

TDWI RESEARCH

TDWI BEST PRACTICES REPORT

SECOND QUARTER 2012

NEXT GENERATION MASTER DATA MANAGEMENT

By Philip Russom

CO-SPONSORED BY

talend*
*open integration solutions

tdwi.org

tdwi

NEXT GENERATION MASTER DATA MANAGEMENT

By Philip Russom

Table of Contents

Research Methodology and Demographics	3
Executive Summary	4
Introduction to Next Generation Master Data Management	5
Defining Master Data Management5
Defining Generations of Master Data Management7
Why Care about Next Generation Master Data Management Now?8
The State of Next Generation Master Data Management	8
Status and Scope of MDM Projects8
Priorities for Next Generation Master Data Management	10
Challenges to Next Generation Master Data Management	11
Best Practices in Next Generation MDM	13
Business Entities and Data Domains for MDM	13
Multi-Data-Domain MDM	15
Bidirectional MDM Architecture	17
Users' MDM Tool Portfolios	19
Replacing MDM Platforms	20
Quantifying MDM Generations.	22
Potential Growth versus Commitment for MDM Options	22
Trends for Master Data Management Options	25
Vendors and Products for Master Data Management	28
Top 10 Priorities for Next Generation MDM.	30



About the Author

PHILIP RUSSOM is a well-known figure in data warehousing and business intelligence, having published over 500 research reports, magazine articles, opinion columns, speeches, Webinars, and more. Today, he's the director of TDWI Research for data management at The Data Warehousing Institute (TDWI), where he oversees many of the company's research-oriented publications, services, and events. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and BI consultant and was a contributing editor with leading IT magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at prussom@tdwi.org, [@prussom](https://twitter.com/prussom) on Twitter, and on LinkedIn at [linkedin.com/in/philiprussom](https://www.linkedin.com/in/philiprussom).

About TDWI

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive Web site, tdwi.org.

About the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new business intelligence technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies and is supplemented by surveys of business intelligence professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving business intelligence problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. Please contact TDWI Research Director Philip Russom (prussom@tdwi.org) to suggest a topic that meets these requirements.

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who responded to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's production team: Jennifer Agee, Candice Christensen, Michael Boyda, and Denelle Hanlon.

Sponsors

DataFlux, IBM, Oracle, SAP, and Talend sponsored the research for this report.

Research Methodology and Demographics

Report Scope. User practitioners of master data management (MDM) face two broad challenges. First, MDM is inherently a multiphase, long-term project, and many organizations need guidance deciding what practices, architectures, and technologies their next generation of MDM should embrace. Second, vendor tools for MDM have added numerous new technologies and functions in recent years, and these offer compelling advantages over the homegrown solutions typical of early-phase MDM programs. The purpose of this report is to accelerate users' understanding of the many new user best practices and vendor tool functions that have emerged for MDM in recent years. It also helps readers map their options to real-world use cases and generations of MDM projects.

Survey Methodology. In November 2011, TDWI sent an invitation via e-mail to the data management professionals in its database, asking them to complete an Internet-based survey. The invitation was also distributed via Web sites, newsletters, and publications from TDWI and other firms. The survey drew responses from 520 survey respondents. From these, we excluded incomplete responses and respondents who identified themselves as academics or vendor employees. The resulting completed responses of 369 respondents form the core data sample for this report.

Survey Demographics. The majority of survey respondents are corporate IT professionals (70%), whereas the others are consultants (20%) and business sponsors or users (10%). We asked consultants to fill out the survey with a recent client in mind.

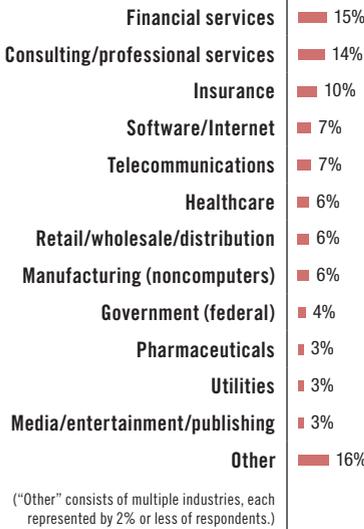
The financial services (15%) and consulting (14%) industries dominate the respondent population, followed by insurance (10%), software (7%), telecommunications (7%), and other industries. Most survey respondents reside in the U.S. (51%) or Europe (22%). Respondents are fairly evenly distributed across all sizes of companies and other organizations.

Other Research Methods. In addition to the survey, TDWI Research conducted many telephone interviews with technical users, business sponsors, and recognized data management experts. TDWI also received product briefings from vendors that offer products and services related to the best practices under discussion.

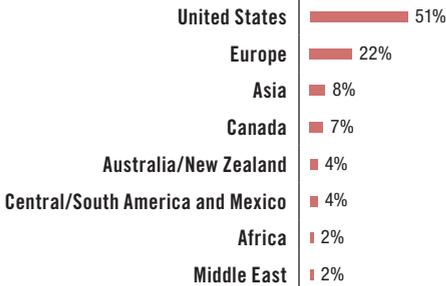
Position



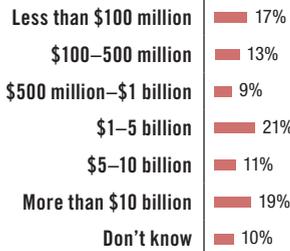
Industry



Geography



Company Size by Revenue



Based on 369 survey respondents.

Executive Summary

Master data management (MDM) is one of the most widely adopted data management disciplines of recent years. That's because the consensus-driven definitions of business entities and the consistent application of them across an enterprise are critical success factors for important cross-functional business activities, such as business intelligence (BI), complete views of customers, operational excellence, supply chain optimization, regulatory reporting, compliance, mergers and acquisitions, and treating data as an enterprise asset. Due to these compelling business reasons, many organizations have deployed their first or second generation of MDM solutions. The current challenge is to move on to the next generation.

Basic versus advanced MDM functions and architectures draw generational lines that users must now cross.

For example, some MDM programs focus on the customer data domain, and they need to move on to other domains, like products, financials, partners, employees, and locations. MDM for a single application (such as enterprise resource planning [ERP] or BI) is a safe and effective start, but the point of MDM is to share common definitions and reference data across multiple, diverse applications. Most MDM hubs support basic functions for the offline aggregation and standardization of reference data, whereas they should also support advanced functions for identity resolution, two-way data synchronization, real-time operation, and approval workflows for newly created master data. In parallel to these generational shifts in users' practices, vendor products are evolving to support advanced MDM functions, multi-domain MDM applications, and collaborative governance environments.

Users invest in MDM to create complete views of business entities and to share data enterprisewide.

According to survey respondents, the top reasons for implementing an MDM solution are to enable complete views of key business entities (customers, products, employees, etc.) and to share data broadly but consistently across an enterprise. Other reasons concern the enhancement of BI, operational excellence, and compliance. Respondents also report that MDM is unlikely to succeed without strong sponsorship and governance, and MDM solutions need to scale up and to cope with data quality (DQ) issues, if they are to succeed over time.

"Customer" is, by far, the entity most often defined via MDM. This prominence makes sense, because conventional wisdom says that any effort to better understand or serve customers has some kind of business return that makes the effort worthwhile. Other common MDM entities are (in survey priority order) products, partners, locations, employees, and financials.

Users continue to mature their MDM solutions by moving to the next generation.

MDM maturity is good, in that 61% of organizations surveyed have already deployed MDM solutions, and over one-third practice multi-data-domain MDM today. On the downside, most MDM solutions today are totally or partially homegrown and/or hand coded. But on the upside, homegrown approaches will drop from 45% today to 5% within three years, while dedicated MDM application or tool usage will jump from 12% today to 47%. To achieve generational change, half of organizations anticipate replacing their current MDM platform(s) within five years.

The usage of most MDM features and functions will grow in MDM's next generation.

Over the next three years, we can expect the strongest growth among MDM features and functions for real-time, collaboration, data synchronization, tool use, and multistructured data. Good growth is also coming with MDM functions for workflow, analytics, federation, repositories, and event processing. Some MDM options will experience limited growth, because they are saturated (services, governance, quality) or outdated (batch processing and homegrown solutions).

This report helps user organizations understand all that MDM now offers, so they can successfully modernize and build up their best practices in master data management. To that end, it catalogs and discusses new user practices and technical functions for MDM, and it uses survey data to predict which MDM functions will grow most versus those that will decline—all to bring readers up to date so they can make informed decisions about the next generation of their MDM solutions.

Introduction to Next Generation Master Data Management

Defining Master Data Management

To get us all on the same page, let's start with a basic definition of MDM and then drill into details:

Master data management (MDM) is the practice of defining and maintaining consistent definitions of business entities (e.g., customer or product) and data about them across multiple IT systems and possibly beyond the enterprise to partnering businesses. MDM gets its name from the master and/or reference data through which consensus-driven entity definitions are usually expressed. An MDM solution provides shared and governed access to the uniquely identified entities of master data assets, so those enterprise assets can be applied broadly and consistently across an organization.

This definition is a good nutshell of what MDM is. However, to explain in detail what MDM does, we need to look at the three core activities of MDM—business goals, collaborative processes, and technical solutions. (These activities are summarized in Figure 1.)

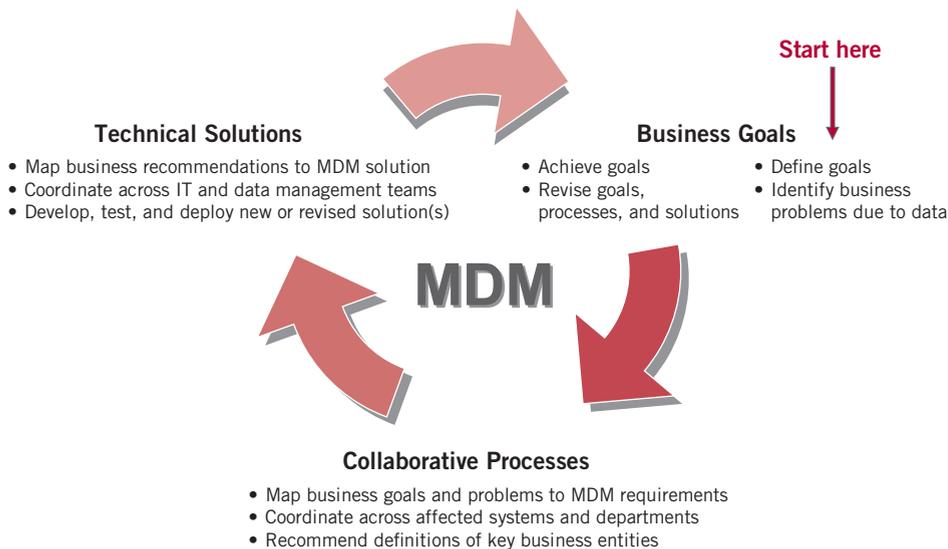


Figure 1. MDM's three core activities: business goals, collaborative processes, and technical solutions. (Source: TDWI)

Business Goals and MDM

Most organizations have business goals, such as retaining and growing customer accounts, optimizing a supply chain, managing employees, tracking finances accurately, or building and supporting quality products. Because most business goals focus on a business entity—such as a customer, supplier, employee, financial instrument, or product—MDM can help organizations achieve all these goals (and other data-driven goals) more easily and accurately. Some goals combine two or more entities, as in customer profitability (customers, products, and finances) or product quality (suppliers and products). MDM contributes to these goals by providing processes and solutions for assembling complete, clean, and consistent definitions of these entities and reference

Most business goals relate to business entities defined via MDM.

data about them. Many business goals span multiple departments, and MDM prepares data about business entities so it can be shared liberally across an enterprise.

Sometimes the business goal is to avoid business problems. As a case in point, consider that one of the most pragmatic applications of MDM is to prevent multiple computer records for a single business entity. For example, multiple departments of a corporation may each have a customer record for the same customer. Similarly, two merging firms end up with multiple records when they have customers in common.

Business problems ensue from redundant customer records. If the records are never synchronized or consolidated, the firm will never understand the complete relationship it has with that customer. Undesirable business outcomes include double billing and unwarranted sales attempts. From the view of a single department, the customer's commitment seems less than it really is, resulting in inappropriately low discounts or service levels. MDM alleviates these problems by providing collaborative processes and technical solutions that link equivalent records in multiple IT systems, so the redundant records can be synchronized or consolidated. Deduplicating redundant records is a specific use case within a broader business goal of MDM, one that is focused on providing complete and consistent data (especially views of specific business entities) across multiple departments of a larger enterprise and thereby enabling or improving cross-functional business processes.

Collaborative Processes for MDM

Consensus-driven entity definitions demand cross-functional collaboration.

By definition, MDM is a collaborative discipline that requires a lot of communication and coordination among several types of people. This collaboration is especially true of entity definitions, because one person rarely knows all the details that would go into a standard definition of a customer or other entity. The situation is compounded when multiple definitions of an entity are required to make reference data “fit for purpose” across multiple IT systems, lines of business, and geographies. For example, sales, customer service, and finance all interact with customers, but have different priorities that should be reflected in a comprehensive entity model. Likewise, technical exigencies of the multiple IT systems sharing data may need to be addressed in the model. Also, many entities are complex hierarchies or have dependencies that take several people to sort out, as in a bill of materials (for products) or a chart of accounts (for financials).

Once a definition is created from a business viewpoint, further collaboration is needed to gain review and approval before the definition can be applied to IT systems. At some point, business and technical people come together to decide how best to translate the definition into the technical media through which a definition is expressed. Furthermore, technical people working on disparate systems must collaborate to develop the data standards needed for the exchange and synchronization of reference data across systems. Because applying MDM definitions often requires that changes be made to IT systems, managing those changes demands even more collaboration.

That's a lot of collaboration! To organize the collaboration, many firms put together an organizational structure where all interested parties can come together and communicate according to a well-defined business process. For this purpose, data governance committees or boards have become popular, although stewardship programs and competency centers may also provide a collaborative process for MDM and other data management disciplines (especially DQ).

Technical Solutions for MDM

An implementation of MDM can be complex, because reference data needs a lot of attention, as most data sets do. MDM solutions resemble data integration (DI) solutions (and are regularly mistaken for them), in that MDM extracts reference data from source systems, transforms it to normalized models that comply with internal MDM standards, and aggregates it into a master database where both technical and business people can profile it to reveal duplicates and noncompliant records.

Depending on the architecture of an MDM solution, this database may also serve as an enterprise repository or system of record for so-called golden records and other persistent reference records. If the MDM solution supports a closed loop, records that are improved in the repository are synchronized back to the source systems where they originated. Reference data may also be outputted to downstream systems, such as data warehouses or marketing campaign systems.

MDM solutions also resemble data quality solutions, in that many DQ functions are applied to reference data. For example, “customer” is the business entity most often represented in reference data. Customer data is notorious for DQ problems that demand remediation, and customer reference data is almost as problematic. We’ve already mentioned deduplication and standardization. Other DQ functions are also applied to customer reference data (and sometimes other entities, too), including verification and data append. Luckily, most tools for MDM (and related disciplines such as DI and DQ) can automate the detection and correction of anomalies in reference data. Development of this automation often entails the creation and maintenance of numerous “business rules,” which can be applied automatically by the software, once deployed.

MDM solutions resemble those for DI or DQ, but applied to reference data and master data sets.

Defining Generations of Master Data Management

Like most IT disciplines, MDM is inherently a long-term, multigenerational discipline that travels through several life cycle stages on its way to maturity. The generations of MDM are most apparent in its technical solutions, which tend to mature into more numerous and more advanced technical capabilities. However, MDM also depends on collaborative processes, which likewise mature by incorporating more people, departments, and business processes. In addition to MDM’s natural maturation, new generations of MDM are also driven by technical advancements in vendor software tools and emerging and evolving business requirements in organizations.

All the above are driving the majority of MDM programs and solutions toward a redesign, major retrofit, or even replacement that we can recognize as a *generation*. TDWI takes the term literally, meaning that an organization’s current generation of MDM will beget its next generation.

What’s next for a given MDM solution can vary tremendously. For example, a next generation solution may tap into leading-edge features, such as identity resolution and event processing. It may simply get you caught up with somewhat more established practices for real-time operation and data services. Sometimes, the next generation addresses platform issues, such as solution architecture and tool types. Even so, most MDM generations are incremental and add more data domains, departments, data management tools, and operational applications into their purview. So, let’s keep in mind that a next generation MDM solution is a relative concept, because it depends on where you’re starting, what new requirements you must address, and how many resources you have.

MDM is inherently multigenerational, so plan accordingly.

The content of your next generation depends on your current situation and future plans.

Update or replace MDM solutions to keep pace with data-driven business goals.

Why Care about Next Generation Master Data Management Now?

Businesses need MDM more than ever. That’s because modern businesses increasingly achieve many of their goals by leveraging data, and MDM helps achieve data-driven business goals with greater ease, accuracy, and breadth, as described earlier.

Most MDM solutions are in serious need of improvement or replacement. Most MDM solutions originate as departmental silos with an intentionally narrow focus, which makes sense as a controllable startup project. But it also means that the first generation (which is often little more than a homegrown, hand-coded prototype) demands considerable rework to enable the usual growth into more business entities, data domains, departments, geographies, sources, and targets. A complete replacement of the original MDM platform may be necessary for advanced MDM capabilities, such as real-time operation and data services. All this requires planning if an MDM program is to keep pace with evolving and expanding business requirements.

There’s probably a new generation in your near future. If you can find the right starting point for MDM that proves its worth, then growing into new generations is highly probable. No matter what life cycle stage a particular MDM solution is in, there is always more it could do to serve departments and the whole enterprise. Given the long list of common generational changes for MDM and MDM’s many business drivers, growth and change are inevitable for successful MDM solutions and programs. This report can help you plan those new generations.

USER STORY

A TRULY COMPLETE CUSTOMER VIEW MAY REQUIRE CONSOLIDATING MDM SOLUTIONS

“A few years ago, the company needed to pursue a 360-degree view of each customer, largely to enable cross-selling a single customer into multiple financial and insurance products,” said Joe Royer, an enterprise architect at financial services firm Principal Financial Group. “Plus, we needed to better institutionalize our policies around customer interactions. The catch was that multiple business units already had their own customer views in place. Although these were good starting points, each view was limited to the priorities of its business unit.

“In pursuit of the business goals, we consolidated preexisting MDM point solutions into a central master repository for customer data. And we enriched customer reference data to include customer policy information, as well as all attributes needed for a truly complete view. Now that we have enterprise-scope customer views, we’ll incorporate more data domains, starting with other parties, namely employees and partnering companies. Then we’ll move on to products, such as financial instruments and insurance policies.”

The State of Next Generation Master Data Management

Status and Scope of MDM Projects

Discussions of MDM and related matters make one wonder whether organizations are actually practicing MDM. To quantify the situation, TDWI’s MDM survey asked: “Which of the following best describes the status and scope of your organization’s MDM practice?” (See Figure 2.)

MDM is common today, but not as common as it should be.

MDM solutions are fairly common today. The majority of surveyed organizations (61%) practice some form of MDM today, regardless of the method, tool type, or generational phase. Although this presence is respectable for MDM, it’s rather modest given that MDM has been a hot topic in the IT press for 15 years and that users are famous for exploring its potential use.

Which of the following best describes the status and scope of your organization's MDM practice?

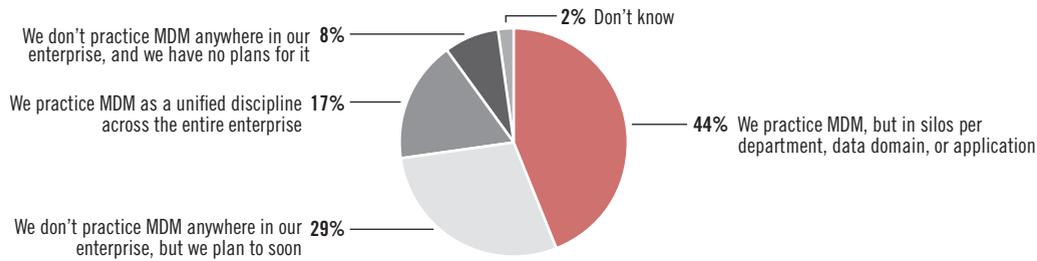


Figure 2. Based on 369 respondents.

Many organizations would like to practice MDM but haven't yet (29%). This report's survey aside, other TDWI surveys have consistently shown a substantial percentage of user organizations stuck in an MDM exploration phase (42% on average; not charted here). Many users see the value of MDM but still can't attain organizational commitment. As we'll see later, the first generation of an MDM program can be stymied by a lack of sponsorship, business case, governance, cross-functional cooperation, MDM expertise, and so on. With so much stacked against it, the first generation of MDM is probably the most critical; without it other generations are not possible.

Most MDM solutions deployed today are silos (44%). Finding just the right starting point for MDM is critical to getting it off the ground. A tried-and-true way to start an MDM program is to narrow its focus to the MDM requirements of a specific department, data domain, or application. For example, first-generation MDM is commonly focused on sales and marketing's need for the customer data domain, to optimize their use of customer relationship management (CRM) and sales force automation (SFA) applications. Data for reports and analytics in a data warehouse is a similarly focused example.

MDM solutions in silos are common. Those that span an enterprise are not.

These siloed starting points tend to succeed, because they have obvious sponsors (the department head) and business cases (the needs of the department), plus they avoid potential showstoppers like cross-functional cooperation and enterprise governance. Success is common when the first generation of an MDM solution has such a narrow focus, which explains why most MDM solutions deployed today are silos. However, the result is a silo with inherently limited visibility into a single data domain within a single department. It also leads to multiple silos that need consolidation or integration in a later generation, when enterprise-scope MDM is required.

Few organizations practice MDM with an enterprise scope (17%). TDWI has encountered user organizations that designed their MDM program to be enterprise scope from the beginning and they built a single enterprise-scope solution—but these organizations are few and far between. Most start with a few silos and then connect the silos later, suffering much rework in the process. To reduce the rework, some start with an enterprise plan and tactically build MDM silos that integrate with each other, usually via common data models and other standards for reference data.

Achieving enterprise reach in MDM is an inevitable goal for most organizations. After all, the sharing of consistent data across departmental and business unit boundaries is the long-term goal of any serious MDM implementation. However, enterprise-scope MDM is a multiphase project that typically takes a few generations to achieve.

Very few organizations have no plans for MDM (8%). This number is a new record low. In other TDWI surveys, “no plans for MDM” has ranged from 21% in 2006 to 13% in early 2011, with steady decreases along the way. The need for MDM has been growing slowly but steadily over the last 15 years, as reflected in the shrinking number of organizations that can ignore it.

Priorities for Next Generation Master Data Management

MDM solutions are common today, but that doesn’t explain why user organizations are deploying them. To get a sense of why users deploy MDM, TDWI’s MDM survey asked respondents to “rank, in priority order, the top five reasons for implementing MDM.” (See Figure 3.)

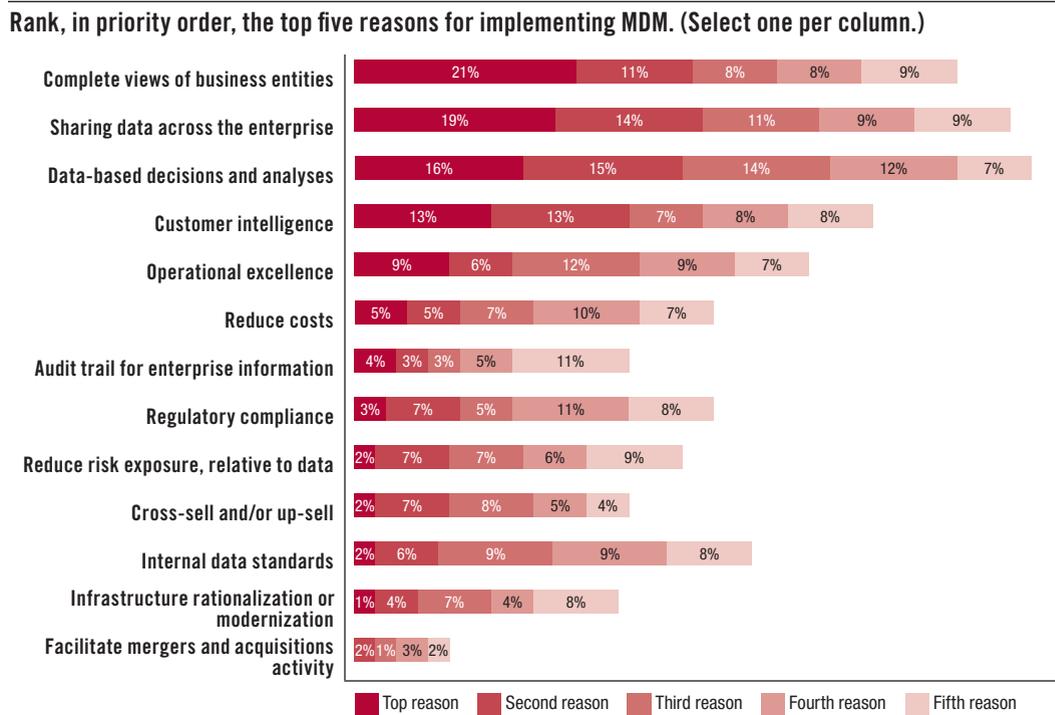


Figure 3. Based on 369 respondents. Results are sorted by the “Top Reason” column.

The top reasons for MDM are to enable complete views and to share enterprise data.

Gaining 360-degree views is the top priority for implementing MDM. More survey respondents chose “complete views of business entities” (21%) than any other reason for implementing MDM. This result makes sense, because a complete view of a customer helps an organization retain and grow that customer, plus provide the customer with better service—no matter which department the customer contacts. Likewise, a complete view of each business partner reveals which ones deliver the best-quality supplies, on time, at the lowest cost. It’s also difficult to efficiently manufacture a quality product on schedule without visibility into all product parts and subassemblies—visibility that MDM can help provide.

Sharing data across the enterprise (19%) is also a top reason for MDM. This priority relates to 360-degree views, because assembling a complete view isn’t possible without data drawn broadly from the IT systems of many business units. If these business units won’t share their data, then entity views are rarely complete. MDM is one way to share enterprise data more easily and apply it more accurately to purposes besides those of its origins. Most approaches to MDM enable organizations

to create consensus-driven definitions of the business entities that are involved in a view, which facilitates the accurate meshing of diverse data into an accurate view. Other MDM techniques, such as standardized reference data and shared data models, also make the assembly of complete views easier and more accurate.

BI is a high priority for MDM. Sixteen percent of survey respondents selected “data-based decisions and analyses” as their top reason for implementing MDM, while another 15% chose it as their second reason. On a related topic, “customer intelligence” ranked highly as both a top reason (13%) and a second reason (13%). BI and analytics rely heavily on views constructed from enterprise data drawn from many sources, which is why BI is one of the most common starting points for MDM programs.

Other leading reasons for implementing MDM include BI, operational excellence, and compliance.

Operations and their costs benefit from MDM. Consistent entity definitions and complete views of those entities can improve most data-driven operational processes, including those for financials, sales, asset management, supplier negotiations, and managing employees.

Compliance issues demand MDM. Noncompliance can lead to an audit, whether you’re audited by a government agency, a business partner, a legal team, or your own employees. Surviving an audit is all about providing credible, bulletproof information to auditors, and the audit trail that MDM provides is instrumental. Furthermore, the accurate use of data (as enabled by MDM) helps to reduce risk relative to data.

Challenges to Next Generation Master Data Management

User organizations tend to have trouble starting a MDM program, as well as progressing from one generation to the next. To get a sense of the most likely challenges, TDWI’s MDM survey asked respondents to “rank the top five challenges to MDM success.” (See Figure 4.)

Business issues are the most common challenges to MDM success. Programs for MDM are similar to those for data quality, in that both require changes to applications, data, and how business people use them. Organizations are more likely to succeed at convincing the owners of applications and data to make required changes to reference and master data if changes are mandated by executive sponsorship (14%). For a sponsor to be committed, he or she needs a compelling business driver or business case (8%). It helps if the sponsor can secure adequate funding (5%).

Sponsorship and governance are the most critical factors for MDM success.

Cooperation can make or break an MDM program. Given the changes that most MDM solutions require, as well as the goal of sharing data across enterprise boundaries, a fair amount of cross-functional cooperation (9%) is required among business managers. Otherwise, the program is hamstrung by turf wars over data ownership (3%). Likewise, the MDM team needs coordination with other data management disciplines (7%); without it, a realistic assessment of MDM technical needs and possible corrections is not possible. To enable and organize the substantial cooperation that MDM demands for success, many organizations are founding or expanding programs for data governance or stewardship (13%). Many IT and business people have told TDWI that MDM (and data quality, too) would not have started successfully, much less grown into further generations, without data governance or an equivalent collaborative organization.

Technical difficulties can drag down an MDM implementation. The current explosion of big data affects most IT systems, including MDM. Ten percent of survey respondents are concerned about scaling up to the growing numbers of reference data sources and targets. Other possible inhibitors to an MDM implementation include poor data quality (6%), integration with business process management (BPM; 6%), lack of MDM experience (5%), and inadequate integration infrastructure (4%).

To succeed, MDM solutions need to scale and to cope with data quality issues.

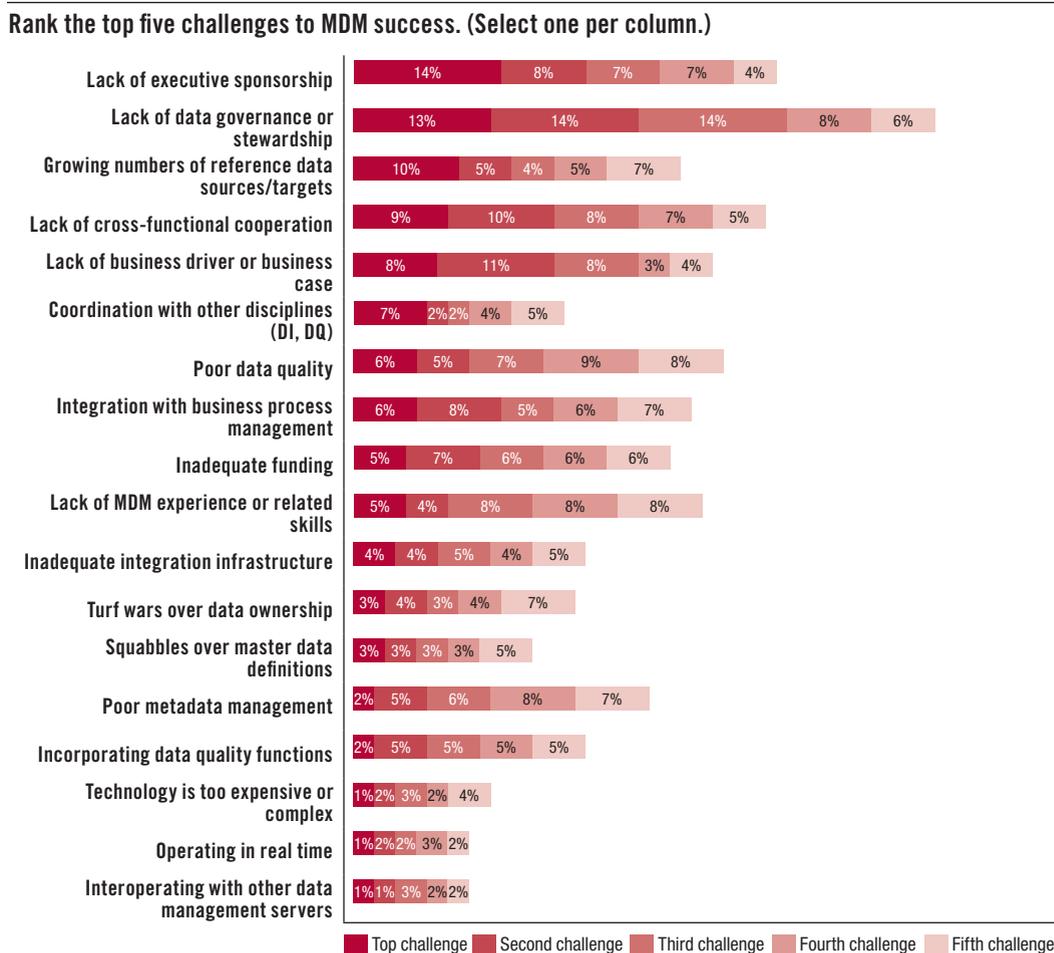


Figure 4. Based on 369 respondents. Results are sorted by the “Top Challenge” column.

USER STORY

DATA GOVERNANCE CAN BE A CRITICAL SUCCESS FACTOR FOR MDM

“I was hired to establish an enterprise data architecture program, which included a program for data governance and strategy,” said the manager of enterprise information architecture at a national retail chain. “I started by focusing on our master data domains, the priority being to identify data owners and stewards among business people. Each business data owner was tasked with selecting a data steward in their area and then identifying the top 10 data elements that are critical to their assigned master data domain, so we could begin to capture the business and technical metadata about each element. To get the business owners started, I had one of my data architects document each domain’s data elements in a spreadsheet, so the business owner would have a starting point and could choose the most critical from the list. That sped things up and helped the business people focus. Most owners had trouble getting their list down to 10, but at least they all knew which elements are critical. Data governance with a stewardship slant was a critical success factor for MDM, because it got the business people involved up front and helped them understand how their business goals are linked to master data elements.”

Best Practices in Next Generation MDM

Business Entities and Data Domains for MDM

Before proceeding, it’s time to clarify the connection between business entity and data domain in the context of MDM. A data domain is a collection of the unique instances of an entity, and the values of the instances are persisted in storage or instantiated, as needed. In other words, a definition of “the customer” as a business entity establishes a schema or data model for the data of specific instances of customers, and the resulting aggregated reference data constitutes the customer data domain.

In MDM solutions, a data domain is a data set that is typically managed as a reference table, repository, registry, or other database structure. Typically, many reference data records are managed in a single master data set per domain. Like any good database, a true data domain has many technical rules and business rules that enforce allowed data types, values, and relationships among them. A data domain’s master data set has a data model that serves as a standard for transformed and aggregated reference data; without this model, reference data standardization, verification, augmentation, and deduplication would not be possible.

An assumption of this discussion is that MDM solutions operate on the reference data domains that result from a focus on business entities. To understand which business entities (and their data domains) receive the most attention from MDM efforts, this report’s survey asked: “Which business entities do you model with master or reference data today?” (See Figure 5.) To ensure that answers to the question are based on real-world implementations, the survey presented the question only to respondents who had reported earlier in the survey that they have direct experience with an MDM implementation; they could select all answers that applied.

A data domain results from each of the business entities that MDM defines.

Which business entities do you model with master or reference data today? (Select all that apply.)



Figure 5. Based on 861 responses from 219 respondents who have MDM experience; 4 responses per respondent, on average.

Customers. As no surprise, the business entity most often modeled in MDM is the customer (77% in Figure 5). After all, many user organizations are in customer-focused industries, such as financial services, insurance, and retail. Instead of mass consumers, many industries are focused on business-

“Customer” is, by far, the entity most often defined via MDM, followed by “product.”

to-business relationships—as in manufacturing or wholesale—where the customer is also a corporation or other kind of enterprise. Other industries have a prominent equivalent to the customer, such as patients in healthcare (10%) and citizens in government (5%). Furthermore, conventional wisdom says that any effort to understand or serve customers better has some kind of business return that makes the effort worthwhile.

Note that few organizations surveyed have only one definition of customer (18% in Figure 6). The norm is closer to 5 definitions (42%) and sometimes 10 or more (25%). One of the consequences of using MDM to enable data sharing across departments is that some departments have unique requirements that demand a separate definition. Many organizations cope with this requirement by creating a broad customer definition that is a superset from which other definitions are drawn—which explains why reference data in the customer domain usually takes the form of a very wide record, sometimes with hundreds or thousands of fields.

Products (54% in Figure 5). The product data domain is obviously important in product-oriented industries that produce products (manufacturing) or distribute them (retail). However, a corporation in any industry (or a school or government agency) will acquire and consume products, perhaps through a procurement process. Even customer-oriented industries offer service-based products, such as policies in insurance (9%). Hence, the product data domain has broader applicability than you might think.

As with the customer data domain, few surveyed organizations have a single definition of product (17% in Figure 7). The norm is approximately 5 definitions (30%) and sometimes 10 or more (36%). Flat, record-based modeling works well with customer reference data. However, product reference data tends to require hierarchical models, because many products are comprised of multiple subassemblies and parts that themselves require definition. The average bill of material (often the subject of MDM efforts) is a case in point.

Approximately how many definitions of “customer” does your organization have?

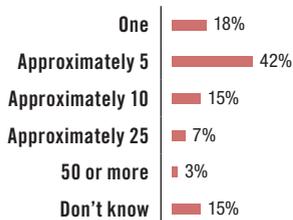


Figure 6. Based on 219 respondents who have MDM experience.

Approximately how many definitions of “product” does your organization have?

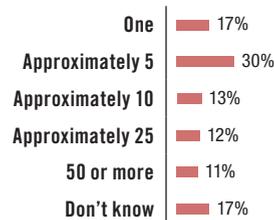


Figure 7. Based on 219 respondents who have MDM experience.

Other common MDM entities are partners, locations, employees, and financials.

Business partners (46% in Figure 5). Companies seeking to optimize their supply chain often apply MDM to data about suppliers, distributors, and logistic partners (in trucking, rail, parcel delivery, etc.).

Locations (44%). Location data is rarely standalone, and it is typically appended to reference data to enhance its content. For example, longitude and latitude coordinates make it easier to find and maintain assets in the field. These and other location identifiers can add attributes to customer data that make deliveries faster and more accurate, enhance customer base segmentation, and provide an accurate geography dimension for customer analytics.

Employees (40%). One reason to apply MDM techniques to the employee data domain is to manage and assign employees in a more agile manner, based on consistent and granular reference data. Some organizations need to support Office of Foreign Assets Control (OFAC) and other “watch lists,” as a matter of compliance and risk reduction. But the most common reason is simply to raise efficiencies in human resources (HR) in general, which may require MDM applied to HR data about benefits, tax deductions, and vacation time accrued.

Financials (39%). Surprisingly, financials ranked rather low in this survey. (It was second only to the customer domain in TDWI’s 2006 survey about MDM.) This data domain is useful on its own, to enable accurate tracking and insightful study of finances across an enterprise. But it’s also very useful when integrated with other MDM domains, to reveal the cost and profitability of customers, products, and partners. As with the product data domain, financials tend to require hierarchical reference data; the average chart of accounts is a case in point.

USER STORY

DON’T SEPARATE REFERENCE DATA FROM MASTER DATA.

“Some professional thinkers about data will show high-level models that segregate reference data from master data, or subsume the former into the latter. I think this is a mistake,” said John O. Biderman, an information architect at Harvard Pilgrim Health Care. “Reference data across an enterprise is as much in need of ‘mastering’ as is the core structural data, such as products or customers. Application systems tend to proliferate their own coding and classification schemes with varying but overlapping taxonomies.

“We felt this acutely in building our enterprise data warehouse, in which reference domains from multiple systems needed to be rationalized into an enterprise standard. This led us to develop a tool we call the Reference Table Utility, and a data store called the Corporate Reference Center, or CRC, and to engage business stakeholders in governance over reference data. The CRC is also the place where data analysts maintain coding schemes that translate or aggregate data into analytically useful terms, taking these reference sets off the end-user’s desktop and making them enterprise accessible. Now, as we build a new operational platform based on service-oriented architecture [SOA], the same processes and toolkits are being leveraged to essentially virtualize the master reference data hub.”

Multi-Data-Domain MDM¹

One of the most common—and challenging—generational changes seen in MDM is the shift from single-data-domain practices (in early generations) to multi-data-domain practices (in mid-life cycle generations). For example, in many organizations, the earliest efforts with reference data focus on one domain (usually the customer domain), expressed by a single, siloed MDM technical solution. As MDM practices in the same organization address other domains—typically financial or product data—they usually do so in separate solutions, with little or no integration with MDM for customer data. Some user organizations have multiple customer-focused MDM solutions, one each for marketing analytics, direct marketing, sales pipeline, customer service, and so on. So much for a single view of the customer! When these organizations have their hands full consolidating customer-data-domain MDM solutions, it delays the next generation, which implements a multi-data-domain MDM solution.

Multi-data-domain MDM on a single platform is technically feasible today. Hence, many of its barriers are organizational. As pointed out earlier, many MDM solutions originate to serve a specific sponsor, department, budget, or application, which is successful for early generations. A multi-domain approach tends to work against this sponsorship model. To make the multi-domain

Multi-data-domain support is a sign of a mature MDM solution.

¹ Some of the material for this section of the report comes from the TDWI blog “The State of Multi-Data-Domain MDM,” available on tdwi.org.

approach more palatable to individuals, central IT may need to provide an MDM platform as infrastructure shared by all business units and departments.

Single-data-domain MDM is myopic.

The problem with single-data-domain MDM is that it's a barrier to having common, consensus-based entity definitions and standard reference data that would allow you to correlate information across multiple domains. For example, single-domain MDM is great for creating a single view of customers. But it needs to federate or somehow integrate with MDM for the product-data domain, if you want to extend that view to include (with a high level of accuracy and consistency) products and services that each customer has acquired or considered. Or you might eventually need to include financial or location data into the single view of customers. Some day, you'll also include data from social media and other Web sources. All these goals are more accurate and easier to achieve with multi-data-domain MDM solutions, built with and integrated through a single tool or platform.

Hence, a common tipping point is when technical and business people have reached maturity with customer-data MDM, and they realize they can't get to the next level of customer intelligence without consistent and integrated MDM about other domains. Another way to put it is that the single view of the customer gets broader as it matures, thus demanding information from other domains. Yet another way to think of it is that multi-data-domain MDM often comes in a later generation, after single-data-domain MDM has proven the concept of MDM, in general. And much of the success of multi-data-domain MDM is not about technology—success depends on having a corporate culture that demands data sharing in support of cross-functional coordination.

To get a sense where the average organization is with this generational shift, this report's survey asked respondents who have had MDM experience: "Which of the following best describes your MDM solutions relative to data domains?" (See Figure 8.)

Which of the following best describes your MDM solutions relative to data domains?

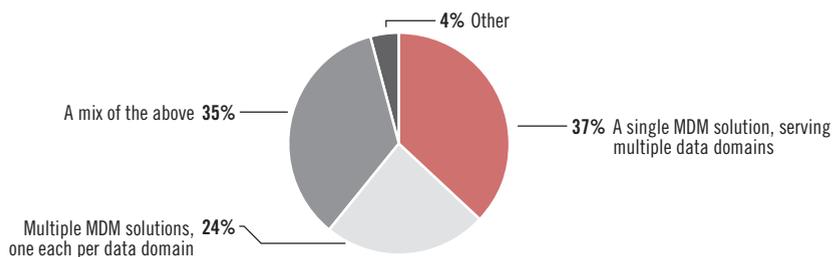


Figure 8. Based on 219 respondents who have MDM experience.

MDM is aggressively addressing multiple domains.

More MDM solutions have gone multi-domain than haven't. Despite its challenges, many user organizations have successfully negotiated the generational shift to a single MDM solution serving multiple data domains (37%), thereby proving that it can be done in a broad range of organizations. This is good news for the cause of multi-data-domain MDM and for MDM maturity, in general.

One-quarter still rely mostly on single-domain solutions (24%). TDWI suspects that most of these organizations are still in early generations of MDM. Even so, some single-domain solutions fulfill the business requirements of organizations that needn't share much data across functions, so they are not compelled to move to multi-domain MDM solutions.

Sometimes, generational MDM solutions coexist. Consolidating numerous, related point solutions is rarely done fully across the board, because it takes time to consolidate systems and some point solutions make sense remaining standalone. Therefore, it's no surprise that many respondents (35%) have a mix of single-domain and multi-domain MDM solutions.

USER STORY

MULTIPLE DATA HUBS MAY BE A CONSEQUENCE OF MULTIPLE DOMAINS.

"I wanted a central MDM hub, so we could source all master data for all domains from one location," said an information architect. "But individual businesses saw that as a time hindrance. So we built a series of small hubs, with one per employee, site, and customer domains. We built the hubs quickly, so that business people could see results quickly, namely that they are in compliance. Reporting and integration tools give us visibility across the hubs. I'm okay with the multi-hub approach now, because it got us into multi-data-domain MDM faster than we would have with a large, single hub."

Bidirectional MDM Architecture²

There's a trap for catching cockroaches called the Roach Motel. TV commercials have made the product famous by using the slogan, "Roaches check in but they never check out!" Oddly enough, some types of user-designed databases are similar to the Roach Motel, such that "data checks in but rarely checks out!" (See Figure 9.)

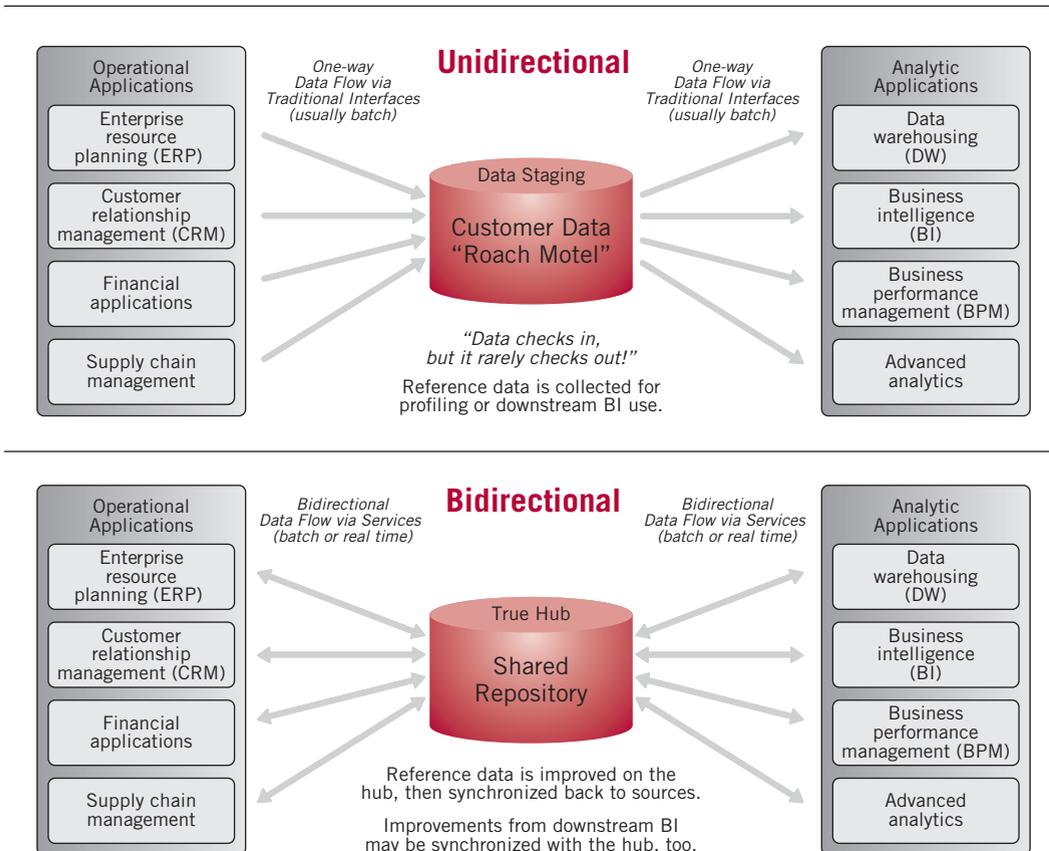


Figure 9. Instead of unidirectional "roach motel" MDM, embrace bidirectional MDM. (Source: TDWI)

² This section of the report is drawn from the article "Avoid Roach Motel MDM Architecture," published in the TDWI Experts newsletter on June 16, 2011. The graphic in Figure 9 comes from a TDWI Webinar in March 2011 titled "Master Data Management for a Single Customer View." You can read the article and replay the Webinar on tdwi.org.

Unidirectional MDM fails to share improvements it makes to reference data.

Roach Motel Approaches to MDM

For example, most homegrown MDM solutions focus on customer data. Central to the solution is a database that serves as the master repository for customer reference data. Typically, customer data flows one way—from multiple operational applications (ERP, CRM, financials) into the database. If data comes back out, it usually goes straight into downstream databases, such as data warehouses or analytic data marts, or it may go into databases used only for sales contacts, marketing campaigns, and direct mail. Rarely does the customer data flow back upstream to enhance the operational applications and databases from which it came.

This kind of “roach motel” MDM architecture is good for profiling and studying customer reference data. It can be handy for documenting the lineage of aggregated reference data, and it’s just fine for customer-base segmentation and other analytic applications. However, roach motel MDM is inherently one way, so it’s bad whenever you need to improve reference data in a central place and then publish it to a wide variety of operational applications.

Although roach motel MDM architectures are sometimes called a customer data hub, it’s not really a hub. A roach motel is merely an operational data store (ODS) or, worse, a data staging area. If you want to reach the full potential of MDM—and especially if you want to operationalize MDM—then you need a real hub that can do more than just aggregate data and pass it to a short list of targets.

It’s not a hub unless information flows both in and out of it freely.

Deploy a True Customer Data Hub

If a data hub is really a hub, data flows bidirectionally, both into the hub and back out to upstream applications and data sources from whence it originated. This is called a *closed-loop data flow*. For this to happen, you usually need a wide range of interfaces, both old ones (database calls in batch) and new ones (data services in real time).

Furthermore, a data hub doesn’t just manage or move data—it improves data. All MDM applications should improve reference data on the hub and then synchronize it back to source systems so they benefit, too. Typical improvements include data standardization, matching, deduplication, and identity resolution. As with the roach motel, reference data from a hub can still flow downstream to BI data stores. Yet these may also add value to reference data and then sync with the hub.

By definition, MDM enables consistent, apples-to-apples sharing of reference data across multiple IT systems, as well as across the business units that own and/or use those systems. As a wise data management professional once said, “If you’re not sharing data, you probably don’t need MDM.” Likewise, if you’re hoarding reference data in a roach motel database, you’re not sharing.

Recommendations for Bidirectional MDM Architecture

Avoid roach motel MDM architecture. Make reference data fully sharable by managing it through a true hub that supports data movement in many directions.

Make MDM’s reference data fully accessible. This approach means supporting a wide range of interfaces, including Web services, perhaps over a service bus.

Improve reference data. Don’t just manage it. Aggregation adds value. Add more value via data quality functions, such as standardization, deduplication, and identity resolution.

Synchronize reference data across multiple applications. After all, the point of MDM is to enable data sharing, not data hoarding as seen in roach motels.

Users' MDM Tool Portfolios

We've already discussed generational trends toward multi-data-domain MDM and bidirectional architectures for MDM. An equally influential trend is the one away from hand coding and other homegrown methods toward greater use of vendor tools. It's natural that users will adjust their tool portfolios to introduce tools that are better suited to MDM or dedicated solely to MDM. To gain insight into organizations' MDM tool portfolios, the MDM survey asked: "Which of the following best describes your organization's MDM portfolio today? For your organization's next generation, how would you prefer that the MDM portfolio be?" (See Figure 10.)

**Which of the following best describes your organization's MDM portfolio today?
For your organization's next generation, how would you prefer that the MDM portfolio be?**

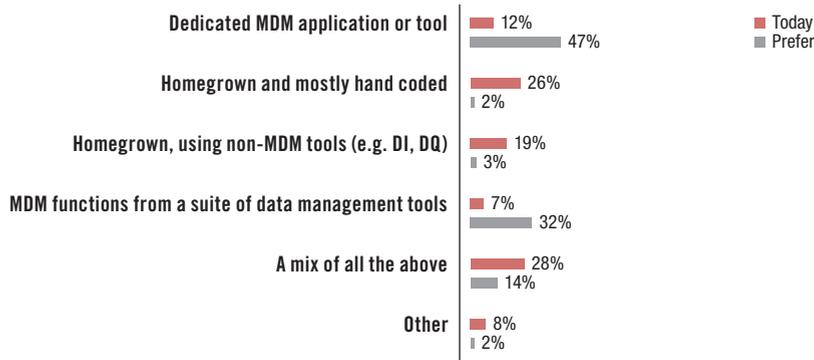


Figure 10. Based on 369 respondents.

Tool Portfolio for Today's MDM Solutions

Most MDM solutions today are homegrown, to some degree:

- Homegrown and mostly hand coded (26%).** In the old days, to prove that an approach to data management should be adopted, we built a prototype using in-house technologies and personnel. If the prototype proved its concept, we threw away the prototype and then acquired tools, personnel, and other resources so we could build a proper solution. Unfortunately, the practice today is to keep the prototype and supply no further resources. More than most solution types, the world of MDM is littered with first-generation prototypes that have outlived their usefulness. These prototypes often consist of nothing more than an ODS for managing aggregated reference data, plus hand-coded SQL for matching and identity resolution and Open Database Connectivity (ODBC) as the only interface. Although these homegrown, hand-coded, first-generation MDM silos proved the usefulness of MDM, they are limited in functionality, and so they hold back organizations wishing to grow into more sophisticated MDM practices.
- Homegrown, using non-MDM tools (19%).** For many technical users, their primary "MDM tool" today is actually a tool for DI or DQ. As noted earlier in this report, an MDM solution may aggregate and transform reference data with a DI tool, as well as verify and deduplicate reference data with a DQ tool. Either tool can assist with standardizing reference data. Although not ideal, DI and DQ tools can automate many or most of the technical requirements of an MDM solution. Most organizations already have a level of competency with DI and DQ they can leverage.

Today most MDM solutions are totally or partially homegrown.

There's pent-up demand for much more tool use.

Tool Portfolio for the Next Generation of MDM Solutions

If surveyed users had their way, they would move en masse away from homegrown methods and toward one of two tool types:

- **MDM functions from a suite of data management tools (32%).** As just explained, there's a strong practice today of co-opting DI and DQ tools into the development and deployment of MDM solutions. However, the DI and DQ tools of leading vendors are today now part of a suite of data management tools, a suite that also includes tools for MDM, data profiling, replication, federation, and even governance and stewardship. As time passes, vendors of such suites more deeply integrate the tools of the suite (and build or acquire more tool types). Consequently, data management practices are moving toward solutions that incorporate DI, DQ, MDM, and related disciplines, all delivered from a single data management platform, whether as a single solution or several that are tightly integrated and coordinated.
- **Dedicated MDM application or tool (47%).** It's ironic that so many MDM solutions today are homegrown, using technologies that are not native to MDM. Yet most users surveyed would prefer to go to the other extreme by using an application or tool that is expressly built for MDM. Users interviewed by TDWI for this report explained that they're struggling to get the resources they need to replace their first-generation prototypes or their second-generation DI/DQ-based solutions. The perception is that a dedicated MDM tool will support more functions than a user organization has the resources or understanding to build themselves.

USER STORY

FOR MDM, SERVING THE BUSINESS IS MORE IMPORTANT THAN TECHNICAL ELEGANCE.

"Everyone in our company—including IT—stays focused on a short list of business goals," said the director of enterprise information management at a financial services firm. "For example, when it comes to MDM, we stay focused on building a single view of each customer, because that's what the business says it needs most. Due to the focus on business goals, no one in IT invests in a new technology until we've proved the business value of what it does. And we can usually provide value with tools and platforms we already have in-house.

"Put all that together, and you can see why we build single customer views with a hodgepodge of tools we already have—mostly tools for data profiling; database management; and extract, transform, and load. Our first-generation homegrown MDM solution isn't elegant, but it does exactly what the business wants today. If, in the future, the business asks for more demanding MDM capabilities—say for real-time reference data updates—we'll consider acquiring new technologies. But that would happen only after we're satisfied that we can't build those capabilities ourselves."

Replacing MDM Platforms

An MDM platform can take many forms, involving any of the tool types and homegrown methods mentioned in the discussion of MDM tool portfolios, as well as dedicated tools and applications for MDM. Regardless of the definition, some users are contemplating a replacement of their MDM platform—and some aren't. (See Figure 11.)

Roughly one-half of surveyed organizations (46%) have no plans to replace their MDM platform. This statistic is a mix of good and bad news. The good news is that some organizations are pleased with their current platform, because it satisfies business requirements for MDM. The bad news (confirmed in other passages of this report) is that other organizations consider their MDM solution inadequate, and they would like to replace it. Alas, they cannot secure approval and funding.

The other half of surveyed organizations plan (50%) to replace their MDM platform. If these plans pan out over the next five years, the average MDM solution will be quite different—and hopefully far better—than today’s average. This is good news, considering that many users are frustrated by the limitations of early-generation solutions.

One-half of users plan to replace their MDM platform within five years.

When do you plan to replace your current primary MDM platform?

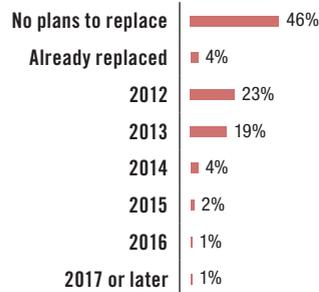


Figure 11. Based on 369 respondents.

To learn why some organizations are making these plans, this report’s survey asked respondents to enter “reasons why you have [replaced] or might replace your primary MDM platform.” In the users’ own words, here’s a sample of their responses to the question, sorted by recurring themes:

New functionality: Many respondents hope to gain “additional functions, plus capacity and design corrections.” For some, this involves “expanding use to serve both analytic and operational MDM needs.”

From new platforms, users want more tools, functions, architecture, governance, domains, and enterprise scope.

Complaints against current platform: “Ad hoc, mishmash of different approaches,” “poorly architected, not documented,” “not fit for purpose,” and a “straw man approach.”

Homegrown issues: The “current solution is a homegrown point solution and cannot be extended to other business objects or data sources.” Our “homegrown MDM tool has a lot of limitations and is not very user friendly.” We’re “moving away from homegrown, hand-coded MDM to a vendor solution.”

Organizational problems: “No governance in place,” “lack of organizational will,” “lacking executive sponsorship,” “no business case,” “turf wars,” “no clear ownership,” “does not provide business stakeholders with value,” and “everything is siloed by business units.”

Enterprise scope: “The current platform just covers part of the entities and applications in the company.” It’s “not integrated across the enterprise,” and we “need an enterprise MDM solution across all data domains.” “MDM is siloed,” and we “need to standardize on infrastructure that can support all enterprise business needs.” “There is no proper sharing of data across the organization, despite the widespread usage of it.” We “need to broaden the scope and reach of MDM.”

Miscellaneous goals: “Scalability,” “better process integration,” “SOA architecture,” “regulatory requirements,” “better performance,” “new custom data model,” and to “expand to other domains.”

The process of replacement: For many organizations, the next generation of MDM is attained through an incremental evolution, not a sudden switchover. “It is not a case of replacing the MDM platform,” said one of the survey respondents, “but rather a question of step-by-step migration to a global platform across different regions and countries.”

Replacing a platform takes time and planning.

Quantifying MDM Generations

At this point in the report, we've defined the terms and concepts of next generation MDM, listed the drivers that push organizations into a new generation, and discussed common generational changes. As you no doubt noticed, the next generation of MDM involves many different *options*, which can be tool features and tool types, user-oriented techniques and methods, and team or organizational structures. Now it's time to draw the big picture, so we can answer questions about these options, such as:

What are the many options that users need to incorporate into the next generation of their MDM solutions? Which ones are users adopting and growing the most? Which are in decline? At what rate is generational change occurring?

MDM options available today are diverse in type and maturity.

To help quantify these and other questions, TDWI presented survey respondents with a long list of options for MDM. (See the left side of Figure 12.) These options include a mix of vendor-oriented product features and product types, as well as user-oriented techniques and organizational structures. The list includes options that have arrived fairly recently (real-time publishing or instantiation of reference data, for example), have been around for a few years but are just now experiencing broad adoption (event processing, identity resolution, BPM), or have been around for years and are firmly established (DQ functions, batch processing). The list is a catalog of available options for MDM, and survey respondents' answers enable us to sort and interpret the list in a variety of ways.

Concerning the list of MDM options presented in the survey, TDWI asked respondents: "For each of the following techniques, features, and practices that may be used in or around an MDM solution, select one button per row [from the three buttons presented] to describe your organization's level of commitment." Each row (representing an MDM option) presented three multiple-choice answers:

1. No plans for using
2. Using today; will keep using
3. Will use within three years

Survey responses show which MDM options are in common use today.

Survey responses for "using today" and "will use" are charted as pairs of bars on the left side of Figure 12. Within each pair of bars, the value for "using today" is the percentage of survey respondents who claim to use that option now. Similarly, the value of "will use" is the percentage of survey respondents who anticipate adopting that option within three years.

Survey responses also indicate how usage of options for MDM will increase or decline.

The pairs of bars on the right side of Figure 12 paint a slightly different view of option usage. The "potential growth" bars calculate the per-option difference between responses for "using today" and "will use"; this metric provides an indication of how much the usage of an MDM option will increase or decrease. An option's commitment value is the percentage of survey respondents who did not select "no plans for using"; this metric provides an indication of how many organizations are committed to using that option, whether today or within three years.

Potential Growth versus Commitment for MDM Options

Figure 12 reveals several interesting things about the use of tools and techniques for MDM. For example, all of Figure 12 is sorted by the potential growth column, in descending order. In this sort order, "real-time publishing of reference data" appears at the top of the chart, because—with a delta of 36% (greater than any other option)—this option exhibits the highest potential for growth.

For each of the following techniques, features, and practices that may be used in or around an MDM solution, select one button per row to describe your organization’s level of commitment.

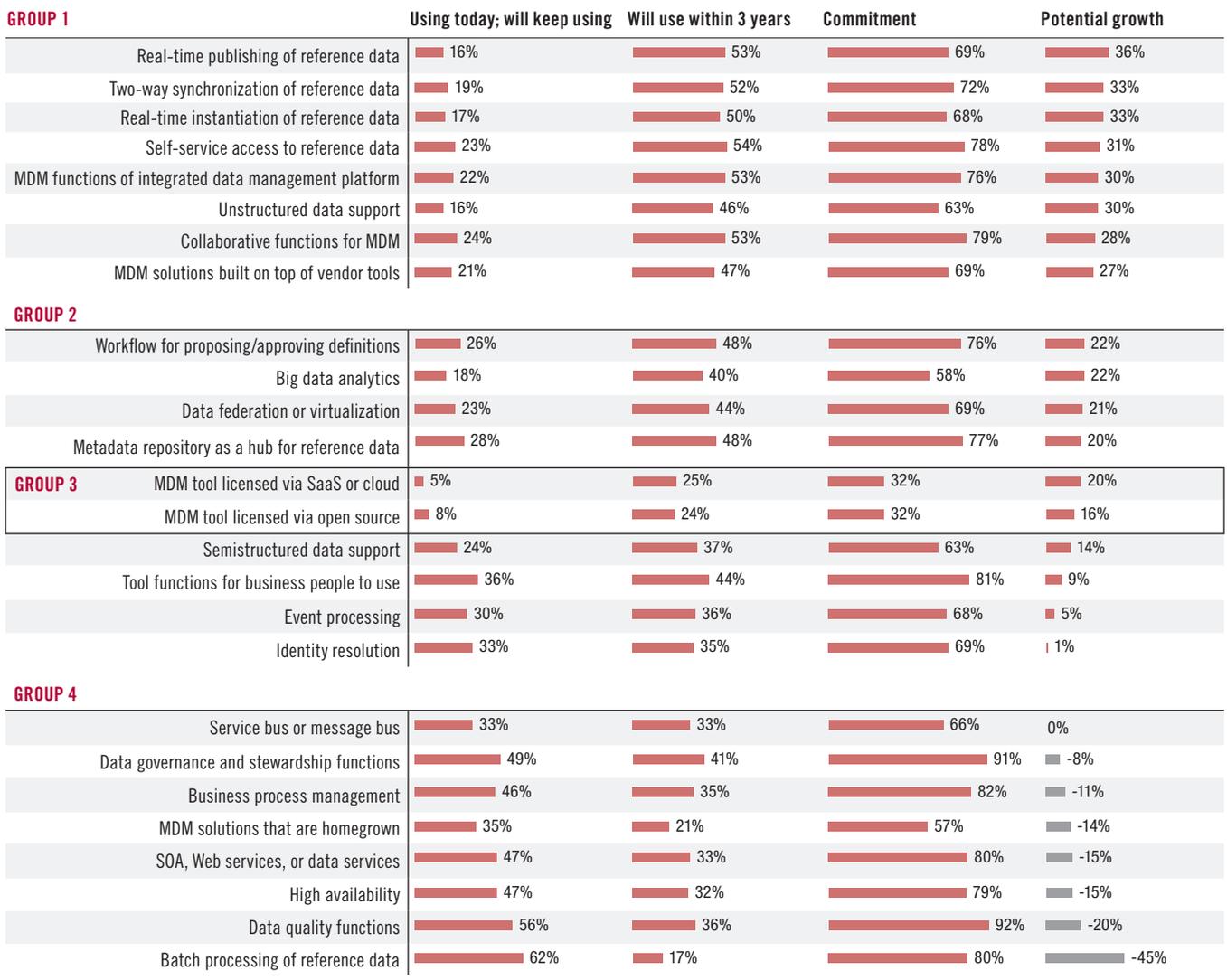


Figure 12. Based on 369 respondents. The charts are sorted by the “Potential Growth” column.

However, not all organizations plan to use the “real-time publishing of reference data” option. In the commitment column, we see that 69% of survey respondents have committed to implementing it, whether today or within three years. By scanning the commitment column, you can see that 69% is a moderate level of commitment for an MDM option. Given the strong potential growth and moderate commitment of “real-time publishing of reference data,” it’s likely that most organizations will include some form of it in their next generation MDM solutions.

From this, we see that there are two forces at work in Figure 12, as well as in the planning processes of user organizations:

Commitment and potential growth are two indicators for quantifying the future of MDM options.

- Potential growth.** The potential growth value is the product of “will use” minus “using today,” and the delta provides a rough indicator for the growth or decline in usage of options for MDM over the next three years. The charted numbers are positive or negative. Note that a negative number indicates that the use of an option may decline or remain flat instead of grow. A positive number indicates growth, and the size of the number suggests a growth rate.
- Commitment.** The commitment value represents the percentage of survey respondents who did *not* select “no plans for using.” Note that the measure of commitment is cumulative, in that the commitment may be realized today or in the near future.
- Balance of commitment and potential growth.** To get a complete picture, it’s important to look at the metrics for both growth and commitment. For example, some features or techniques may have significant growth rates, but within a weakly committed segment of the MDM user community (e.g., federation, open source, semistructured data). Or they could have low growth rates (or even flat or declining rates), despite being strongly committed through common use today (e.g., data governance, homegrown solutions, batch processing). Options seeing the greatest activity in the near future will most likely be those with strong ratings for both growth and commitment (e.g., real-time publishing or instantiation, two-way synchronization, or self-service access to reference data).



Figure 13. Options for next generation MDM plotted by growth and commitment. Plots are approximate, based on values from Figure 12.

To visualize the balance of growth and commitment, Figure 13 includes the potential growth and commitment numbers from Figure 12 as opposing axes of a single chart. MDM options are plotted in terms of growing or declining usage (x-axis) and narrow or broad commitment (y-axis).

Trends for Master Data Management Options

Figures 12 and 13 reveal that most MDM options will experience some level of growth in the near future. The figures also indicate which options will grow the most versus those that will stagnate or decline. Four groups of options cluster together based on combinations of growth and commitment. (See the groups circled, numbered, and labeled in Figure 13.) Furthermore, the groups are indicative of trends in MDM and other data management disciplines.

Rates of growth and commitment identify four groups of options for MDM.

Group 1: Strong to moderate commitment, strong potential growth

Options that have the highest probability of altering best practices for MDM are those with a strong potential growth (according to survey results), coupled with a moderate or strong organizational commitment. Group 1 meets both of those requirements, and it includes tool types and techniques that TDWI has seen adopted aggressively in recent years. In many ways, Group 1 is the epitome of next generation MDM, because of its mix of leading-edge options supported by real-world organizational commitment. Furthermore, today's strongest trends in BI, data warehousing, and data management are apparent in Group 1.

Real-time operation. The movement toward real-time operation is the strongest trend in data management and BI today, and it's affecting vendor tools and user practices in MDM, as well as in DI, DQ, data warehousing, reporting, and analytics. This trend is evident in Figure 13, with the two highest projected growth rates for "real-time publishing of reference data" and "real-time instantiation of reference data." The former instantaneously applies reference records to applications as they are created or updated (instead of in batch later). The latter assembles referential records from multiple sources in real time (instead of aggregation in batch, which is today's norm).

Expect strong growth with MDM options for real time, collaboration, data synchronization, tool use, and multistructured data.

Bidirectional operation. One of the most dramatic generational changes seen in MDM is the shift from one-way aggregation of reference data (usually to profile and study it) to the "two-way synchronization of reference data" (which applies improved reference data back upstream to source systems where it originated). Without this capability, MDM is an academic exercise in data profiling, instead of the value-adding enrichment of applications data that it should be.

Collaborative functions. Data management specialists today regularly collaborate with their comrades in related IT disciplines and with business people, to assure that data management work aligns with business goals and enterprise-scope data standards. Automating collaboration requires "collaborative MDM tool functions" and "self-service access to reference data." Keeping business people in the loop demands "tool functions for business people," which is nearby in Group 2.

Vendor tools instead of homegrown methods. One of the strongest trends across data management disciplines is the movement from homegrown and hand-coded methods to using vendor tools as the primary medium for building and deploying MDM solutions. This trend is seen in Group 1 in Figure 13 as "MDM functions of integrated data management platforms" and "MDM solutions built atop vendor tools." Conversely, "MDM solutions that are homegrown" (in Group 4) are poised for flat growth or decline.

Unstructured data. Another prominent trend in data management involves embracing multistructured data, instead of just structured data. This trend is represented by "unstructured data support" in Group 1 and "semistructured data support" in Group 2.

Group 2: Strong to moderate commitment, good potential growth

Although not poised for the high rate of growth that we can expect from Group 1, Group 2 options for MDM should experience good growth, backed up by a strong to moderate commitment from user organizations.

Good growth is coming for MDM functions for workflow, analytics, federation, repositories, and event processing.

Workflow. Early in this report, we saw that MDM involves (among other things) a complex collaborative process. Much of the process is a series of steps for proposing, reviewing, developing, and policing the business entity definitions and reference data standards that are the product of MDM. To automate this, some MDM tools now support a “workflow for proposing/approving [MDM] definitions.” This kind of workflow usually works within a single tool. As a more sophisticated alternative, some organizations use tools for “business process management” (Group 4), which can reach across multiple departments and business processes.

Analytics. MDM tools and techniques have long been applied to data warehouses, where MDM provides the consistent data usage and source audit trail that BI applications need. Although it’s still early days for “big data analytics,” users no doubt will eventually apply MDM to the more advanced forms of analytics associated with big data.

Virtual MDM. MDM is on the cusp of incorporating virtual techniques, as seen with the option “real-time instantiation of reference data” (Group 1). One way to implement this kind of instantiation is with “data federation or virtualization” techniques (Group 2).

Master data sets. In almost all MDM architectures, there is a master data set at the hub where reference data is aggregated for profiling, improvement, and publication. For this purpose, users seem to prefer a repository approach, as opposed to the registry style of MDM architecture. As homegrown MDM solutions fade out, so do their hand-coded master ODSs. Because MDM professionals usually have experience with metadata repositories (especially those built into DI and DQ tools), it makes sense to use a “metadata repository as a hub for reference data,” an approach that already has a strong commitment from users.

Events. An MDM solution experiences a meaningful event every time a referential record is created or updated. Likewise, changes to entity definitions are significant. “Event processing” has a moderate commitment as a technology for handling events involving reference data. In some configurations, event processing can operate in real time—a direction in which MDM is going.

Group 3: Weak commitment, good growth

Despite weak commitment today, expect growth from MDM on open source, SaaS, and clouds.

Data management tools based on open source, software-as-a-service (SaaS), and clouds are still relatively new; MDM tools based on them are even newer. This newness explains the weak organizational commitment to the options of Group 3, namely “MDM tools licensed via SaaS or cloud” and “MDM tools licensed via open source.” Even so, potential growth is good within committed organizations, so we can expect more organizations to use these options soon.

Group 4: Strong to moderate commitment, flat or declining growth

Group 4 includes essential options, such as services, governance, quality, homegrown solutions, and batch processing. In fact, these options are some of the most common ones in use today for MDM and other data management disciplines. If they are so popular, though, why does the survey show them in decline or suffering flat growth?

Sometimes an established tool type or user technique reaches a saturation point, because most organizations that need it have already deployed it in most of the situations where it's needed. After this point, deployments receive maintenance but little or no new development. This may be the case with MDM options for services, governance, and quality. The case with batch processing and homegrown solutions is different; they are legacy methods that are being replaced by more modern options for real-time operation and vendor tools, respectively.

Services. The MDM option “SOA, Web services, or data services” scored strong commitment, and “service bus or message bus” scored moderate commitment. To be honest, these varied approaches to services turned out to be much more established than TDWI thought they'd be. That's good, because modern MDM solutions need a wide range of services to enable real-time operation, reach operational applications, integrate with Web environments, and leverage enterprise messaging infrastructure. Furthermore, a generalized service tends to get more reuse than older approaches that develop a hard-coded interface or unit of work.

Governance. According to survey results, “data governance and stewardship functions” scored extraordinarily high commitment (91%), but poor prospects for growth (-8%). The high commitment is good news, because (as discussed earlier in this report) the presence of a governance board (or an equivalent program, like stewardship) increases the likelihood of MDM acceptance and eventual enterprisewide coverage.

Quality. Despite low potential growth (-20%), the option “data quality functions” scored the second-highest score for “using today” (56%) and the highest score for commitment (92%). For good reasons, DQ functions are deeply ensconced in MDM solutions and will continue to be. As noted earlier, many MDM solutions resemble DQ solutions, because improving MDM's reference data so often relies on DQ functions, such as matching, deduplication, standardization, verification, and data append.

Batch processing and homegrown solutions. Ironically, “batch processing of reference data” had the highest score for “using today” (62%), yet the lowest score for potential growth (-45%), which indicates decline. Despite what the survey suggests, this strongly established technology and its attendant practices won't go away. They're incredibly useful, as well as too entrenched to dislodge. Even so, survey respondents are probably thinking ahead to when they will put more effort into real-time operation than older technologies like batch processing. Likewise, as organizations use more vendor tools, the use of “MDM solutions that are homegrown” will conversely decline.

Old or saturated options won't go away, but they will be a smaller percentage of MDM options when joined by new ones.

USER STORY

MDM TAKES MANAGING EMPLOYEES TO A HIGHER LEVEL.

“I think we're pretty innovative in how we apply MDM to employee data,” said the director of enterprise information management at a consulting firm that serves the financial services industry. “You see, when a new consulting engagement comes in, we have to very quickly pull together a team drawn from multiple consulting practices across our firm. And we have a lot of practices, many of them autonomous, because they came in through mergers and acquisitions.

“To assist with this, we've developed highly detailed reference data about each employee—or associate or consultant, as we call them. And the reference model is applied to all IT systems across our diverse, internal practices. This way, an engagement manager can run a few cross-system queries to locate consultants with a specific expertise, availability, language skill, and geographic location. Without this capability, the manager would spend weeks assembling a team through e-mails and phone calls. So, MDM gives us the speed and agility we need to pull together just the right team and get them out there serving our corporate clients as soon as possible.”

Vendors and Products for Master Data Management

Because the firms that sponsored this report are all good examples of software vendors that offer tools, platforms, and services for MDM, let's take a brief look at the product portfolio of each. The sponsors form a representative sample of the vendor community, yet their offerings illustrate different approaches to MDM.³

DataFlux

As we've seen in this report, MDM solutions often resemble solutions for DI and DQ. In fact, many users rely heavily on DI and DQ tool functions for achieving their MDM goals. In a related issue, data governance is a critical success factor for data management practices that require a collaborative process and change management, as DQ and MDM do in spades. Years ago, DataFlux recognized the inherent relationships and overlaps among user practices and required technologies for DI, DQ, data governance, and MDM. In response, they built a unified platform that provides integrated implementation tools, deployment infrastructure, and collaborative functions for all these and other data management disciplines. As part of the DataFlux Data Management Platform, DataFlux qMDM includes a bidirectional, real-time data hub that enables users to create custom master views of multiple business entities and their data domains. Other notable functionality includes a data stewardship console, automation for exception processing, hierarchy management, and a patented matching engine for entity resolution.

IBM

For years, IBM offered multiple MDM applications and tools, each with strengths in particular industries or focused on specific data domains or business functions. However, in response to the strong trend among users to support many entities, domains, and processes through a single platform, IBM recently unified its MDM offerings in IBM InfoSphere Master Data Management V10. In addition to product consolidation, V10 includes workflows for collaborative authoring, and it leverages common probabilistic matching and linking technology. V10 also includes new functionality, such as an application toolkit for GUI development, so users can create their own MDM-powered applications that facilitate access to master and reference data for a broad range of user types. One of the bolder innovations of IBM InfoSphere MDM V10 is that it includes IBM BPM Express. This makes sense, because the point of MDM is to improve business processes, and enterprise-scope MDM reaches across multiple business processes and departments. IBM BPM Express enables users to implement policies and coordinate multistep, cross-functional workflows for data stewardship and data governance.

Oracle

Oracle offers several hub-based MDM applications, under the assumption that each data domain has unique requirements. But these are *not* siloed solutions. They all sit atop Oracle Fusion Middleware, which provides a common infrastructure for integrating Oracle's MDM solutions and other Fusion-based applications. Furthermore, Fusion provides functions that enhance MDM, including shared services for application integration, BPM, DQ, DI, metadata management, business rules, event-driven architecture, Web services, user identity management, and analytics. These services make the Oracle MDM Hubs operate bidirectionally in real time, within workflow processes for the approval and publication of reference data. Furthermore, the diverse services of Fusion enable users to deploy MDM in a variety of MDM styles and architectures. Oracle Fusion Middleware is now available on Oracle Public Cloud, which in turn makes Oracle's multi-data-domain MDM solutions cloud based.

SAP was one of the first vendors to offer a dedicated MDM application. The commitment continues today with SAP NetWeaver Master Data Management, which is now part of a comprehensive suite of enterprise information management (EIM) tools that are integrated in both development and deployment environments. The EIM suite and NetWeaver give MDM the integration with other data management disciplines and operational applications that is a top priority for next generation MDM. SAP NetWeaver MDM is inherently multi-data-domain via one platform, and it provides a single version of master data for supplier, product, customer, or user-defined data objects in heterogeneous environments, not just SAP. SAP Master Data Governance (MDG) is a domain-specific governance application, natively integrated with SAP Business Suite to centrally create, change, and distribute master data. SAP MDG users can leverage investments in their SAP data model, user interface, and existing business logic and configuration for the creation and validation of master data. The solution enables governance, compliance, and transparency through integrated staging, approval, and central audit trail—for both SAP and non-SAP environments. SAP NetWeaver MDM and SAP MDG integrate with SAP BusinessObjects Information Steward, a business user interface for profiling, DQ assessment, remediation, and continuous monitoring.

SAP

TDWI has seen open source tools build up a presence in data management in recent years. Open source for DI came first, and then DQ and now MDM. Talend's holistic integration solutions encompass tools for all three, plus enterprise service bus (ESB), BPM, and data profiling and stewardship. Furthermore, all Talend tools are built atop a shared platform—the Talend Unified Platform—with a common metadata and artifact repository, only one server to deploy, and all development GUIs presented in Eclipse. In this unified environment, work done with Talend Open Studio for MDM (the open source edition) or Talend Enterprise MDM (the commercial enterprise edition) is easily coordinated and integrated with data management work done with Talend tools, which is a next generation requirement. Both tools present an active data model and a modeling interface so users can flexibly model and manage master data for any domain, as well as include built-in DQ and data transformation capabilities. With the recent release of v5, Talend Open Studio for MDM and Talend Enterprise MDM are now cloud ready and therefore suited to deployments on-premises, on a variety of cloud types, via SaaS licenses, and on hybrids of these.

Talend

EXPERT COMMENT

MDM CANNOT SOLVE YOUR DATA GOVERNANCE PROBLEMS. ONLY A DATA GOVERNANCE PROGRAM CAN DO THAT.

“There is a growing body of empirical evidence which says that attempts to deploy a master data management solution are all too often diverted into an extended introspection of enterprisewide data governance disciplines,” according to David Loshin, the president of data management consultancy Knowledge Integrity, Inc. “In other words, many organizations on the road to MDM find that the process- and semantics-based challenges to resolving multiple views into a conceptual ‘single source of truth’ far outpace the technical challenges of MDM.

“As a case in point, our consulting practice regularly sees organizations that bought a vendor tool for MDM, thinking that it would solve all their problems. Indeed, a tool ably solves MDM implementation issues, but it cannot resolve a lack of strategies, policies, and standards for enterprise data.

“Data governance is a desirable practice, but not when done ad hoc in the middle of an MDM project. Instead, user organizations should seek to simultaneously mature the master data program with a data governance initiative, using an evolutionary program plan that delivers incremental tactical value. This is less likely to stall the MDM program (as compared to ad hoc governance or none at all), while still improving enterprisewide information sharing.”

Top 10 Priorities for Next Generation MDM

The news in this report is a mix of good and bad. Half of the organizations interviewed and surveyed are mired in the early life cycle stages of their MDM programs, unable to get over certain humps and mature into the next generation. On the flip side, the other half is well into the next generation, which proves it can be done.

Ten priorities can guide organizations into next generation MDM.

To help more organizations safely navigate into next generation MDM, let's list its top 10 priorities, with a few comments why these need to replace similar early-phase capabilities. Think of these priorities as recommendations, requirements, or rules that can guide user organizations into the next generation.⁴

1. **Multi-data-domain MDM.** Many organizations apply MDM to the customer data domain alone, and they need to move on to other domains, such as products, financials, and locations. Single-data-domain MDM is a barrier to correlating information across multiple domains.
2. **Multi-department, multi-application MDM.** MDM for a single application (such as ERP, CRM, or BI) is a safe and effective start. But the point of MDM is to share data across multiple, diverse applications and the departments that depend on them. It's important to overcome organizational boundaries if MDM is to move from being a local fix to being an infrastructure for sharing data as an enterprise asset.
3. **Bidirectional MDM.** "Roach motel" MDM is when you extract reference data and aggregate it in a master database from which it never emerges (as with many BI and CRM systems). Unidirectional MDM is fine for profiling reference data, but bidirectional MDM is required to improve or author reference data in a central place and then publish it out to various applications.
4. **Real-time MDM.** The strongest trend in data management today (and BI/DW, too) is toward real-time operation as a complement to batch. Real time is critical to verification, identity resolution, and the immediate distribution of new or updated reference data.
5. **Consolidating multiple MDM solutions.** How can you create a single view of the customer when you have multiple customer-domain MDM solutions? How can you correlate reference data across domains when the domains are treated in separate MDM solutions? For many organizations, next generation MDM begins with a consolidation of multiple, siloed MDM solutions.
6. **Coordination with other disciplines.** To achieve next generation goals, many organizations need to stop practicing MDM in a vacuum. Instead of MDM as merely a technical fix, it should also align with business goals for data. MDM should also be coordinated with related data management disciplines, especially DI and DQ. A program for data governance or stewardship can provide an effective collaborative process for such coordination.⁵
7. **Richer modeling.** Reference data in the customer domain works fine with flat modeling, involving a simple (but very wide) record. However, other domains make little sense without a richer, hierarchical model, as with a chart of accounts in finance or a bill of materials in manufacturing. Metrics and key performance indicators—so common in BI, today—rarely have proper master data in multidimensional models.⁶

⁴ See "Master Data Management: Rules for the Next Generation" on the TDWI blog (available on tdwi.org) for an earlier version of this list.

⁵ For more information, see the TDWI blog article "Master Data Management Can Learn from Data Quality," available on tdwi.org.

⁶ For more information, see the *TDWI Experts* article "MDM for Performance Management," available on tdwi.org.

8. **Beyond enterprise data.** Despite the obsession with customer data that most MDM solutions suffer, almost none of them today incorporate data about customers from Web sites or social media. If you're truly serious about MDM as an enabler for CRM, next generation MDM (and CRM, too) must reach into every customer channel. In a related area, users need to start planning their strategy for MDM with big data and advanced analytics.
9. **Workflow and process management.** Too often, development and collaborative efforts in MDM are mostly ad hoc actions with little or no process. For an MDM program to scale and grow, it needs workflow functionality that automates the proposal, review, and approval process for newly created or improved reference data. Vendor tools and dedicated applications for MDM now support workflows within the scope of their tools. For a broader scope, some users integrate MDM with BPM tools.
10. **MDM solutions built atop vendor tools and platforms.** Admittedly, many user organizations find that homegrown and hand-coded MDM solutions provide adequate business value and technical robustness. However, these solutions are usually in simple departmental silos. User organizations should look into vendor tools and platforms for MDM and other data management disciplines when they need broader data sharing and more advanced functionality, such as real-time operation, two-way synchronization, identity resolution, event processing, service orientation, and process workflows or other collaborative functions.



Talend Enterprise MDM

Talend Enterprise MDM provides a comprehensive set of functions for enterprise-scale master data management. It delivers easy-to-deploy, end-to-end functionality for integration, quality, mastering, and collaborating on enterprise data.

Data is critical for the success of every organization. However, analyzing and mastering this critical asset can be difficult as it often relies on e-mail, shared spreadsheets and databases, or proprietary, black-box applications. Further, this data is found in disparate sources in heterogeneous formats and with no process to maintain its accuracy and viability. Even if a common definition is found, duplicates and inconsistencies can lead to erroneous analytics or conflicting policy.

Talend Enterprise MDM presents a single source of the truth for this key asset, your master data. MDM also helps enforce data governance policies and enables a collaborative workflow so that business stakeholders can share responsibility over complete, consistent, and accurate data. It provides a system of record no matter the disparity of source systems, and ensures that master data stays clean and is made available to those who are authorized.

Ultimately, MDM helps organizations meet corporate objectives for increased revenue, decreased costs, risk management, and regulatory compliance. The solution includes the following:

- **Active data model**—Allows you to quickly model and master any domain.
- **Business user interface**—View data in standard composite views to gain a 360-degree view of any master entity or investigate a hierarchy. Customize any view for any user with a custom form designer.
- **Fast search**—Talend Enterprise MDM provides a search capability to search across all domains and tens of millions of master records in less than a second.
- **Workflow and master data process**—Allows you to define and track master data through a process or series of steps using tasks. A graphical trail of process steps provides contextual history as tasks are being resolved.
- **Events management**—Talend Enterprise MDM implements a structure of triggers and processes that allow you to coordinate and orchestrate any process for your master data.
- **Data quality**—Profile data, standardize, and identify duplicate records within the MDM studio. Dedicated survivorship components define common business rules to apply to sets of duplicate records and automate the creation of a single master.
- **Much more**—See our Web site for a complete list.

Until now, typical MDM projects have been extremely expensive and have taken months, if not years, to complete. Talend simplifies MDM with a flexible and open approach to master data projects. It provides a complete solution at a fraction of the cost of other solutions on the market. Best of all, our open source version, Talend Open Studio for MDM, is free to download and try from our Web site, www.talend.com.

TDWI RESEARCH

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



1201 Monster Road SW T 425.277.9126
Suite 250 F 425.687.2842
Renton, WA 98057-2996 E info@tdwi.org

tdwi.org

WP156-EN