map business needs to technology acquisition.

6 SUMMARY AND CONCLUSIONS

6 REFERENCE

7 ABOUT OUR SPONSOR

7 ABOUT THE TDWI CHECKLIST REPORT SERIES

7 ABOUT THE AUTHOR

7 ABOUT TDWI RESEARCH

© 2011 by TDWI (The Data Warehousing Institute®), a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or in part are prohibited except by written permission. E-mail requests or feedback to info@tdwi.org. Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.
There is evidence that the master data management (MDM) movement is maturing, as organizations consider the incorporation of data quality and data governance processes to supplement the emergence of master data systems. Yet according to research firm The Information Difference, “MDM programs appear to have a mixed track record, with only 24 percent [of respondents] reporting their implementation as successful or better” (The Information Difference, 2011). Clearly, there is room for improvement, and for those intrigued by the potential benefits of MDM, it is worthwhile to understand the potential pitfalls and potential “stalls” in a successful deployment.

Some of the barriers to success are “people related”—a lack of understanding who the consumers of master data are or their expectations, and failing to ensure that the MDM framework meets their needs. Others are “data related,” where data consolidation greatly magnifies the impacts of minor differences across data subsystems. Some barriers are systemic issues related to modeling, standards for information sharing, or general data governance and oversight.

This Checklist Report reviews opportunities where anticipating potential roadblocks can help improve planning, identify issues early in the program, and generally eliminate barriers to MDM success.

There is a misperception that MDM is all about data consolidation. MDM projects are often driven by the expectation of a “single source of truth” for one or more specific data domains (such as customer or product) that emerge as a result of consolidating the various source data sets presumed to represent the same conceptual domain. However, this approach of “data dumping” introduces risks to program success, especially when the expectation is that the resulting records are suitable for all the downstream consumers of master data. In many cases, the data domains for mastering have been already selected, the target master data models have been crafted, and consolidation has begun before there is a general understanding of where the unified view will provide value.

Targeting a data domain for mastering before determining which business processes will benefit and how they will employ a master representation may undermine the requirements, design, and implementation of the overall master data program. An approach that incrementally layers new business processes into the infrastructure means the analysts must review and possibly modify the existing master model for each new consumer application, with potential side effects negatively impacting existing master data consumers.

Having greater visibility into the collective requirements for the uses and quality expectations of master data domains will positively influence master model designs while reducing the need for continuous adjustments. As the first step of your MDM program, identify the pool of consumers and solicit their input regarding master data consumption:

- What master domains does the business process use?
- How do the business processes use the master domains?
- What are the implicit or explicit definitions for the shared data elements?
- What are the data quality expectations in terms of accuracy, completeness, consistency, uniqueness, etc.?

Work with the enterprise scope of master data consumers to assess their data quality and usability expectations, and document their perceived understanding of the definitions for shared master data. This effort will help to identify any potential data conflicts whose resolutions can be addressed as part of the data consolidation plan.
Even those organizations with the most comprehensive governance and IT control still have complex architectures patched together from systems, software, and design philosophies that have evolved and changed over an extended time frame. At the same time, individuals selectively “cut and paste” data from one report, download data extracts into desktop productivity tools, and sometimes even propagate their local data sets into production systems.

Unifying master data from known data sources can be a challenge, yet when reports, extracts, and desktop databases and spreadsheets eventually work their way into pseudoproduction processes, it significantly complicates master data governance. Uncontrolled data repurposing enables ungoverned data copies; because these copies fly under the radar, they are particularly prone to inconsistencies with master data sets. In addition to surveying the create/read/update characteristics and access locations across the portfolio of sanctioned business applications, define and execute processes to assess dependencies that bypass standard IT oversight and control.

There are two main issues to consider: ungoverned uses and ungoverned sources. The set of candidate sources for master data provides a reasonable starting point for both. For each of these sources, monitor any accesses (such as queries, extracts, etc.) that are not part of sanctioned applications to determine if and when the creation of copies of master data bypasses standard procedures. Engage the individuals involved to understand how the data is used and whether the extracted data is used as an input to other applications, either directly (such as a sweeper application that “picks up” table files for integration) or indirectly (such as through publication within shared environments, including wiki Web sites or other common collaboration platforms).

Collect comprehensive visibility by mapping the flow of data from a variety of sources into “information products” such as tables, reports, and even documents and presentations. Backtracking through the creation of information products consumed by end users helps identify intermediate ungoverned data sets and implicit data dependencies. The information production map will help isolate opportunities for reducing or eliminating instances where ungoverned replicas of master data feed into production results.

The overwhelming marketing messages for MDM focus on “creating a single view” of one or more desired master data domains. Consequently, many MDM initiatives are focused on data consolidation to create and populate the master data repository. However, they fail to consider how existing applications will adopt the use of the data in the newly created master environment. Although budgets are approved for the design and development of an MDM environment, not surprisingly, there is little consideration of the effort necessary for enabling its use once it is operational. The risk is that resources spent on consolidation will not yield a benefit if there are no applications using the master data.

Instead, envision an MDM program as a three-stage activity to renovate business processes. Phase One looks at assessing immediate and future needs and creating the infrastructure to support those needs, while Phase Two focuses on operational efficiencies gained through reduced functionality and improved data quality from consolidation into the master repository. The forward-thinking design feeds into Phase Three, where business processes are continuously reengineered to adopt the emerging master data system.

That reengineering may span multiple years, and over that time frame, reengineered business processes must be developed to use the master data environment while existing business processes migrate from their own data subsystems to use the master data system.

Having two layers of master data services will allow existing applications to transition to using the master environment. The first layer provides direct data access, and can be fashioned in lockstep with the development of the master models to enable applications to preserve existing functionality while migrating to the master environment. The second layer enables the evolution of business process services that reduce functional replication horizontally across multiple business processes, such as creating new customer records or introducing new products.

Developing a plan for migrating existing applications to the master data repository using master data services allows for scoping the appropriate levels of effort to go beyond the creation of the consolidated repository and ensure its adoption across the application infrastructure.
NUMBER FOUR

PLAN PROCESSES FOR SYNCHRONIZATION WITH PROPRIETARY SYSTEMS.

All organizations seek to differentiate themselves from their competitors, yet when functional similarities are found within each industry, companies rely on vendor-provided systems to meet their general needs so they can concentrate on what makes their company’s products or services unique. However, this model introduces a critical reliance on proprietary environments, which are often hosted at other physical locations.

From an operational standpoint, proprietary systems perform admirably, but may challenge integration with a master data environment. These vendor system data sets are often hidden, obfuscated, or in proprietary formats. Underlying models are designed to meet the application’s needs and are not necessarily aligned with the organization’s business definitions. When creating a unified view of critical master data domains, these issues raise questions about the ability to extract these data sets, resolve them with the master view, and then share the results back out to the proprietary system.

Companies will need dependable and repeatable methods to consolidate many different sources of master data and make that master version available for different consumers, including the copies of master domains that are stored in external or proprietary systems. To plan specific processes for synchronizing the data in proprietary systems with the master environment, if possible, engage the vendor to determine the best methods for enabling a bidirectional link for exchanging shared master data. This process will include:

- Sharing the conceptual, logical, and physical models
- A review of differences and similarities of shared data elements to assess any need for harmonization
- Access routines or an API to extract copies of master data for integration with the master systems, as well as procedures to publish updates from the master environment back to the proprietary system
- Consistent transformations into and out of the master model
- Exposure of master data services (including lookups and matching) to the proprietary application to reduce unnecessary replication
- Additional services for interoperability with other applications that use the master data

A reasonable approach will institute the additional requirements analysis, planning, integration, and interoperability to ensure that vendor-provided system data is consistent with the master repository.

NUMBER FIVE

ALIGN SUCCESS CRITERIA FOR MASTER DATA MANAGEMENT ACROSS THE ORGANIZATIONAL CHART.

Leaders of the organization may recognize and tout the conceptual value of unifying data domains into a master environment, but that enthusiasm may not trickle down consistently to those at different levels of management. Functional areas of any business are traditionally measured with performance metrics specific to the function, and enterprise initiatives may often be perceived to increase the workload for some project teams while benefiting other projects.

Area managers are more likely to be concerned with their own fiefdoms than with allocating precious resources to support activities outside of their administrative domain. Criteria for project success are often directly related to meeting project objectives, and technical approaches suited to achieving those goals may conflict with the MDM approach. Although it may take time for budget allocations to accurately reflect the increased levels of effort, project timelines and deliverable schedules are not adjusted to include the increased effort.

Resolving misaligned success criteria must involve both the leaders of the MDM program and senior managers in a position to influence behavior across the corporation. It must also focus on addressing the key objections. Some ideas include:

- Create incentives—for instance, recognition or financial compensation—for contributions to enterprise activities such as data governance and master data management.
- Provide additional support to the function manager, to assure him or her that achieving acute project milestones will not be compromised.
- Adjust existing project timelines to reflect any increased expectations for supporting the MDM program.

Finally, it is important to socialize the value of collaboration and to continuously demonstrate that contributions to the improvement of shared data resources provide business benefits to all the staff. Augmenting line-of-business performance scorecards with metrics related to enterprisewide contributions will reinforce the importance of this collaborative effort.
Structural inconsistencies appear at the data element level, and occur when the same data element concept is represented in different data sources using variant data element lengths and data types. As an example, consider a conceptual data element for a postal code implemented in a variety of ways to accommodate different structural variations in different data sets. One application might presume five characters that map to the five-digit U.S. Postal Service ZIP code. Another might use a nine-character field supporting the ZIP+4 format, while a ZIP+4 field in another system uses a ten-character field to include both the digits and the intermediate hyphen. Even worse, data often is misrepresented and you end up with a five-digit code in an expected nine-character field.

Aside from complicating the data integration process, these structural inconsistencies also introduce a risk of applying inconsistent validations when the same conceptual data element has multiple, variant representations. The presumption of adopting the “least common denominator” of the most flexible representation may address the consolidation complexities, but attempting to share the results back to systems with incompatible models will be inconsistent at best and incorrect at worst. Imagine the ramifications these inconsistencies present on analytics and reporting.

A good practice is to collect information about similar data elements to determine whether they truly represent the same underlying business concepts and, if so, influence the originating business applications to modify their representations to conform to a defined standard. Use a data discovery process to collect data element metadata along with attribute names and corresponding business definitions. Use a metadata discipline for assessing similarities among conceptual data elements and for cataloging their formats, structures, and data types.

Any structural discrepancies that are discovered can be shared with the business application owners to help resolve or differentiate variations that may or may not be relevant. Managing these processes under the auspices of an enterprise data governance council empowers the team to define data element modeling standards as well as influence process owners to comply with defined standards.

From a high level, everyone presumes that specific data sets can be collapsed into a master data domain such as customer or product. However, because data models for business applications are organized to address specific functional objectives, the meanings of data concepts may vary slightly, in addition to the structural variations.

These variations may be irrelevant in isolation, but the impacts of these semantic variations are greatly magnified when consolidating data into a single unified view. For instance, what the sales department means by customer may differ slightly from its meaning for customer support (or finance, or fulfillment, etc.); as a result, the ultimate view in the master repository is inconsistent with each of the original sources.

When looking at data domains that are targeted for mastering, determine if the definitions within the candidate data sources are aligned or not, similar to how you would resolve structural variations through data element harmonization. The potential for conflict decreases when semantically consistent sources are brought together. Alternatively, if there are potentially conflicting definitions, or if representations and underlying meanings from different business functions are dramatically distinct, following a process to qualify the master concepts can help in crafting a canonical model for consolidation and sharing.

A model-driven approach to MDM addresses this challenge. Instead of basing a master model on one or more existing data domain representations, consider the hierarchical nature of the underlying entities and the ways those entities can be mapped to the different roles they play across different business processes. Design master models in a way that prevents the introduction of inconsistencies at the entity level as a result of different business contexts. This approach allows core identifying attributes to be associated with the underlying entity but distinguishes the attributes that are specifically affiliated with the roles those entities play.
Our motivation for organizational preparedness is to avoid the pitfalls of aggressively pursuing the technical aspects of master data consolidation, without considering how shared master data will benefit the user communities. Additionally, most of our checklist items strive to develop awareness of complicating factors, to influence proper planning of the MDM system, and to establish good data management practices and processes as part of the program.

A solid first step is to ensure that proper processes are in place to identify consumers of master data, analyze requirements, complete data discovery, and resolve structural and semantic variation—as well as procedures for modeling, integrating data, and migrating applications. Then, you can review the actual business needs for technology and determine what types of tools and methods address the immediate business goals. In some cases, a comprehensive suite of products might be the right approach, but in other situations, the underlying models and services might be insufficient to meet your needs.

To determine what technologies and processes your business needs, wipe the slate clean and envision which tools and technologies are necessary, when they will be needed, and how that affects the acquisition strategy. Data discovery and metadata capture and management tools support the early processes of requirements analysis and structural and semantic harmonization. As the master data environment is designed, a model-based approach to MDM demands good modeling capabilities tightly coupled with the evolution of the master data services layer. Data quality tools will supplement the identity resolution procedures for consolidation and the master data services. In short, obtain a holistic view of MDM and consider all the technology requirements as a whole.

Be sure to allocate time and budget for proper training in the use of acquired technology and for appropriate guidance in leveraging technology to solve business problems. Realize that the tools are most effective when integrated into defined data management and governance policies. People who make purchasing decisions should remember that effective allocation of resources demands a good delineation between the allocation of funds for technology acquisition and those assigned for expertise, guidance, and implementation.

As with any complex system, a successful MDM implementation depends on people, process, and technology. Guidance in this checklist suggests reviewing the program plan and ensuring that the processes and the technologies accommodate the needs and expectations of the people:

- Engage the potential user community and solicit their requirements and expectations, while examining how they may have “gamed the system” by accessing and managing data through ungoverned methods to accomplish their objectives.
- Institute a framework for data governance that actively rewards collaboration and contributions to the enterprise activities.
- Develop a platform that enables both data and application migration.
- Examine the candidate data sources and evaluate differences in structure and meaning that might impact shared use of the master repository downstream.
- Harmonize where possible, but differentiate when necessary.
- Link technology acquisition milestones to the right points in the program plan.

Talend is the recognized market leader in open source middleware solutions. More than 2,500 paying customers around the globe—including eBay, Virgin Mobile, Sony Online Entertainment, Deutsche Post, and Allianz—subscribe to Talend’s solutions and services to minimize the costs and maximize the value of data integration, ETL, data quality, MDM, and application integration. Unlike the small (and quickly consolidating) number of traditional vendors offering proprietary, closed solutions—which can only be afforded by the largest and wealthiest organizations—Talend offers a completely new vision. The company shatters the traditional proprietary model by supplying open, innovative, and powerful software solutions with the flexibility to meet the needs of all organizations. Talend makes middleware solutions available to organizations of all sizes, for all their integration needs. With more than 15 million downloads, Talend’s solutions are the most widely used and deployed data management solutions in the world. The company has major offices in North America, Europe, and Asia and a global network of technical and services partners. Within the last year, Talend was named a Network World open source company to watch and has received a number of other prestigious honors, including The Practitioner’s Guide to Data Quality Improvement (see www.dataqualitybook.com). David is frequently invited to speak at conferences, Web seminars, and on sponsored Web sites and channels, including www.b-eye-network.com. His best-selling book, Master Data Management, has been endorsed by data management industry leaders, and his valuable MDM insights can be reviewed at www.mdmbook.com.

David can be reached at loshin@knowledge-integrity.com.

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.