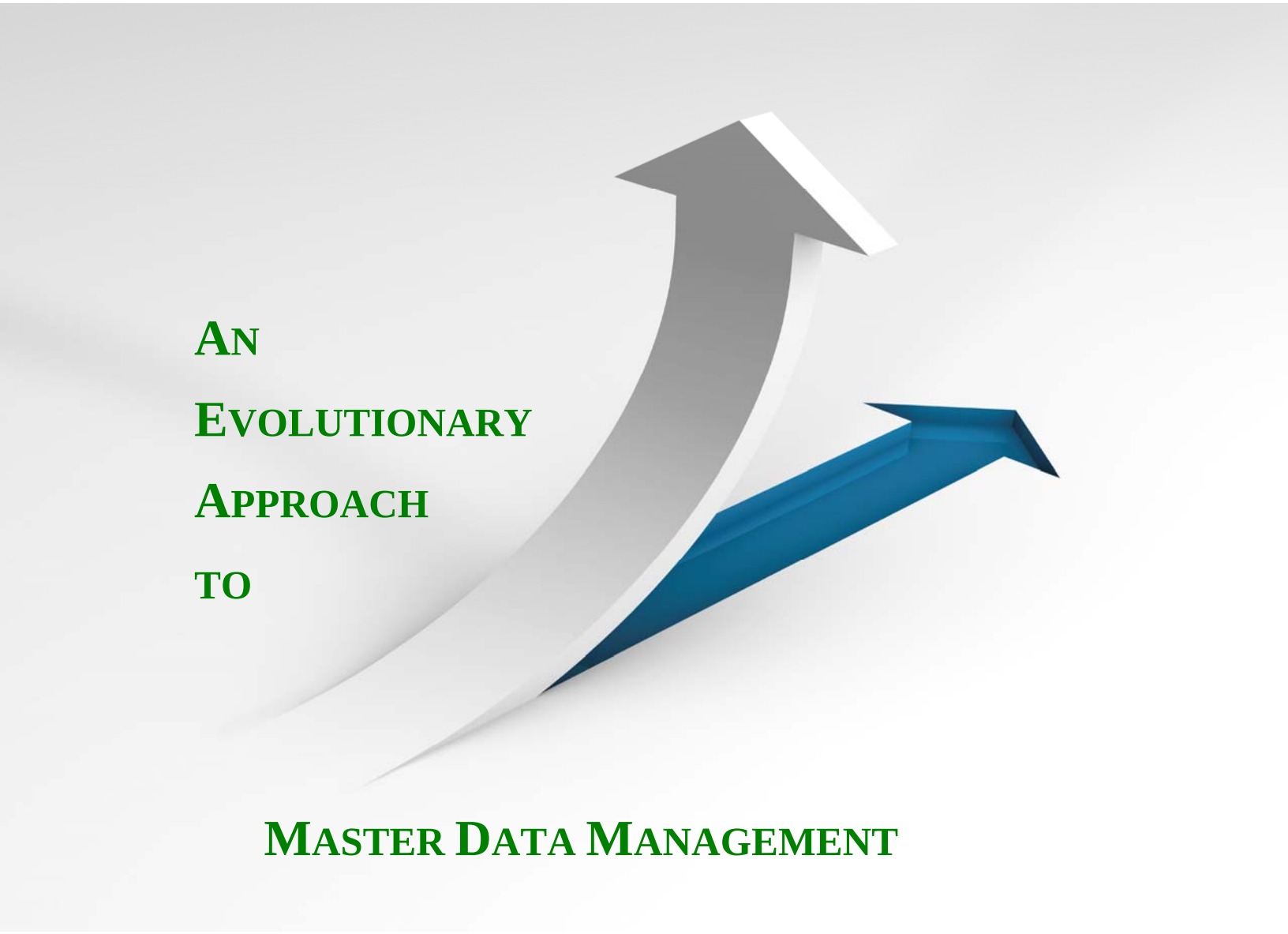
BeyeNETWORK

Global coverage of the business intelligence ecosystem



AN EVOLUTIONARY APPROACH TO MASTER DATA MANAGEMENT

BeyeNETWORK Research Report Prepared for IBM

By Claudia Imhoff and Colin White

January 2009

TABLE OF CONTENTS

EXECUTIVE SUMMARY

A little more than two years ago, we wrote our first research paper on master data management (MDM). At that time, it was a fairly new concept, and our report was one of the more in-depth studies of the new initiative. Since that time, hundreds of articles, presentations, blogs, white papers and case studies have been created and published. The discipline has become more mature and the technology supporting it more sophisticated. Unfortunately in the intervening time, more confusion surrounding master data implementations has also cropped up, leading to conflicting and disjointed implementation advice.

We have taken a fresh look at where MDM is today with this second edition of the research report, pushing the ideas and ideals farther than ever before. Our goal is to establish a solid baseline of understanding for MDM initiatives, the technological environment needed to fully implement a global store of master data, reasonable benefits that can be expected and how you can evolve your MDM initiatives to move toward a fully integrated and functioning MDM environment.

We define MDM as: *A set of disciplines, technologies, applications, policies and procedures used to manage, harmonize and govern the master data associated with an organization's main business entities.*

A fully functioning MDM environment can overcome many of the challenges faced by businesses today – poor data governance processes, redundant and inaccurate master data, inefficient business processes, lost business opportunities, etc. Some of the benefits we found in our research include a reduction in risk, easier auditability of master data, lowered administration costs and improved collaboration across the enterprise. In addition, the system of record, system of entry and system of reference for master data are all clearly defined and conflict-free.

A mature MDM environment includes 3 major technological components: master data applications for maintaining and managing master data in a global master data store; master data services for use by master data applications; and data management services that interoperate with the master data services. These, in turn, support 3 types of MDM processing: operational MDM, collaborative MDM and analytical MDM.

The report closes with practical advice for starting your MDM initiative. There are several starting points for an MDM project. Some organizations begin by gradually upgrading operational applications to directly access a global master data store of the MDM environment. Other companies evolve a data quality management project into a full MDM initiative. Many companies start MDM projects by extending a data warehousing and business intelligence (BI) environment to support master data requirements. Ultimately, the strategic goal is to move toward an enterprise MDM environment whose purpose is an integrated MDM environment that supports all forms of MDM.

WHAT IS MASTER DATA MANAGEMENT?

We define master data as reference data directly associated with an organization's core business entities

Master data has taken on increasing importance in enterprises today, especially those organizations that are implementing ERP, CRM, BI or other enterprise initiatives. Master data forms the fundamental foundational "glue" for these sweeping programs. But what is master data? Unfortunately, there seems to be much confusion surrounding the definition of this basic and essential building block. To that end, we define master data as *reference data directly associated with an organization's core business entities*. Examples of these business entities are:

Person – including customers, employees, suppliers and partners. These may have attributes in common as well as those that make them unique. An example of the person entity is shown in Figure 1.

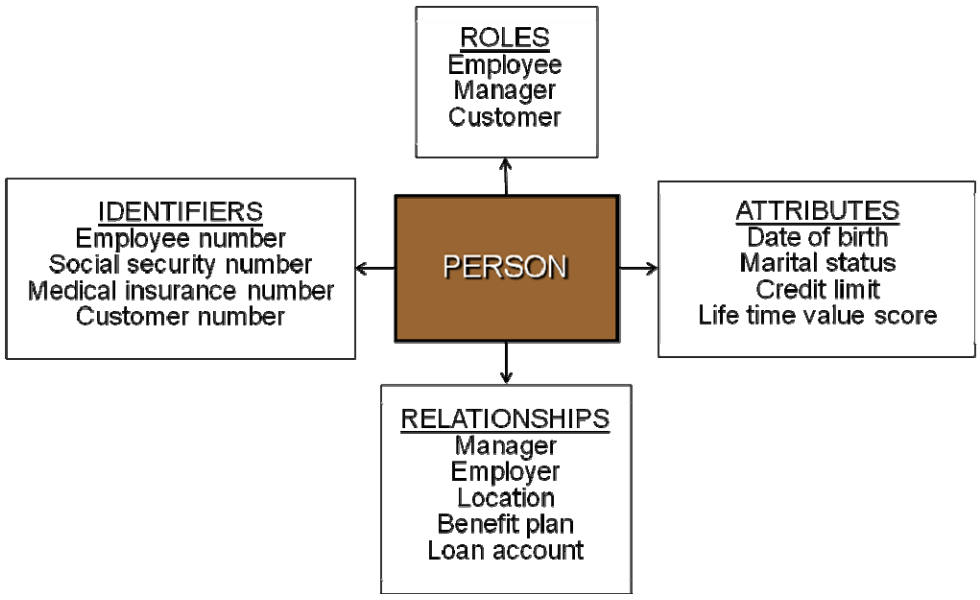


Figure 1: Example of Person Master Data

Things – including products, finances (ledgers) and other assets. Each has its own set of distinctive attributes for each organization.

Places – such as locations, regions, stores or depots. Geographic entities may also share some attributes but the combination of attributes makes each one unique.

In our survey of readers, we found that the *customer* master data entity was the most popular one being implemented (75% of the respondents were planning to implement or already had implemented this entity) with product not far behind (63%). Third place belonged to the financial master data entity (see Figure 2).

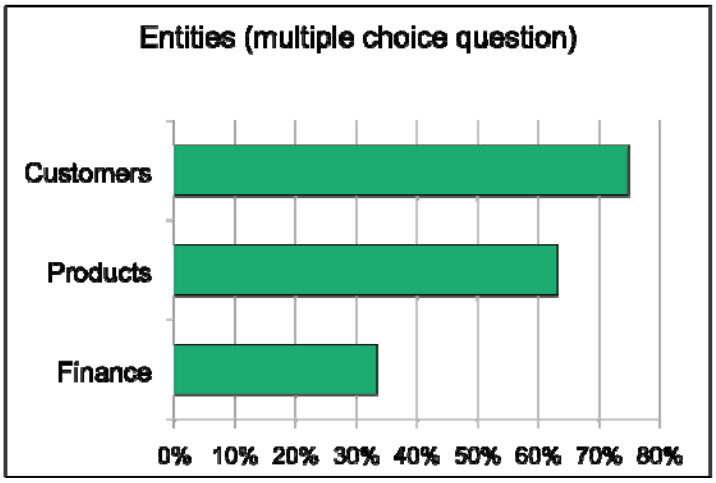


Figure 2: Most Popular Master Data Entity Being Implemented

The MDM environment is a cross-functional and technologically complex one, governing all aspects of master data business entities

While these examples seem simple to understand, the ability to create an environment for their creation and maintenance is anything but simple. It requires a full-blown data management discipline that actively *administers* and *governs* master data enterprise wide. The resulting environment is an organizationally cross-functional, technologically complex, procedurally oriented one, governing all aspects of these business entities. This environment is MDM, and its creation is what delivers true value to enterprises. Therefore, for the purposes of this report, the formal definition of MDM is:

> *A set of disciplines, technologies, applications, policies and procedures used to manage, harmonize and govern the master data associated with an organization's main business entities.*

BUSINESS REASONS FOR MDM

An MDM initiative has to solve business problems to be of value. So what problems can be solved by the implementation of an MDM environment?

DATA REDUNDANCY

It is rare that the data-gathering process produces consistent customer data

In the absence of an MDM function, each system, application and even department within the enterprise collects its own version of key business entities. A good example comes from the collection of customer data. Key attributes such as customer name and address information are collected repeatedly throughout the enterprise – in order entry, billing, customer service and other systems. Unfortunately, it is rare that this gathering process produces the same or consistent data about customers. This leads to the critical difficulty (aside from the storage costs of it) with such redundant data – poor data quality.

DATA INCONSISTENCY

Because of the fractured nature of the master data in operational systems, enterprises spend enormous resources (time, money and people) doing a function that leads to minimal business benefit – reconciliation. Determining a customer's *real* address or name does nothing to enhance the corporation's business revenues. And unfortunately, the reconciliation process must be frequently repeated because there is no mechanism or process in place to capture the data assets garnered from the first or succeeding reconciliations.

BUSINESS INEFFICIENCY

All along the workflows within an enterprise, fractured master data causes ineffectiveness and inefficiencies

Low productivity, inefficient supply chain management, inconsistent customer treatment, customer dissatisfaction and wasted marketing efforts are examples of the types of business inefficiency generated from fractured master data. A customer service representative who must scramble through multiple operational systems to determine the status of a customer's order is not only inefficient, but also risks causing dissatisfaction or alienation of the customer by being perceived as incompetent. All along the workflows within an enterprise, fractured master data causes massive ineffectiveness and inefficiencies, rendering much of the manual effort useless or wasted.

BUSINESS CHANGE

Organizations are constantly changing as new products and services are introduced and withdrawn, companies are acquired and sold, and new technologies appear and reach maturity. These disruptive events cause a constant stream of changes to master data. Without a way to manage these changes, the issues of data redundancy, data inconsistency and business inefficiency are exacerbated. Without MDM, organizations lack a complete view of their customers, products and other critical business entities, which can have far-reaching consequences. Each of the four case studies analyzed for this research paper had very definitive business purposes for their MDM projects. These revolved around two main objectives: improving the productivity of the enterprise, and increasing its revenues or decreasing its costs.

LOCAL MASTER DATA CHANGES LOST

Changes to the master data in applications are often lost to other systems that require those changes

For most organizations, the operational transactional systems (order entry, billing, general ledger, etc.) contain local master data, i.e., master data specifically needed for that application. Changes to the master data in the application are often lost or not transferred to other systems that require the changes.

For example, a customer may tell the order entry clerk of a change in address, but that change may not be forwarded to the billing system. The customer's bill gets delivered to the old address, increasing the customer's frustration (they have to call the company again!) and increasing the inefficiency of the operational processes. This is an example of a local change to master data that should have been incorporated into the global master data.

POOR MASTER DATA GOVERNANCE

With no MDM processes in place, coordinated data governance efforts are greatly hindered. At best, the enterprise may have pockets of governance occurring – each having minimal contact with the other pockets. The efforts to clean up master data are duplicated, the processes are redundant or inconsistent, and the overall efficacy of the governance program is greatly diminished.

Without MDM, the bottom line is a high likelihood of confusion and disorganization across the enterprise. Simple questions such as "How many customers do we have?" become nightmares to answer. Inefficiencies in supply and demand chains go unchecked. Lost revenues in terms of real dollars or in opportunities lost are substantial and unrecoverable.

MDM MISCONCEPTIONS

We found in our research that many enterprises have misconceptions about MDM projects. The more popular ones are described here:

MDM IS PART OF A DATA WAREHOUSING OR BUSINESS INTELLIGENCE PROJECT

Submerging the MDM initiative within the larger BI and data warehousing program is a mistake

Untrue! MDM is not a BI and data warehousing project. Certainly, the BI and data warehousing team may be the first to see the benefits from an integrated global master data store, but to submerge the MDM initiative within the larger BI and data warehousing program is a mistake. When MDM is part of the data warehousing initiative, the tendency is to create master data for usage with BI applications only – that is, historical dimensional data. A separate and properly created MDM environment becomes an essential source of master data or *system of record (SOR)*, once it is cleaned, integrated and loaded into the global master data store, for *all* systems requiring such data – including the data warehouse, operational data store (ODS) or data marts and operational systems. Figure 3 lists the differences between an MDM project and a BI/data warehousing one.

Data Warehouse	Master Data Management System
Purpose is to analyze historical data	Purpose is to maintain accurate MD
Contains transactional and MD	Contains only MD
Data is historical	Data is current and possibly historical
Is generally a consumer of data	Is a consumer and a producer of MD
Technology is oriented toward analysis	Technology is oriented toward data management
Focus is on aggregated and summarized data and MD	Focus is exclusively on MD
Dimensional data not used by other non-BI applications	MD may be used by multiple operational and BI applications

Figure 3: Differences between MDM and Data Warehouse

It is true that MDM and data warehousing initiatives are synergistic, but they may also have some redundancy between them. There is no doubt though that data warehouse environments become simpler to implement and sustain if a global master data store exists. The designs, processes and data management procedures are easier to create and are less intricate to perform.

Data warehouse implementers can focus their efforts on integrating only transactional data from various operational system data sources since the heavy lifting of master data integration has already been done. The implementation team has a ready source of current, consistent, integrated business entities from which to develop their historical dimensions for analytic purposes. The remaining hurdle for dimension creation is the establishment of reliable interfaces between the global master data store and the data warehousing environment.

MDM IS SOLELY FOR MAINTAINING DATA CONSISTENCY ACROSS OPERATIONAL APPLICATIONS

MDM is about actively managing master data across the enterprise rather than simply *collecting* it in each transactional system

Again untrue. MDM is all about actively managing master data across the enterprise rather than about simply *collecting* it in each transactional system. MDM does help improve the consistency of master data by coordinating it between operational systems, but it has a much broader role throughout the enterprise. MDM creates its own environment, separate and distinct from both data warehousing *and* operational processing. MDM administers the global or enterprise version of master data for use by both operations and data warehousing systems. Figure 4 lists the difference between MDM and the operational applications.

Operational Systems	Master Data Management System
Purpose to process daily transactions	Purpose to maintain accurate MD
Contain transactional and master data	Contains only master data
Data is current	Data is current and possibly historical
Is generally a producer of data	Is a consumer and a producer of MD
Technology is oriented toward transaction (Tx) processing	Technology is oriented toward data management
Focus on processing detailed transactional data – MD is a byproduct	Focus is exclusively on MD
Tx and MD may be used by multiple operational and BI applications	MD may be used by multiple operational and BI applications

Figure 4: Differences between MDM and the Operational Systems

Just as in the previous section, having an MDM environment makes life simpler for the operational side of the house as well. These applications become simpler to implement and sustain because they can access the global master data rather than managing it locally (and redundantly) in their own databases.

For some operational systems, there may be a need to maintain some attributes of master data locally through their systems of entry, while global master data is maintained within the MDM environment. We discuss global and local master data later in this report. If this is the situation in your organization, then IT must determine how to control the creation of master data in potentially two locations (operational applications and the MDM environment).

For mature MDM environments, the master data processing may be replaced completely by the MDM applications

For mature MDM environments, the master data processing may be replaced completely by the MDM applications. This generally means new operational applications are easier to design and implement, but could require major changes to existing ones. The trick is to ensure that each operational application is properly interfaced with the appropriate master data in the global master data store. This is certainly not trivial, but is easier and requires less maintenance than trying to coordinate multiple operational applications and their associated inconsistent sets of master data.

MDM IS JUST A DATA INTEGRATION PROJECT

MDM initiatives must develop disciplines such as governance policies and procedures in addition to selecting data integration technologies

MDM projects consist of so much more than just data integration technologies and processes. From our definition of MDM, you can see that MDM initiatives must develop disciplines such as governance policies and procedures in addition to selecting appropriate data integration technologies. From our case studies, we learned that the most significant effort in MDM projects came from definition resolution and the creation of master data procedures and policies. Creating standard definitions for customers, products, suppliers, etc. is very time-consuming and gets right to the heart of severe political and cultural disputes and disparities in any enterprise.

Further, MDM requires that many difficult procedural questions be answered before an organization can garner its full value. For example, consider the following: Who can create, change or delete master data? What relationships and hierarchies are valid? Who can access the master data? These questions lead to the ultimate procedures for who controls these critical and important business entities.

For MDM to reach its full potential, our research shows that it should be integrated with other enterprise functions and initiatives such as CRM and ERP systems, and the data governance function. Otherwise, it is likely that the MDM effort will result in yet another silo of integrated data, only adding to the growing list of silos and contributing to the confusion within the enterprise.

MANY COMPANIES THINK MDM IS JUST ABOUT DATA QUALITY TECHNOLOGY

There can be no doubt that data quality is a major process or service in the creation of a global master data store. However, a fully functioning MDM environment consists of more than data quality services. It must also have robust data integration services, overarching database management functions in addition to several horizontal and vertical services involved in an MDM environment. (We have an entire section devoted to the components of an MDM environment later in this document). The environment must contain specific master data applications (see next section) and interface with other systems requiring access to the global master data store.

TWO POPULAR MDM APPLICATIONS – CDI AND PIM

Today, most organizations implement their first MDM initiative in one of two flavors by developing either a customer master or product master data store. These two business entities have the biggest impact on the enterprise and demonstrate the value of MDM with hard dollar ROIs. Because of this, most MDM vendors have focused on customer data integration (CDI) and/or product information management (PIM) for their MDM application offerings. The applications integrate these business entities through syntactic and semantic processes (see below for the definition and examples of these forms of processing).

CDI applications have a heavy focus on traditional data quality capabilities

CDI applications began with a heavy focus on traditional data quality capabilities – name and address hygiene, merge/purge processing and so on – using predominantly syntactical processing. The technologies come with hundreds of prepackaged rules or

algorithms to help them perform the syntactical processing. These applications can be used for other *Person* areas such as supplier and employee, giving them more depth and breadth.

PIM applications must be able to capture the expertise from product managers or product subject matter experts

PIM applications must support predominantly semantic processing since product standards are not as prevalent as customer ones. These MDM applications must be able to capture the expertise from product managers or product subject matter experts, resolve any discrepancies and then integrate the product data based on rules generated from the captured expertise.

Please note, however, that all MDM applications should be able to use both semantic and syntactic processing regardless of whether they are CDI or PIM offerings. Without doubt, the MDM initiative will require that the supplied MDM solution be customized or added to regardless of whether the project involves customer or product data. Input from company experts must always be easily accommodated and maintained.

THE BENEFITS AND HURDLES OF MDM INITIATIVES

MDM benefits can be split into the demand side and the supply side of the business

MDM benefits can be split into the demand side and the supply side of the business. Demand side benefits come from customer-oriented master data projects and include better coordinated customer communications, improved service, better determination of potential discounts for high value customers, more accurate billing, increased understanding of a customer's *next best product* as well as improved up-sells and cross-sells. Ultimately, the biggest benefit may come from the boost in the customer's overall satisfaction with your company.

The supply side benefits are generated from product-oriented master data initiatives. These benefits include more efficient supply chain management, increased speed to shelf, reduced merchandizing or inventory costs, more accurate profitability analysis and reduced errors in order fulfillment. For manufacturers, a large benefit comes from a better understanding of product life cycles and processes leading ultimately to the reduction in product failures.

Regardless of the type of master data project, the first step must be to create a business case stating not only the business need for integrated master data, but also the benefits that will be generated from its creation. The business case will, of course, differ according to the master data entity you pick. The following benefits were documented from various MDM projects and vary from intangible or hard-to-measure benefits to quantifiable, tangible ones. When developing a business case, incorporate as many of these benefits as are reasonable along with the means for demonstrating the benefit.

MORE ACCURATE REPORTING AND DECISION MAKING

Integrated master data is the basis for many strategic and operational decisions and

reports. Master data is particularly important for regulatory and financial reporting since discrepancies in these reports can have dire consequences on the executives as well as the enterprise as a whole. Consistently created and higher quality master data ensures that reports are populated with correct data and that the decisions based on these reports are reliable.

More accurate reporting and decision making seem to be the most often cited benefits from MDM case studies

While difficult to quantify, more accurate reporting and decision making seem to be the most often cited benefits from MDM case studies. Given the importance of regulatory and other forms of compliance, it is easy to understand why this benefit is one of the more important, though unquantifiable, ones.

EASIER AUDITING

Most enterprises have entered the age of corporate accountability, and many executives are motivated by auditing paranoia. Unfortunately, limited visibility into data lineage inevitably leads to auditing nightmares. A significant benefit of MDM environments then is the creation of a single source of master data. This leads to the ease with which audits of business entity data origins are performed.

For your business case, determine the cost of a data audit before the implementation of the MDM initiative. Then determine the savings in time and effort for the same audit after the implementation.

REDUCTION IN RISKS

An error in master data can have significant impacts throughout the enterprise

Because master data is used in many critical applications, an error in it can have significant impacts throughout the enterprise. For example, an incorrect customer address can mean orders, invoices, product shipments and bills will all be sent to wrong addresses. Likewise, an incorrect product price can be a marketing and sales catastrophe, and incorrect account numbers in a company's general ledger can lead to huge fines or even jail sentences for the CEO. Consistent master data has the obvious benefit of reducing an organization's overall risk by eliminating the types of errors listed here. Again the implementation team would be wise to document these types of problems before embarking on their MDM project and measure their reduction after the implementation.

REDUCED ADMINISTRATIVE COSTS

There can be no bigger drain on an enterprise than the unproductive effort of reconciling different versions of business entities. Creating a global master data store eliminates most if not all of this inefficient and wasteful exercise. If enterprises actually tracked the costs of reconciliation activities, these would often outweigh the cost of implementing an MDM environment. Therefore, the business case for MDM should consider measuring the pre-implementation costs of reconciliation.

A single source of master data reduces the number of interfaces that must be maintained throughout the enterprise

In addition, having a single source of master data reduces the number of interfaces that must be maintained throughout the enterprise. The costs of maintaining these interfaces should be examined before implementing the MDM initiative as well.

A tangential benefit is the creation of a master data governance and best practices

function as a result of MDM initiatives. The need for this function becomes obvious almost immediately after an MDM project is funded.

IMPROVED COLLABORATION

Because MDM success is heavily dependent on cooperation between business units and with IT, collaboration becomes a natural component. We recommend an MDM implementation include the creation of some form of master data collaborative workbench for use by the various factions. This collaboration will also greatly enhance the master data governance and best practices function as well.

The benefit to the organization is improved communication, better cross-functional cooperation, reduced redundant work efforts and so on.

HURDLES

Any new initiative will run into potholes or hurdles that it must overcome; MDM is no exception. We found in our survey of BeyeNETWORK readers that nearly 60% said poor governance was the biggest challenge to the success of their MDM initiative. Business cooperation was a close second with more than 50% voting for it. Figure 5 shows the responses for all five hurdles suggested.

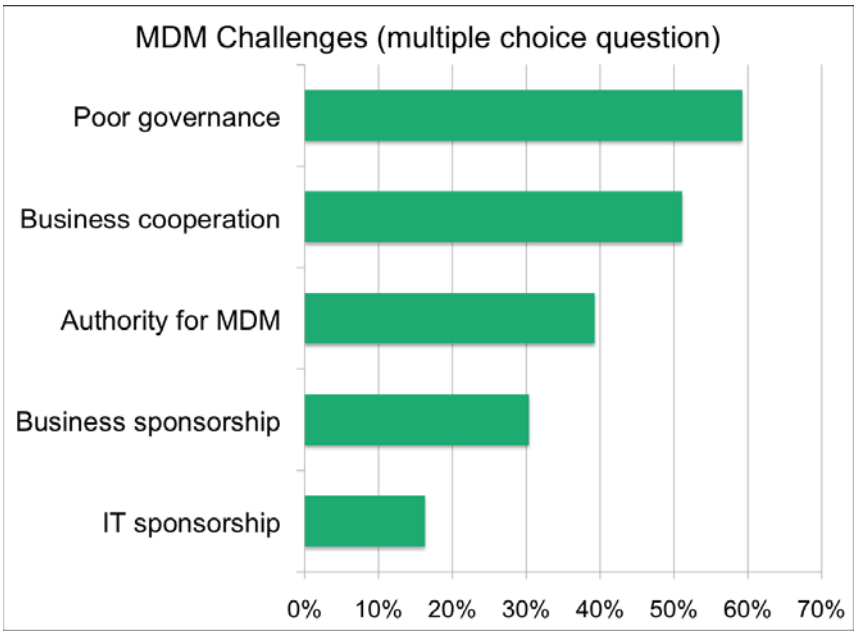


Figure 5: Challenges Faced in an MDM Implementation

Many of these hurdles are cultural in nature and may constitute the biggest challenges. The rest are technical ones that can often be addressed using technology

and products from vendors such as the sponsors of this research paper.

CULTURAL HURDLES

GENERATING THE BUSINESS CASE FOR MDM

The absence of master data results in many visible difficulties such as the inability to reconcile data across multiple business units, poor information quality, long decision-making cycles, loss of revenue opportunities or the reduction in profits. A sound business case for MDM can be built around any of these business problems. In creating this case, you should focus on tangible benefits such as improved operations, streamlined IT and business processes, and improved profits. These have real dollars associated with them that can be tracked and reported upon. The intangible benefits are also useful to mention such as higher customer satisfaction, improved supplier relations and higher quality data; but they may not have real dollar metrics to back them up.

A second part of the business case should be devoted to the creation of the governance and best practices function for the master data entity involved. The team must set aside time to develop the procedures and policies governing the creation and maintenance of the master data being implemented and to assign the appropriate personnel to the function.

OBTAINING AND SUSTAINING EXECUTIVE BACKING

In any enterprise-focused initiative, there are bound to be turf wars, disagreements between divisions or departments, or arguments over the definitions of business entities. An MDM initiative brings all of these conflicts into focus. An executive steering committee, fully committed to the ideals of the MDM initiative, should be considered if the MDM team knows these problems will be difficult to solve. The committee's purpose is to ensure that the initiative moves forward with full executive support.

Moving from a LOB view of master data to the enterprise view requires the entire organization accept the global master data store as the master data system of record

The executives themselves may be actively involved in dispute resolution and in assuring compliance to the new master data, at least during the first few projects. Overcoming these political barriers can be the most difficult hurdle of all. Moving from a line-of-business (LOB) view of master data to the enterprise view requires that the entire organization accept the global master data store as the master data system of record. This means that there must be full agreement or at least acceptance of the common business definitions for these critical business entities throughout the enterprise.

An integration competency center and a data governance and best practices function can play important roles in overcoming these barriers. These functions are responsible for defining and managing the governance policies and procedures for master data. A comprehensive governance policy is an essential MDM success factor as well, and executives must show active support of this effort. These activities will ensure that the MDM initiative has the proper authority to enforce MDM compliance throughout the organization.

Ensure Sufficient Funding to See the Initiative Through to its Completion

Although there are significant benefits to implementing a full scale enterprise-wide MDM environment, such a project is an expensive undertaking. MDM should be viewed as a multiyear program that is implemented as a series of incremental projects. In many cases, tactical projects may be needed to meet short-term business needs. The overall long-term objective, however, is to build a full enterprise MDM environment; and organizations should plan the funding with this strategic objective in mind, even if they deploy MDM applications bottom up in small incremental steps.

The IT sponsor (CIO or VP of IT is recommended) plays an important role in communicating the funding needs for the initiative. This executive should be part of the steering committee and will represent the technological needs to the team.

Technical Hurdles

Many of the technology issues associated with MDM have already been discussed. However, a summary of the key technology issues follows.

Creation of Flexible Master Data Business and Data Models

You will not – cannot – know all the master data your organization will ever need

These models are the road map to success. You will not – cannot – know all the master data that your organization will ever need. This need evolves over time, just as the business changes and grows with market shifts and cultural changes. Therefore, the designers responsible for the master data models must take extra care to create models that are flexible and can accept additions without causing significant perturbations to the rest of the environment. Some MDM vendors provide customizable industry templates and data models that can give you a quick start to an MDM project.

Management of Master Data Quality

Data quality involves identifying and defining the master data attributes, assessing the quality of those attributes and correcting problems where possible. This can be a significant task, particularly when it comes to reconciling definitions across operational data sources and gaining agreement from the business units where definitions or usage conflict (see cultural hurdles).

Data quality activities are done either on all sources of data when the MDM solution is first implemented or on a source-by-source basis as sources are brought into the MDM initiative. These activities are also applied to the data in each source as it changes on an ongoing basis. There are a variety of different data quality and data profiling tools that can aid in this process. For specific types of master data entity, there are specialist tools that can assist with address pattern matching and syntax, or with analyzing the business semantics of products and parts documentation.

Constantly Changing Master Data

MDM applications must handle complex changes

Just as the master data models must remain flexible to accommodate new master data elements, the MDM applications must handle complex changes, not only to the content of the data, but also to the hierarchies, relationships and business rules established within and between the data. Change is inevitable; therefore, when choosing an MDM technology, make sure that it can accommodate changes.

Upgrading Operational Systems to Directly Use the Global Master Data Store

The challenge is to ensure that this transition is smooth in terms of the switch over from operational systems (the SOEs) to using the MDM environment. As more and more of the SOEs are replaced with the MDM applications, each operational application with data entry capability for that master data must be either altered to eliminate potential dual entry of that data or redirected to the global master data store for entry and update capability.

If it is not possible to modify or replace an SOE to directly using the global master data store, then facilities are required to integrate the SOE local master data into the MDM environment so that its system of record is kept current.

Identity Management Especially for Customer Master Data

MDM environments must be able to create global identifiers to handle complex and evolving customer relationships

Because customers can be global in nature, have myriad relationships with the enterprise as well as with each other and are constantly changing these relationships, it is imperative that the MDM environment be able to create global identifiers that can handle these complex and evolving situations. This is often a cultural issue as well. The enterprise must agree to, and adhere to, the new global identifiers and make sure that the MDM function is aware of changes or updates to these identifiers and relationships.

Building a Sound MDM Infrastructure

There are a great number of companies that offer partial or incomplete MDM solutions. These may be considered best-of-breed vendors. They have focused a great deal of research and development on a very specific technological aspect of the overall MDM environment such as the ability to improve the quality of customer or part data. While they may have the best solution for a particular aspect of MDM, you will have to ensure that each piece of technology can integrate into the overall MDM infrastructure. Also, because of the number and complexity of the disparate data environments and systems involved in MDM processing, it is crucial that MDM solutions employ a common master data integration architecture. This architecture should be consistent with your enterprise data integration strategy and solutions.

Vendors offer an end-to-end solutions encompassing all or most of the aspects found in a mature MDM environment

Many vendors now offer an end-to-end solution encompassing all or most of the aspects found in a mature MDM environment. In many instances, these companies have merged or acquired the needed technology to complete the MDM technological landscape. While these suite solutions may not have the individual depth found in the best-of-breed approaches, they do have the advantage of reducing the need to create

custom adapters and bridges to integrate products from different vendors. You must determine what works best in your current IT environment.

MDM CONCEPTS AND TECHNOLOGIES

WHAT IS MASTER DATA?

As we discussed earlier, master data is *reference data* about an organization's core business entities. However, not all reference data is master data. A high percentage of reference data consists of lookup, code and domain value tables that are used to encode and validate information in IT systems. An example is a two-column *Marital Status* code table where one column contains a single digit status code and the second column details what that code means – 1 for married, 2 for single, 3 for divorced. These types of reference tables consist mostly of stand-alone and single-level data structures.

Master data is reference data, but not all reference data is master data

Master data, on the other hand, is more complex. It not only documents key business entities, but also the relationships within and between those entities. Product and organizational hierarchies are good examples of typical master data structures. Compared with other types of reference data, master data attributes, hierarchies and relationships change more frequently; and this is why good tools for managing and tracking those changes are crucial in MDM.

There are of course gray areas between the two types of reference data. An example is a two-column *USA State* table where one column contains the two-letter code for a state and the other column details the full name of that state. The code *OR*, for example, stands for the state of *Oregon*. Is this a code table, or a data table for customer address master data? In these types of cases, developers have to decide based on business usage. Code tables are typically *local* in nature, whereas master data is more likely to be *global* data that applies to the whole organization.

Although most master data is used throughout the enterprise, there are situations where master data is unique to a specific application, business group or location. In this situation, the master data may not be migrated to the global master data store, but will instead continue to be maintained locally by a system of entry operational application.

SYSTEMS OF RECORD, ENTRY AND REFERENCE

Three key terms in MDM are system of record, system of entry and system of reference (see Figure 6).

The global SOR is the golden copy of master data

A master data *system of record* (SOR) is an application system that is responsible for managing a current and consistent version of any piece of master data and its associated metadata. This master data is the so-called *golden copy* that represents the globally accepted true version of master data in an organization. The ultimate goal of an enterprise MDM environment is to manage the SOR for all enterprise-wide master data – this enterprise-wide view of master data is called the *global SOR*.

An SOE may maintain local master data or the global SOR

A master data *system of entry* (SOE) is an application system responsible for creating, updating and deleting master data. The master data associated with an SOE may be stored and controlled locally or managed in conjunction with the master data software responsible for the global SOR. When controlled locally, the SOE is managing *local* master data. ERP and CRM systems are examples of applications that manage local master data. The problem with local control is that there is no globally accepted SOR; and when conflicting and overlapping master data occurs in multiple SOEs, there is no golden copy to identify which piece of master data is the correct version.

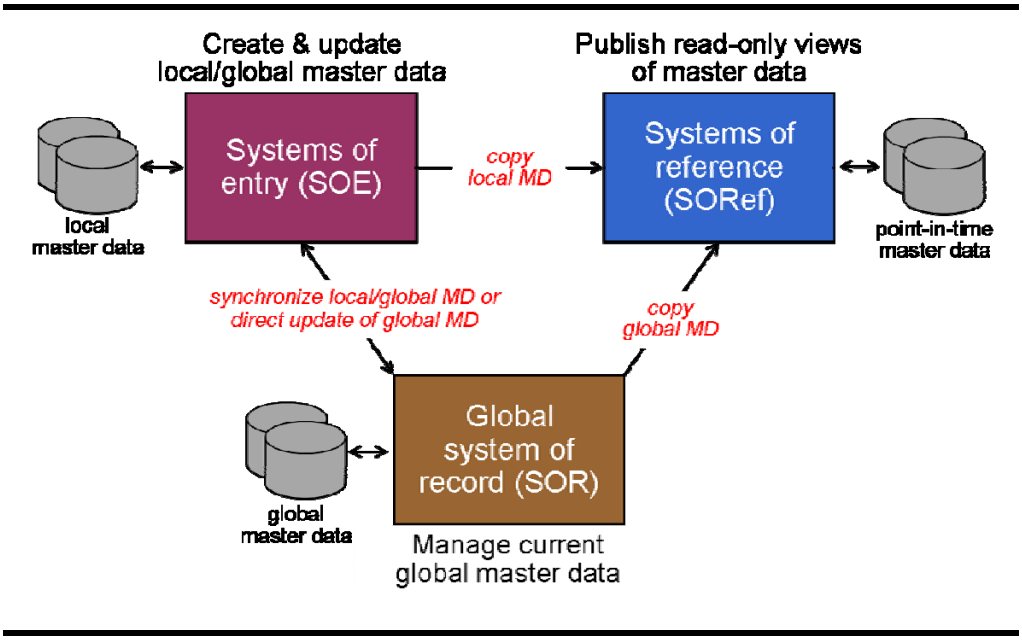


Figure 6: Systems of Record, Entry and Reference

In an MDM environment, custom or vendor-supplied master data software is responsible for ensuring that the master data maintained by the SOE is consistent with the global SOR. In this research, we use the term *MDM system* to describe this master data software. The SOE can continue to maintain its master data locally, in which case the MDM system synchronizes the local master data with the global SOR. Over time, the SOR may be modified to directly create and update the global SOR managed by the MDM system, which eliminates the need for synchronization.

A SORef is a point-in-time and read-only copy of master data

A *system of reference* (SORef) is an application system that publishes point-in-time views of master data. By definition, the master data published by a SORef is read-only historical master data. A data warehousing system is an example of a system that can act as a SORef.

Figure 7 summarizes the main characteristics of the three types of master data systems. An SOE produces and maintains local and global master data. The focus of an SOE is supporting the operational processing of an enterprise.

A global SOR consumes and synchronizes relevant local master data and manages global master data in the master data store. The focus of a global SOR is on master data management.

A SORef consumes local and global master data and publishes historical views of that master data. The focus of a SORef is on reporting and analysis.

SOE	Global SOR	SORef
Producer of local/global MD	Consumer of local MD & producer/controller of global MD	Consumer of local/global MD
Focus is operational processing (including MD maintenance)	Focus is master data management	Focus is MD publishing for operational data reporting and analysis
Current MD	Current & possibly historical MD	Historical MD
Read/update MD	Read/update MD	Read MD

Figure 7: Summary of Types of Master Data Systems

BUSINESS ENTITY MDM VERSUS ENTERPRISE MDM

Master data solutions may address a specific business entity or multiple business entities. We have already discussed how CDI solutions focus on customer master data and PIM solutions are for handling product master data. In an ideal world, we would like to have a single enterprise MDM solution that can manage all of our business entities. However, is such a solution really possible?

A single MDM solution is unlikely because each business entity has its own unique requirements

A single MDM solution is unlikely because each business entity has its own unique requirements. What is needed instead is a set of horizontal MDM services that are common to all business entities (hierarchy management, for example) and a separate set of vertical services that are specific to each particular business entity (name and address cleanup for customer data, for example). Vendor and in-house developed

MDM vertical applications can then use these two sets of services based on business requirements.

An example of how master data processing may differ by business entity can be found in the master data reconciliation processes of CDI and PIM products.

An example of business entity master data differences: syntactic and semantic master data analysis

Master data reconciliation may involve syntactic (from the Greek word *suntaxis* meaning *put in order*) and/or semantic (from the Greek word *semantikos* meaning *significant*) data checking. CDI solutions typically use a business rules-driven syntactical process to perform name and address cleanup. The business rules in this case define a set of standards for matching and arranging a group of related attributes (names and addresses) in a systematic and consistent way.

PIM solutions, on the other hand, employ semantic analysis to examine the relationships and usage of words and phrases in product information. This enables the matching of product records where different values are used to mean the same thing. An automated learning process that examines word patterns in sample input data creates the business rules in this situation.

Figure 8 shows examples of both syntactic and semantic data checking. The top example shows how three different records for the same person, *Claudia Imhoff*, can be reconciled into a single master data record with a standard name and address format.

In the bottom example, the meaning of the term *HP* can only be determined based on how it is used with related words. In the first three records, the term means *horsepower* because it is associated with motor *voltage*, whereas in the fourth record it means the company *HP* because it is associated with a *printer*.

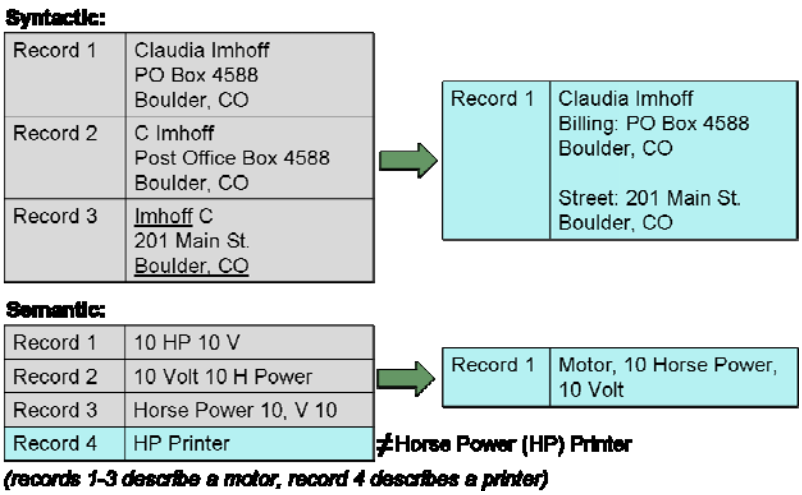


Figure 8: Syntactic and Semantic Master Data Reconciliation

THE ULTIMATE GOAL: ENTERPRISE MDM

Figure 9 shows an example of a typical starting point for an MDM environment where operational production systems continue to maintain current master data and transaction data in operational databases.

The operational MDM environment in the figure is responsible for managing a global master data store – the global SOR – and synchronizing the data in this store with local master data in the operational product systems (i.e., the SOEs). This synchronization process may, or may not, require changes to the SOE applications depending the complexity of those applications and on whether the local master data can be accessed directly by a data interface or has to be accessed indirectly by application interfaces.

A data warehouse publishes historical master data

A data warehousing system publishes historical master data copied from the operational MDM environment and also possibly from operational production systems. The master data is then combined with historical transaction data for reporting and analysis purposes.

This type of MDM environment is the least disruptive to existing operational and analytical applications and enables an organization to gradually evolve to a full enterprise MDM environment.

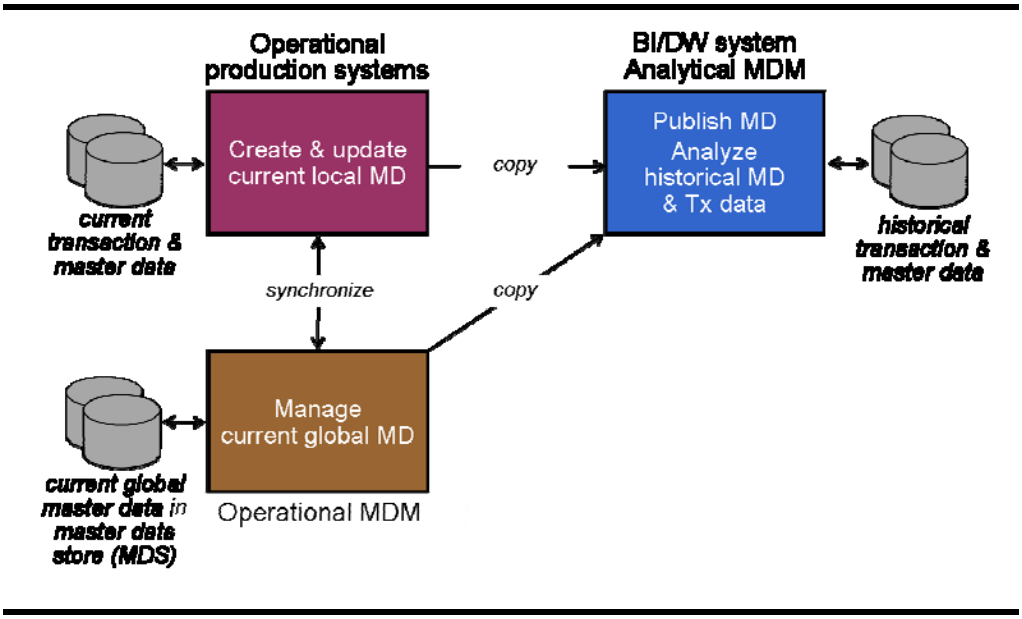


Figure 9: An Initial MDM Environment

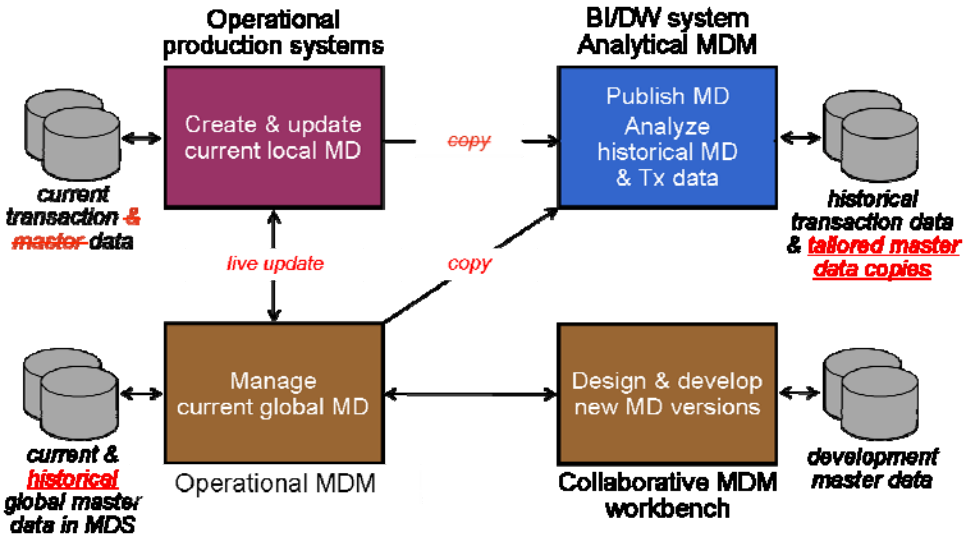


Figure 10: Enterprise MDM

Figure 10 shows the ultimate environment for managing master data, an enterprise-wide MDM environment. In this type of environment, operational production applications no longer maintain local master data, but instead directly create and update master data in the global master data store. This approach eliminates the need for the MDM system to synchronize local and global master data, and for the data warehousing system to copy master data from operational production systems.

Migration to an SOA may provide an opportunity for enterprise MDM

In reality, it is unlikely that all operational systems can be updated to directly access the global master data store. However, as operational applications are replaced, this type of approach to MDM should be considered. Migration to a services-oriented architecture (SOA) often provides the opportunity to evolve toward an enterprise MDM approach.

In enterprise MDM, BI and data warehousing systems obtain required master data from the global master data store. In some cases, depending on performance and ease-of-use requirements, historical master data may also be managed in the master data store, which simplifies the data warehousing environment.

Another option in an enterprise MDM environment is the addition of a collaborative master data workbench that enables business and IT users to design and develop new versions of master data. An example here would be the master data required to support the introduction of a new range of products.

TYPES OF MDM PROCESSING

Types of MDM processing

We can see from the previous discussion that there are three types of MDM

processing: operational MDM, collaborative MDM and analytical MDM. These three terms are used in this report because they are frequently employed by vendors, other consultants and analysts. In reality, however, MDM is really an operational process since the primary objective of an MDM environment is to manage the current global SOR for master data. It's better to address and reconcile master data issues and inconsistencies where the data is created, rather than in downstream systems such as a data warehouse.

Operational MDM is the main focus of an MDM environment. It separates the operational environment into two types of processing. One type manages and maintains local transaction data, while the second type maintains global master data. The goal of operational MDM is to eliminate the need for SOEs to manage and control redundant local master data in operational systems. Another, but less important, goal is for the operational MDM environment to also maintain historical master data in addition to current master data.

An interactive and **collaborative MDM** workbench that supports the development of new master data structures and the maintenance of existing ones can be used to extend operational MDM. This type of workbench is often categorized as collaborative MDM because it enables groups of users to cooperatively work together to create new versions of master data.

An **analytical MDM** environment is a system of reference (such as a data warehousing system) that publishes historical master data for use by reporting and analysis applications. The master data published by a SORef can be a complete historical copy of an SOR (single view of the customer, for example) or a subset of an SOR (regional view of customers, for example) that has been tailored to suit the requirements of specific business users and user groups. These requirements are based on local needs and the security requirements of the organization.

Analytical MDM may derive additional master data

Additional master data may be derived by the analytical MDM environment. An example would be customer lifetime value scores that are created by an analytical application. These metrics should be stored with the SORef, rather than with the SOR. This is because the derived master data metrics are based on the point-in-time SORef version of the master data, rather than on the current SOR master data.

Of the BeyeNETWORK MDM survey respondents who were planning, implementing or have implemented an MDM project, 20% were using analytical MDM, 26% were using operational MDM and 54% were using both types of MDM processing.

COMPONENTS OF AN ENTERPRISE MDM ENVIRONMENT

There are many different approaches to maintaining and managing master data. Before discussing these approaches in detail, let's first identify the ideal architecture of an enterprise MDM environment. As we shall see later, this architecture is achieved through an iterative and evolutionary development process.

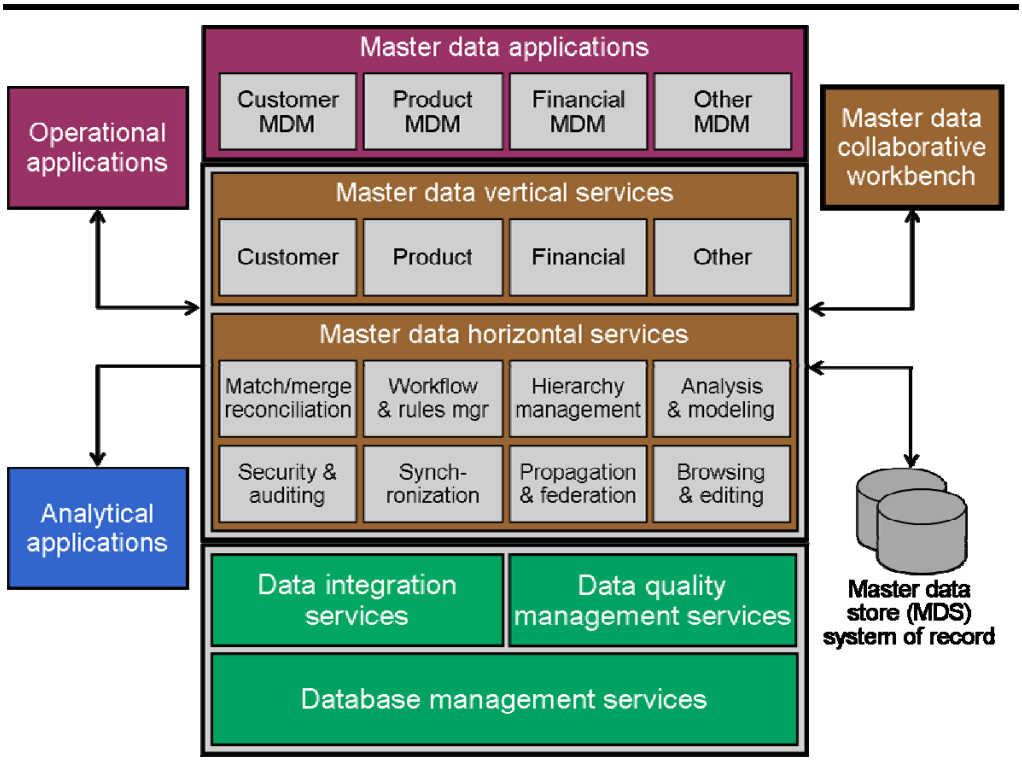


Figure 11: Components of an MDM Environment

Three main enterprise MDM components

Figure 11 shows the main applications, tools and services of an MDM environment. Such an environment consists of three main components.

Applications for Maintaining and Managing Master Data in a Master Data Store

These applications may be new applications provided with the MDM system, new applications developed by IT to take advantage of the tools and services provided by the MDM system, or operational applications that have been upgraded to work in conjunction with the MDM environment. A master data collaborative workbench as well as analytical applications that consume master data from the MDM environment can also be considered a master data application.

Horizontal and vertical master data services component

Master Data Services for Use by Master Data Applications

Master data vertical services support the processing needs of specific master data business entities. Name and address data cleansing for customer data is an example of such a service. Master data horizontal services support any type of master data entity. The most important services are detailed in Figure 11. Some of these services, such as data analysis and modeling, are used during application development, while others, such as master data reconciliation and synchronization, are used at runtime.

Underlying data management services component

DATA MANAGEMENT SERVICES THAT INTEROPERATE WITH THE MDM ENVIRONMENT

Most of these data services (data integration, data quality management and database management) are likely to be already in place because they are used by other IT systems. However, in some cases these services may need to be upgraded to support the needs of an MDM system. This may especially be the case in the area of data quality management.

We discuss selecting vendor products for building an MDM environment in a subsequent section of this study.

MDM PROCESSING TECHNIQUES

Of the many approaches for building an MDM environment, some enable operational MDM, while others focus on analytical MDM. The ideal approach is one that enables the organization to evolve to a full enterprise MDM environment that supports both operational and analytical processing.

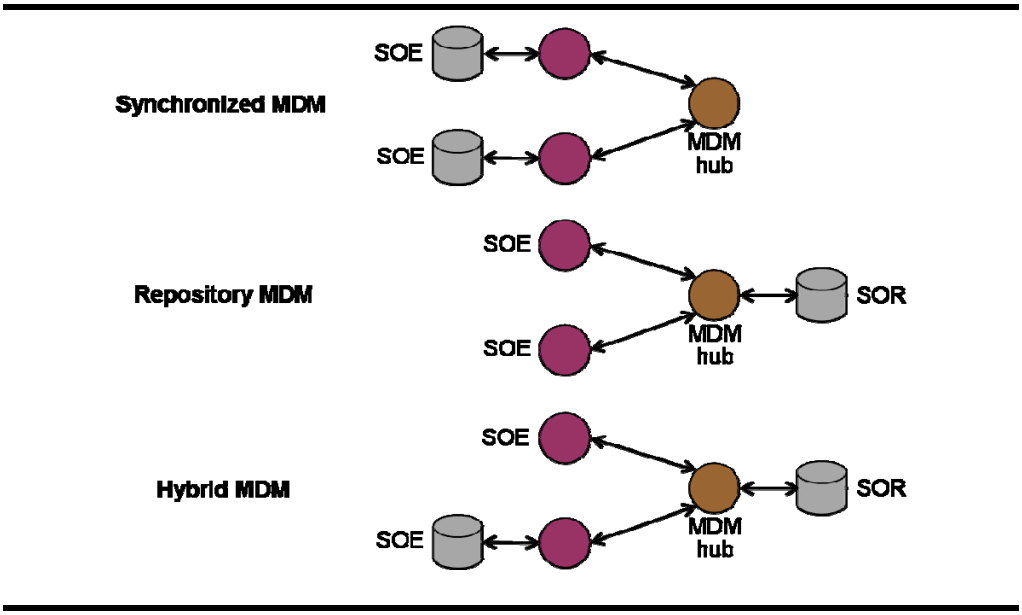


Figure 12: Approaches for Supporting Operational MDM

OPERATIONAL MDM

Figure 12 shows three main approaches for supporting operational MDM. Each approach employs an MDM *hub* that provides vertical and horizontal master data services that operate in conjunction with underlying data management services for data integration and so forth. The actual services provided by the hub will vary by which of the three approaches is used and the product supporting that approach.

With **synchronized MDM**, the hub propagates and synchronizes local master data changes between SOEs. With this approach, there is no global master data store or global SOR. The objective here is to maintain the consistency of local master data across SOEs. In addition to not providing a global SOR, this approach also suffers from the problem of the hub having to deal with conflicting business rules across SOEs.

Repository MDM extends synchronized MDM by adding a global master data store, supporting a global SOR, and providing services that enable SOEs to directly access and update the global SOR. As discussed throughout this report, this architecture represents the main goal of enterprise MDM – a global SOR.

Hybrid MDM employs a combination of synchronized and repository MDM where some SOEs may directly access and update the global SOR, while others rely on the MDM environment to synchronize the global SOR with local master data. This is the most flexible of the three operational MDM approaches and may be the appropriate one to use for most operational MDM projects.

All three operational MDM approaches can work in conjunction with a SORef to publish historic master data for reporting and analysis purposes.

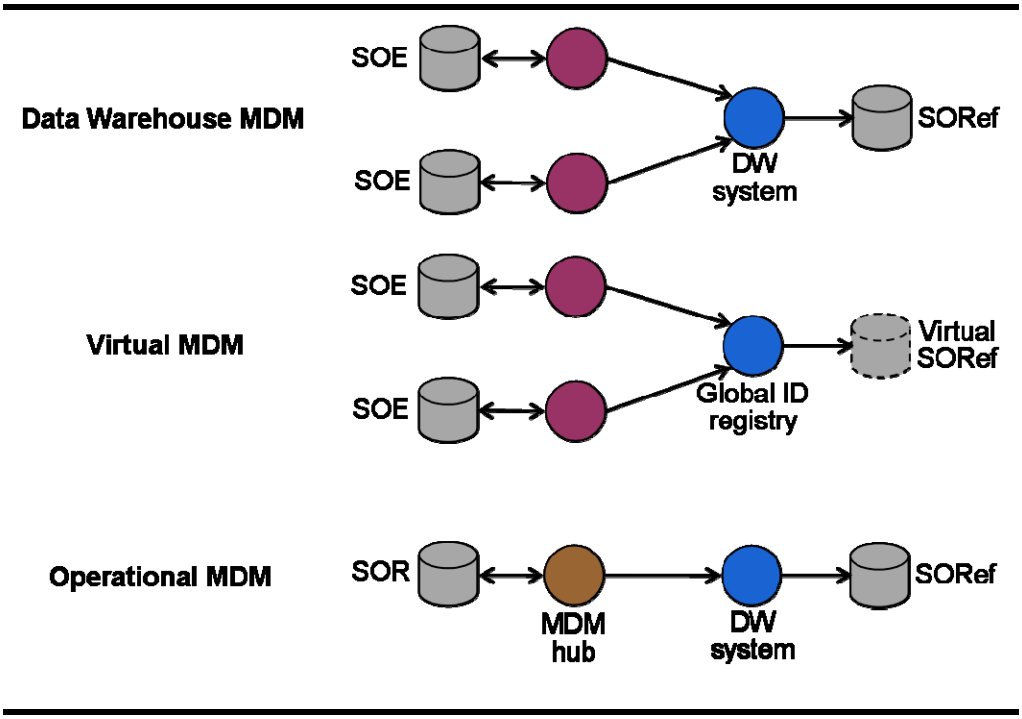


Figure 13: Approaches for Supporting Analytical MDM

ANALYTICAL MDM

Figure 13 illustrates the three main approaches for supporting analytical MDM.

Data warehouse MDM extracts local master data directly from SOEs, reconciles and transforms that master data and then loads the results into a data warehouse. This approach represents the traditional approach used by organizations to support analytical MDM. Its main disadvantage is that it attempts to retroactively solve some (but not all) operational master data problems, rather than solving those problems in the operational systems themselves.

Virtual MDM uses data federation technology to provide a virtual and consolidated view of local master data that is dispersed across multiple SOEs. The benefit of this style of MDM is that it provides a view of current master data. The main issue to consider, however, is the ability of the data federation technology to dynamically resolve master data inconsistencies when analytical applications use the virtual view to access multiple SOEs at runtime. This latter issue can be partially resolved by extending the virtual view with a global identifier registry (built by an identity management application) that maintains links related local master data in each of the SOEs. This is sometimes called **registry MDM**. It can be used by a data warehousing environment to capture master data from operational systems and load it into a data warehouse. The registry helps the data warehouse operational data extract and transformation programs capture related master data from multiple sources.

Operational MDM is simply an extension to one of the three styles of operational MDM that propagate master data from the MDM hub to a SORef, which in most cases is a data warehousing system. This obviously is the better of the three approaches for supporting analytical MDM.

Of the BeyeNETWORK MDM survey respondents who were planning, implementing or have implemented an MDM project, 12% were using synchronized MDM, 16% were using repository MDM, and 25% were using a hybrid of the synchronized and repository MDM styles. Some 33% of respondents were using data warehouse MDM and 11% were using virtual MDM.

SELECTING MDM PRODUCTS

When evaluating an MDM vendor product, it is important to consider the vertical master applications solutions it offers, its technical architecture, and the MDM horizontal features and services it provides. Its support for underlying data management services should also be considered.

Choice is often between a complete vendor MDM platform or a master data services solution

Some vendors package vertical master data application solutions and master data services into a complete MDM platform, while others may provide just a master data horizontal services component that enables third-party vendors and customers to develop their own vertical master data applications and services.

The master data application solutions provided by a vendor may consist of vertical master data applications, a starter set of master data templates and models, or a set of vertical master data services. Regardless of which of these options are offered, the solution should allow an organization to adapt and customize the solution to match business needs, and to evolve and integrate it into a full enterprise MDM environment. To avoid silos of master data, vertical master data applications should have a clean interface to underlying horizontal master data services, rather than bundle these services into individual MDM applications.

There should be open interfaces between master data applications, master data services and underlying data management services

From a technical perspective, the MDM environment should have an open and web services-based master data hub and global master data store. The hub should support both master data propagation and federation, and provide easy interfaces for importing and exporting master data and metadata into the global master data store. Where possible, it should interoperate with, and exploit, the capabilities of the existing data integration and data quality management environment.

Horizontal master data features and services in products vary considerably. Some of the keys ones that a product should provide include:

- A collaborative workbench for master data analysis and modeling, and browsing and editing
- Automatic versioning
- Hierarchy management
- Data record match, merge and reconciliation
- Reporting facility with data lineage and where-used reports
- Role-based security and granular access controls
- Auditing and traceability
- Workflow facility with rules management and notification

BUILD OR BUY?

To date, most MDM applications have been custom-built. Whereas these in-house developed applications may support virtual MDM or data warehousing MDM, they rarely provide full enterprise MDM. They are instead focused primarily on supporting master data integration, rather than master data management.

Vendor solutions usually provide better support for enterprise MDM

Like any *build versus buy* decision, the trade-off is between the cost of in-house development and maintenance versus the cost of vendor licensing and maintenance fees. There is no question, however, that most MDM vendor solutions provide additional capabilities compared with custom-built solutions, especially for enterprise MDM. They are also usually better integrated with packaged operational applications.

Avoid vendor monolithic MDM solutions supporting a single business entity and bundled master data services

Vendor MDM solutions are often specific to a business entity (customer entity, product entity, etc.) and/or an industry vertical (retail, banking, telecommunications, healthcare, etc.). For certain business entities and vertical industries, this can be an advantage because the built-in business models and templates usually provided enable a quick-start to MDM application development. Customer MDM is a good example here. In other areas, such as products, complexity and lack of industry standardization make the use of predefined templates and models less beneficial. When buying industry-focused solutions, organizations should be careful not to create multiple MDM silos that can be caused by monolithic applications with poor technical architecture.

Of the BeyeNETWORK MDM survey respondents, 38% were using a vendor MDM product, 35% custom built their MDM environment, while 27% used a combination of vendor and in-house solutions.

THE ROLE OF AN OPERATIONAL DATA STORE IN MDM

One of the main goals of an enterprise MDM environment is to manage a global master data store of integrated and consistent operational master data. One question that often arises concerning the global master data store is how it is different from an operational data store (ODS) that is often a component of a data warehousing system.

A traditional ODS integrates transaction and master data

In a non-MDM environment, the role of an ODS is to provide an integrated store of transaction and master data that has been acquired from multiple operational source systems. Although, an ODS presents a consistent view of operational data, this view is rarely up to date. There is always a time lag, or latency, between operational systems being updated and those updates being propagated into the ODS. This latency may be a few seconds or many hours, depending on business needs and supporting technologies. An ODS can be used for operational reporting and as an operational data source for downstream systems such as an enterprise data warehouse.

In an enterprise MDM environment, the global master data store subsumes part of the ODS functionality by providing an integrated store of operational master data.

In enterprise MDM, the ODS integrates only transaction data

The ODS in an MDM environment could continue to provide its traditional role, but the integration of master data into the ODS is redundant. Instead, it is more appropriate for the ODS to provide an integrated store of transaction data as shown in Figure 14. An enterprise data warehouse can then obtain global master data from the global master data store and integrated transaction data from the ODS.

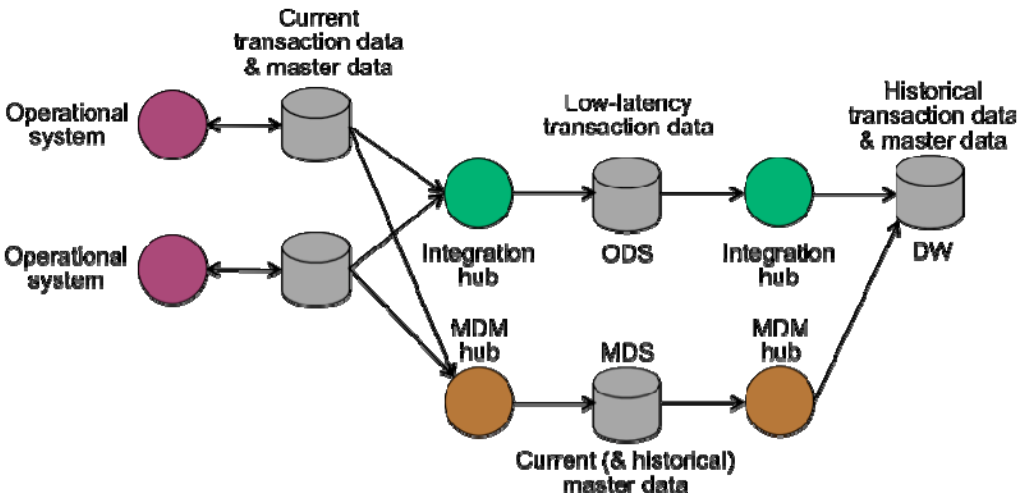


Figure 14: The Role of an ODS in an MDM Environment

GETTING STARTED ON AN MDM INITIATIVE

Enterprise MDM is a major initiative that must be implemented in small iterative steps

MDM is not a new concept: IT organizations have been building MDM solutions for several years. Many of these MDM initiatives, however, are a component of other projects such as building a data warehouse or improving data quality. Often these initiatives address a specific business need such as creating a single view of the customer or a single parts catalog. The issue here is that although these projects are faster and less costly to implement than full enterprise MDM, there is a risk of multiple master data silos being deployed in an organization. This is analogous to the data mart and enterprise data warehouse issues that occur in business intelligence.

Although most organizations realize that it is better in the long term to implement an enterprise data warehouse, they nevertheless build independent data marts because they are faster and cheaper to deploy. After building several data marts, companies realize they are creating data silos and then embark on an expensive project to consolidate their data marts. The same situation of data silos is beginning to occur with master data. The solution is to iteratively implement MDM using small projects, while at the same time having a strategic plan that allows the organization to move toward an enterprise MDM environment in a managed and cost-effective manner.

This strategic and evolutionary approach to enterprise MDM extends the focus and capabilities of more parochial MDM initiatives. It considers master data disciplines, policies and procedures, and not just products and technologies. It focuses on a hybrid enterprise MDM solution, rather than on stand-alone CDI, PIM and data warehouse approaches. It also emphasizes data governance, rather than just data

cleansing and reconciliation. Lastly, it requires organizational changes and change control management, rather than just executive sponsorship.

A long-term strategic master data plan is required

Enterprise MDM initiatives can, however, evolve from existing CDI, PIM, data quality and data warehousing projects provided there is a detailed long-term plan to achieve this.

FIRST STEPS

A master data governance and best practices function is essential for enterprise MDM success

How then do companies get started on an MDM initiative? The best approach for balancing the needs of short-term master data projects and the long-term strategic goal of enterprise MDM is to create a master data governance and best practices function that is responsible for managing the tactical and strategic master data goals of the organization. This group should develop a realistic master data plan that executives can support and that allows the organization to begin with small projects and progress incrementally to an enterprise MDM environment.

Specific MDM requirements that are important here include a well-documented inventory of master data and metadata sources, processes, producers and consumers; a strong master data governance and best practices function with political clout and a sound data integration and data quality infrastructure.

Enterprise MDM can evolve from existing SOA, data warehouse and data quality initiatives

There are several starting points for an MDM project. Some organizations begin by gradually upgrading operational applications to directly access the global master data store (the global SOR) of the MDM environment. As already mentioned, this may be done as a part of an SOA implementation. Other companies may evolve a data quality management project into a full MDM initiative. Many companies start MDM projects by extending a data warehousing and BI environment to support master data requirements such as a single view of the customer.

Ultimately, the strategic goal is to move toward an enterprise MDM environment. It may be the case that an organization never achieves full enterprise MDM because barriers such as older legacy applications prevent this. The purpose, nevertheless, is an integrated MDM environment that supports operational, collaborative and analytical MDM.

BEST PRACTICES

The people interviewed for this research suggested a number of practical, no-nonsense best practices to embrace before embarking upon an MDM initiative.

CREATE A VISION STATEMENT AND PLAN FOR ENTERPRISE MDM BUT ENSURE A SHORT-TERM DELIVERABLE

MDM is a long-term initiative that requires forethought and planning

MDM is a long-term initiative that requires forethought and planning. However, it is also true that to garner continuing support and funding, the team must implement interim deliverables demonstrating significant value to the enterprise. Advice from

the case studies was for the team to determine the overarching business reason(s) for this environment, and then develop a *big picture* road map while focusing on short-term projects.

They also suggested that the team do a thorough job of assessing and understanding four critical areas: people, process, technology and data. They recommended that each area be documented and procedures established for dealing with the issues that may come up in each area.

The last advice in this area dealt with the project scope. A business case for each project is needed. The business case should delineate the project's scope in terms of what is in and what is out of scope. The team must also be prepared to deal with scope creeps that will inevitably crop up.

CREATE A FORMAL MASTER DATA FUNCTION

While the first MDM implementation may not require a formal master data governance and best practices function, at some point it will be mandatory. Best practices indicate that the team should push for a formal function from the very first project. However, difficult questions will have to be answered quickly and the answers socialized. Who does the function report to? How will issues be resolved? Will it consist of dedicated staff? What authority, if any, will the function have in operations and data warehousing processing?

INCORPORATE BUSINESSPEOPLE INTO THE MASTER DATA FUNCTION

Business involvement is critical to the overall success of MDM

All the case study interviewees stated that businesspeople involvement was critical to the overall success of MDM. These business users included data stewards, data governance specialists, subject matter experts, etc. IT also has a critical role – it plays a supporting one by ensuring that the technological needs of the MDM environment are met. IT must also assure that master data can be easily accessed, transferred, updated, etc.

Other advice in this area was to establish incentives for businesspeople to support and promote MDM. These incentives should be a part of the employee's compensation plan and based on measurable events such as the reduction of master data interfaces, increased acceptance of a global master data store by the business, compliance with MDM procedures and so on. Of course, this best practice must assume another best practice – that there is active and vocal executive support (business and IT) for the initiative.

DETERMINE THE CONSUMERS AND PRODUCERS OF MASTER DATA

Document the workflows for consuming and producing master data systems to determine the interfaces needed

This best practice means the team must understand the full impact of their MDM environment on both the operational systems and the BI environment. How will MDM become integrated with operational systems and the BI environment? The MDM team must document the workflows for consuming and producing master data systems in order to determine the interfaces needed (and those that can be eliminated) as well as the ultimate master data SOR, SOE and SORef.

ESTABLISH REACTIVE AND PROACTIVE PROCESSES FOR MASTER DATA

There are two processes involved in an MDM environment – reactive and proactive. Reactive MDM occurs when the master data is extracted or propagated from operational system producers of local master data, cleaned up and integrated as global master data in the MDM environment.

Proactive MDM occurs when operational systems or data warehouse processes access (consume) master data directly from MDM environment. In some cases, the systems will use the master data from the MDM environment directly forgoing their own local storage. This may happen with operational systems directly accessing global customer and product master data. In others, the systems will copy the master data into their own stores. The data warehouse environment is a good example of when global master data is extracted from the MDM environment and reformatted into dimensions.

Most clients are still in the reactive phase of MDM

In researching the case studies for this report, we found that most clients were still in the reactive phase of MDM. However, several were considering the move to proactive MDM. In all likelihood, these two processes will co-exist in most enterprises. The tricky part will be to ensure that conflicting policies and procedures are completely ironed out.

MONITOR AND REPORT MDM BUSINESS PERFORMANCE IMPROVEMENTS (ROI, TCO)

All of our case studies mentioned the need to demonstrate real and significant value to the enterprise in order to sustain momentum and funding for MDM. The interviewees all stressed the need to determine meaningful and measurable metrics for each master data project. They also stated that these metrics must generate recurring returns rather than just a quick one-shot benefit at the end of the project. This means that the team may need to tie the MDM initiative to a larger initiative like CRM or to the strategic goals of company in order to maintain an ongoing return on investment.

CASE STUDY: FINANCIAL SERVICES ORGANIZATION

ORGANIZATION BACKGROUND

In these uncertain economic times, how comforting it is to read about a financial company committed to improving its information delivery capabilities as a core foundational step to operational efficiency and performance management. For this study, we looked at the strategic data management steps taken by an organization committed to improving these capabilities in support of its customer service goals.

THE BUSINESS PROBLEM

The business challenge facing the company today is their evolving market. Until recently, the company enjoyed a significant competitive advantage with the clientele they serviced, but this has shifted dramatically. Due to regulatory changes, many challengers have now entered their market space, causing it to be very competitive. While their investment portfolio has increased, their participant base has actually shrunk.

The executives in the company believed that the company's vast amount of data on their customers could give them a competitive advantage; but access to that data was not consistent, timely or as accurate as it could be. The company had overlapping architecture, projects and systems causing significant redundancy in data and processes for each office. They also had built a custom-code delivery system for customer data which was massive and uncoordinated. Overall, the business problem was to decouple the custom code system, develop a real customer data integration (CDI) solution and create an automated mechanism to deliver the data.

THE MASTER DATA MANAGEMENT SOLUTION

The data management team determined that the first step toward integrated customer data was to create an overarching enterprise architecture consisting of front, middle and back-office alignment, architectural layers and technology. The front-office systems are where the company's customers can perform inquiries on their investments and access their personal data.

The back-office systems manage the various investment instruments and keep track of each customer's investments. Though customers do not directly interact with the back office, this is where their books and records reside.

A middle office was needed to ensure the integration and accuracy of the customer data flowing between the front and back-office systems. The middle office would be where the customer's data was transformed into information and knowledge. It would be where transaction processing and customer profiling occur.

The team used this architecture to determine where the deficiencies were in the existing environment. After assessing the current situation, they discovered that each office had its own silo of front- and back-office processes for handling its customers' information and investments. They determined that there were more than 100 business processes in customer channels that all made changes to customer and customer account data.

The team needed to create an abstraction layer to cope with the complexities of the customer data and a process and delivery platform to deliver the data to applications. These became the components of the middle-office layer used to coordinate the customer data between the front- and back-office applications. The middle office would consist of a master data management (MDM) environment with an enterprise service bus (ESB) to deliver customer data to the front and back offices as well as for on-demand access and on to the enterprise data warehouse.

The company chose IBM's MDM Server for its CDI solution to create this service layer abstraction or middle office to ensure proper delivery of the customer data to appropriate applications throughout the enterprise. The first step toward this architecture has been to create a CDI hub of read-only current customer data. In this current configuration, the back-office systems contain both the system of record and the system of entry for customer master data. The customer hub is updated daily and used for the front-office accesses and inquiries.

The back office also supplies customer data to the enterprise data warehouse. The enterprise data warehouse is responsible for maintaining the historical customer data. By 2009, the goal is to have the customer MDM repository be read *and* write – that is, customer data will be directly updated within the MDM repository. This repository will then be responsible for supplying customer data not only to the front- and back-office applications but also to the enterprise data warehouse.

IMPLEMENTATION ADVICE

The team's advice to others starting an MDM project was straightforward and insightful.

Their advice is to understand the current and future state set of capabilities required to deliver and support MDM, and they recommend four areas be assessed and understood:

- People – Determine the roles, skill sets and experiences of the team members. The organizational structure itself can have a dramatic impact on an MDM project. Is the organization willing to step up to the help with the data quality processing, definitional creation and work flow determinations?

- Process – Determine what processes are in place or need to be in place to support the governance function so important for MDM projects. These processes should also include the establishment of standards for master data, metrics to measure its progress and utilization throughout the enterprise, etc.

- Technology – Choose appropriate technologies to support the enterprise architecture developed for master data, integration and quality processing of the master data, its storage in a repository and technologies that ensure its access and availability to all consuming applications.

- Information – Create appropriate business and governance rules in collaboration with the business community. This information includes all the metadata for master data as well as the enterprise data model documenting the entities, attributes and relationships established for all data including master data.

Their second piece of advice is to build an actionable MDM road map that aligns the required capabilities with strategic business priorities and the business architecture. They recommend that you manage the maturity of each capability against the overall road map. Doing this helps to:

- Ensure that you plan for growth and development in a controlled fashion. It is easy to get out of control fast.

- Make sure you design and build for reuse of components. Support the reuse of ETL and data quality processes, data models, metadata, etc. It is also useful to create templates for reuse as well.

- Evolve the MDM strategy based on the organization's changing priorities. Change is inevitable so your MDM road map must accommodate change with minimal perturbations to the existing environment.

- Plan (and architect) for future integration of additional data types. CDI is just one type of data for an overall MDM environment. The MDM repository should accommodate other forms of master data – product, account, geographic information, etc.

The final piece of advice is to watch the scope of your MDM projects. In this case, the team tried to put a lot of foundational pieces into place first and to go slowly. However, they faced big hiccups with customers, and the team took on more of the technology implementations than they should have. They did have the expertise or skill sets to ensure a smooth implementation. They recommended that MDM managers do an honest assessment of their team skills and determine if they would be better off outsourcing some of technology installation and implementation to skilled implementers.

CHALLENGES

The MDM team faced some challenges when they undertook the CDI project. The first became clear after their assessment of the existing environment. Understanding all the processes where customer data could enter the enterprise was no small feat. However, the team was able to document each of these processes and will eventually eliminate most of these as the MDM repository expands to include read and write

capabilities. In addition, the business rules for the creation, validation and maintenance of customer data were embedded in the back-office systems. These had to be deciphered and moved to the MDM environment.

A second challenge was ramping up the skill sets of their employees. At first, the team did not have a suitable methodology in place for MDM creation, nor did they have the requisite skills for a CDI project. Therefore, they found themselves overly reliant on outside consultants. Transfer of knowledge is a key aspect of their relationship with their vendors and acquiring the appropriate skills is occurring.

A final challenge comes from the executive sponsorship itself. The sponsoring executive left the company in the middle of the CDI project. The new management may have a different focus, which could result in a loss of traction for the CDI effort. The team is currently demonstrating the value of the MDM environment to the new executive to ensure ongoing support.

Overall, the challenges were greater than they anticipated. At a high level, their architecture was sound, but the devil is always in the details. Documenting the various systems was much harder and messier than expected. They also found that understanding and creating new data flows was difficult and remains problematic today.

BENEFITS

In terms of hard dollar ROI, the team estimated the savings generated from retiring several dozen customer stores as well as the removal of the APIs and interfaces for these stores. The breakeven point of 27 and 30 months is within current benchmarks.

Perhaps the bigger benefits are more intangible. The overall improvement in customer data quality has been significant, resulting in better customer service and more satisfied customers. They have succeeded in their stated goals:

- To deliver quality information to the front-office channels, reduce the number of customer complaints and improve the ability to drive more self-service through web.
- To be able to deliver more functionality with less time using in-house resources and partners.

Ultimately the CDI solution will enable a well-defined middle office to bridge the customer-centric front office to the back-office books and records. According to the Data Integration Services VP, the solution will:

- Insulate customers from operational/IT complexity and maximize data delivery efficiency to other consuming applications;
- Enable swift and timely processing, provisioning and presentation of customer account information;

- Orchestrate, integrate and execute customer requests and transactions;
- Share "services" across lines of business, as appropriate; and
- Build extensibility and flexibility to realign "services" with different business configurations.

FUTURE ISSUES

The next projects for the MDM team will ensure that the system of record and system of entry for all customer data migrates to the MDM repository. To do this, they will integrate the repository's interfaces with the actual systems of entry and eliminate potential conflicts with the front- and back-office customer sales order entry and sales order report capabilities.

They also have plans for a more proactive data quality service, not just in the ETL processing for the data warehouse. They plan to consume IBM's Quality Stage as a SOA-callable service for both the MDM and the data warehouse environment and reuse common data quality rules. Their goal is to reduce the time for data quality processes by making it more automated (versus the current manual processing).

SUMMARY

This company is well on its way to creating a first class MDM environment with the implementation of its CDI hub. The road map, architecture and technology support expansion into other master data areas while ensuring flexibility as the company's requirements shift.

By creating a cohesive customer repository, the team accomplished its goal of establishing consistent, reliable customer data for usage by both the front-office and back-office systems. Customers can readily access and update their information while the company can quickly manage their transactions and other requests. The CDI implementation will give the company the competitive advantage it was looking for.

IBM MASTER DATA MANAGEMENT SOLUTION

COMPANY BACKGROUND

International Business Machines was founded in 1896 as the Tabulating Machine Company and has evolved over the past 100 years to become a leading multinational computer technology and consulting company. It is headquartered in Armonk, New York.

IBM manufactures and sells computer hardware and software, and offers infrastructure services, hosting services and consulting services to a broad spectrum of organizations ranging from large enterprises and state and local governments to small and medium businesses. IBM's reported income in 2007 was $14.5 billion based on revenues of $98.8 billion.

The company's major operations cover four main areas: global services, global financing, systems and technology, and software.

IBM offers five main brands of software

The software group develops and sells operating systems, middleware and information management software. This software can be broken down into the following segments or brands:

- **WebSphere**: applications servers, process servers and transaction processing
- **Information Management**: business intelligence and performance management, information integration (including master data management), content management and data services
- **Lotus**: collaboration, social networking and portals
- **Tivoli**: systems management, storage and security
- **Rational**: architecture and construction, software life cycle management and quality management

IBM INFORMATION MANAGEMENT STRATEGY

IBM Information Management software supports the management, integration and analysis of business data and content

IBM's information management strategy is to "optimize enterprise performance by unlocking the business value of information," states Jim Caldwell, IBM Director of Software Group Technical Strategy. The software that supports this strategy is illustrated in Figure 1.

The DB2, Informix and FileNet software layer, shown at the bottom of the figure, provides a foundation for managing business information over its life cycle. The Cognos software layer, at the top of the figure, enables business users to leverage

information to better understand and optimize business performance. Sandwiched between these two layers is InfoSphere software, which integrates DB2, Informix, FileNet and other information sources to establish accurate and trusted information for use by Cognos software, third-party products and customer applications.

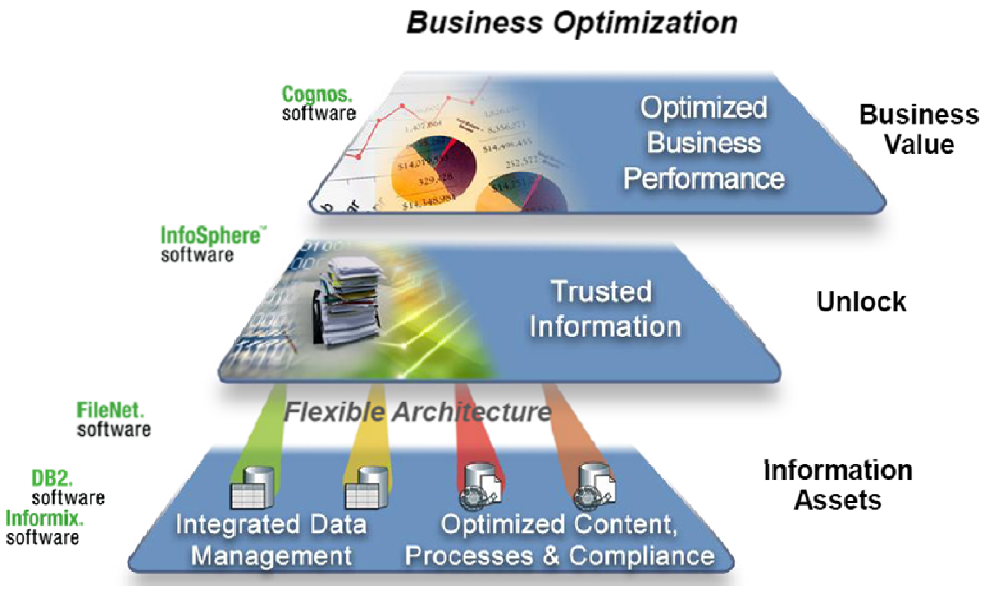


Figure 1: IBM information Management Software

InfoSphere provides products for managing accurate and trusted information including master data

InfoSphere consists of four main product families:

- **IBM Information Server** is a platform for accessing, integrating and governing heterogeneous data and metadata.
- **IBM InfoSphere Warehouse** includes a set of offerings that support the creation and maintenance of an enterprise data warehousing environment for both large and small enterprises.
- **IBM Industry Models** offer a collection of industry specific solutions that deliver prebuilt enterprise data, process and integration models for accelerating enterprise IT initiatives.
- **IBM InfoSphere MDM** provides products for managing master data business entities.

IBM INFOSPHERE MDM SOLUTIONS

HISTORY AND POSITIONING

IBM supports operational, collaborative and analytical MDM

IBM has invested heavily in master data management (MDM) over the past several years by acquiring products, and enhancing and integrating them with other InfoSphere software.

"IBM's MDM strategy is to provide a set of solutions that support operational, collaborative and analytical master data management for multiple business entities, including party, product, account and location," said John Gairhan, IBM's WW Manager in InfoSphere MDM Product Marketing. "IBM defines operational MDM as enabling real-time line-of-business (LOB) application access to master data. Collaborative MDM provides MDM-specific solutions for authoring and managing master data, while analytical MDM is intended for master data that will be used in an analytical processing environment, such as a data warehouse."

Two major MDM products

IBM offers two MDM products:

- IBM InfoSphere Master Data Management Server (IBM InfoSphere MDM Server)
- IBM InfoSphere Master Data Management Server for Product Information Management (IBM InfoSphere MDM Server for PIM).

IBM INFOSPHERE MDM SERVER OVERVIEW

InfoSphere MDM Server provides a master data hub for operational MDM

The **IBM InfoSphere MDM Server** manages and maintains master data for multiple business entities including party, customer, account, location, contract and product. The product has evolved from earlier customer data integration (CDI) offerings, namely DWL (a company acquired by IBM in 2005) and IBM WebSphere Customer Center. It is particularly well suited to helping LOB applications gain operational real-time access to trusted customer information for CDI projects, while at the same time supporting real-time access to product and account data for other initiatives.

Comes with more than 800 prebuilt business services

The product uses a services-oriented hub architecture for maintaining and propagating master data across systems. It comes with a library of more than 800 prebuilt and customizable Java business services. These services cover subject areas such as party demographics, roles, party relationships, events and customer insights, customer service and sales, locations, party identification, contracts, accounts, products, data stewardship, and history and audits.

A model-driven MDM Workbench is provided for customizing the IBM InfoSphere MDM Server and for building new MDM applications that employ MDM Server services. For collaborative MDM processing, the MDM Workbench also contains modeling tools for generating user interfaces for creating and maintaining master data.

Integrated with the IBM Information Server

The IBM InfoSphere MDM Server can be employed in conjunction with the IBM Information Server and other third-party products for additional data quality services. The IBM Information Server is also used to provide a *rapid deployment package* for the initial loading of existing master data into the MDM system. "The IBM rapid deployment package provides the technologies and services for the initial phases of an MDM project to help accelerate MDM implementation and provide faster time to value," said John Gairhan.

IBM MDM Server for PIM Overview

InfoSphere MDM Server for PIM is a collaborative MDM solution for product data

The **IBM InfoSphere MDM Server for Product Information Management (PIM)** provides a platform for creating, integrating, and managing product and supplier master data over the complete life cycle of a product. It supports a collaborative approach for managing product master data through the use of a flexible data model, users' interfaces and workflows.

The product imports and consolidates product information from ERP systems and other internal and external (D&B, for example) data sources. It enables business users to add additional information to the consolidated master data for key business processes such as new product introduction for web and brick-and-mortar stores, adding new suppliers, price and promotion management, and global data synchronization with trading partners. Consistent and up-to-date product information can then be exported to customer-facing systems such as web store fronts, kiosks and business portals, and also to external suppliers and vendors.

Uses workflows for authoring product master data

Key technology features include an extensible data model, master data versioning, incremental data import and export, field level security, workflows for collaborative tasks and a browser-based user interface for master data authoring and search. The product can also import and export business definitions from the business glossary component of the IBM Information Server. The Information Server also provides a MDM for PIM Loader that accelerates the importing of data into IBM InfoSphere MDM Server for PIM.

Summary

InfoSphere MDM Server is the main IBM product for enterprise-wide master data

The IBM InfoSphere MDM Server provides organizations with an operational MDM solution for consolidating and propagating customer-related master data across operational business applications. The recent addition of support for product and account master data indicates that this offering is evolving to become IBM's key solution for horizontal MDM of multiple business entities. Other recently added features also demonstrate the direction of the product to supporting collaborative MDM and the direct authoring of master data.

InfoSphere MDM Server for PIM is a more specialized solution for product master data

For more specialized PIM applications, IBM offers the IBM InfoSphere MDM Server for PIM. This product provides a collaborative MDM environment for consolidating and enhancing upstream product master data during the complete life cycle of products. This consolidated master data can then be exported to downstream

applications as required.

Both products can deliver master data to a data warehouse

Both products are gradually being integrated with other InfoSphere products, specifically the IBM Information Server, for data integration, data quality management, data loading, and business and technical metadata integration. Master data from both products can be exported to a data warehousing environment for the analytical processing of master data.

AUTHOR INFORMATION

CLAUDIA IMHOFF

Claudia Imhoff, Ph.D., is the president and founder of Intelligent Solutions, a leading consultancy on data warehousing and business intelligence technologies and strategies. She is a popular speaker and internationally recognized expert, and serves as an advisor to many corporations, universities and leading technology companies on these topics. She has co-authored five books and more than 100 articles on these topics and has a popular blog on the BeyeNETWORK at www.b-eye-network.com/blogs/imhoff/.

Claudia Imhoff
President
Intelligent Solutions
P. O. Box 4577
Boulder, CO 80306

Office: (303) 444-6650
Email: cimhoff@intelsols.com
Expert Channel: http://www.BeyeNETWORK/imhoff

COLIN WHITE

Colin is the Founder of BI Research. He is well known for his in-depth knowledge of leading-edge business intelligence and business integration technologies, and how they can be used to build a smart and agile business. With more than 35 years of IT experience, he has consulted for dozens of companies throughout the world and is a frequent speaker at leading IT events. Colin has written numerous articles on business intelligence and enterprise business integration. Colin has a popular blog on the BeyeNETWORK at www.b-eye-network.com/blogs/white/.

Colin White
President
BI Research
204 Logan Drive
Ashland, OR 97520-6600

Office: (541) 552-9126
Email: cwhite@bi-research.com
Expert Channel: http://www.BeyeNETWORK/white