

Data Integration

# White Paper

## Ten Mistakes

---

to Avoid When Planning Your CDI/MDM Project

by Jill Dyché and Evan Levy

April 2007

# baseline

---

C O N S U L T I N G

Data Mastered. Value Unleashed.

### Contents

---

❖ Foreword .....	3
❖ One: Misusing the term “MDM” .....	4
❖ Two: Confusing CDI with Data Warehousing .....	5
❖ Three: Overlooking CDI and MDM Development Complexity.....	6
❖ Four: Relying on Source System Data Accuracy.....	7
❖ Five: Over-Emphasizing Business Requirements .....	8
❖ Six: Treating CDI like ETL.....	9
❖ Seven: Assuming Your ERP Package Can Handle MDM .....	10
❖ Eight: Putting Too Much Faith in the “Golden Record” .....	11
❖ Nine: Pitching MDM as an Application.....	12
❖ Ten: Treating Data Governance as a Luxury, Not a Necessity.....	13

## Foreword

---

In his classic business book *The Fifth Discipline*, author Peter Senge advocates the concept of the “shared vision,” explaining that a vision really takes hold when stakeholders communicate clearly. Senge counsels companies to actively reinforce their vision by fostering stakeholder excitement and encouraging early success.

The promise of customer data integration (CDI) solutions to (finally) deliver a single version of the truth about customers has certainly caused excitement. And, as we discuss in our new book, *Customer Data Integration: Reaching a Single Version of the Truth*, there have been some high-profile successes. Companies like Amgen, Intuit, XO Communications, and Royal Bank of Canada have all delivered early wins with customer master data, achieving quantifiable and high-profile business breakthroughs. Consistently, the sustained management of cross-functional master data begins with CDI.

In this *Ten Mistakes to Avoid*, we’ll outline what we’ve seen to be the biggest barriers to successful customer data integration and master data management (MDM) programs. Hopefully, this will help you shape a shared vision for master data at your company, one that can propel you forward.



### ONE

#### *Misusing the Term "MDM"*

Master data management is a hot topic. And as with most emerging trends, it's accompanied by its own mythology, as companies begin their research and decide to adopt it.

But there's already noise around MDM. Watch a Webcast or open a trade magazine, and you'll see MDM used synonymously with data quality, data governance, or data stewardship—all legitimate practices in their own right. BI and data quality vendors have begun embracing the term to rebrand incumbent technologies, irrespective of the breadth of their functionality. A large bank we work with recently appointed a director of master data management. How large is her organization? There's just her.

Her first task? Create an enterprise data model.

In reality, master data management is both a noun and a verb. We define MDM as the set of disciplines and methods used to ensure the currency, meaning, and quality of a company's reference data within and across various data subject areas. It should be used to connote the day-to-day-to-day tactics of managing enterprise master data, including developing data quality and correction processes; implementing privacy standards; ensuring that data availability reflects regulatory guidelines; and, yes, maintaining an integrated data model. MDM tools should be part of—but not all of—a sustained MDM effort. CDI is a subset of MDM that focuses on the customer subject area. It's where most MDM initiatives typically start.

And that's a very big job in itself.

## TWO

### *Confusing CDI with Data Warehousing*

Because CDI is a software solution—in the form of a centralized “hub” that is a clearinghouse for data reconciliation and deployment—comparisons with data warehouses are inevitable. After all, both aim to deploy clean, meaningful information to the enterprise. Both a CDI hub and a data warehouse have clear, often quantifiable business benefits. And both mandate a solid business-IT partnership.

But comparing the two solutions is dangerous, not only in terms of their positioning, but also their usage. Data warehouses are designed and built to support business intelligence, and are meant for use by business people. Best practice data warehouses are those that have been planned around a set of business requirements that inform a series of applications—we call this the BI Portfolio—that are deployed incrementally to the business over time.

CDI, however, is purpose-built for operational data integration. The CDI hub is the ultimate home of customer master data that has been matched, reconciled, and certified, and is available to a series of business applications and systems (not end users). And, unlike with the data warehouse, whose data quality depends on add-on tools, CDI has data quality “baked in” to its processing. Unlike the data warehouse, which usually stores historical detail, summarized data, and time-variant information, the CDI hub stores or points to certified master data about a customer—hence the hub’s role as the bona fide single version of truth for a range of data needs across the enterprise.



### THREE

#### *Overlooking CDI and MDM Development Complexity*

In order to enable an MDM or CDI solution to serve other operational applications, you need to do some programming. To connect a CDI hub to other applications, the interface code—the code that submits and retrieves customer data from the de facto data store—must be modified. There are two basic approaches to modifying interface code to support new CDI or MDM technology. One approach is to modify the operational application to send and retrieve data to and from the hub.

The second approach is specific to environments already leveraging a messaging server—for instance, enterprise application integration (EAI), enterprise service bus (ESB), or other application messaging technology. This approach involves modifying the interface code (once again, relating to an application's submission and retrieval logic) to communicate with the CDI hub. This alternative is transparent to the operational application's software.

It's hard work, since it requires the IT organization to be intimate with the transaction processing logic of individual applications and the accompanying interface or messaging code. And the organization should be savvy enough to retain technical expertise to support the API set and service-oriented architecture (SOA) environment that the CDI or MDM product relies on. Implementing a master data hub is more complex than simply loading a file; it requires the skills to modify transaction processing logic to interface to the hub.



## FOUR

### *Relying on Source System Data Accuracy*

If you've had any involvement with your company's enterprise data warehouse, you've probably encountered the challenge of operational system accountability: that is, convincing source system owners that it's their job to address data quality.

CDI technologies allow the merging of content from multiple sources to create a master record about a customer. While any data quality tool can correct a customer address, it can't identify and resolve duplicate or disparate records and reconcile them into one when subordinate attributes are different.

The quality of the master record is not dependent on the accuracy of the data from an individual source system, since the CDI or MDM technology can spot synonyms, duplicates, and errors in the source data. For instance, when an operational system has duplicate customer entries because of inconsistent descriptive detail (for instance, the customer goes by both "Bob" and "Robert," or has different home addresses), it can selectively match other details to determine which descriptive attribute is best to include in the master record.

The good news about CDI is that the hub can identify unique customers without affecting the day-to-day development activities of operational system programmers. When the time comes and the operational system team decides to correct its data, it can leverage CDI to identify duplicate or disparate customer records.

# 4

### FIVE

#### *Over-Emphasizing Business Requirements*

That's right, we said it. Business requirements—so critical to the success of BI—are a bit overrated when it comes to CDI and MDM deployments. Believe us when we say we never thought we'd see the day.

However, we've already watched a few CDI efforts get tripped up over painstaking conversations with business users about how they would use integrated data. (Variation: "What data do you need that you don't have and what would you do with it?") The inevitable backlash is swift and fierce: "Haven't we already had these conversations? Didn't I give you my requirements when we were building that financial analysis app? If you make me go to another JAD, I'll go ballistic..." and so on.

With CDI, the requirements of individual business users cede to the processing requirements of the applications that need access to customer master data. The requirements stakeholders are the developers of the company's applications, not the business users of those applications. Thus, many initial CDI projects begin with developers who understand what data the target application needs for consistent and reliable processing. Issues such as response time, latency, availability, and data formatting—not history, storage, or breadth of content—are paramount. CDI requirements resemble those of OLTP implementation, not BI development.

This means that, unlike with BI, a CDI hub can rely on the data requirements already identified within the operational applications, not those articulated by business users. The CDI hub improves data access, delivers accurate data, and interfaces with disparate applications, ultimately rendering a more literal single version of the truth.



## SIX

### *Treating CDI like ETL*

While both CDI hubs and ETL (extract, transform, load) tools support the conversion of data for matching and integration, CDI is not ETL. In many ways, CDI is “smarter” than ETL, since it has customer subject area knowledge and the accompanying data matching and validation processing built in.

Because a CDI hub has subject area knowledge, it understands the concept of a customer and the associated attributes—and their association with other attributes. For instance, the CDI hub knows that a given customer can have multiple addresses, and it can easily spot inaccurate data values and distinguish synonyms.

With ETL, it's left to the developer to create this logic to enable integration. One of the biggest challenges of data integration is understanding the demographics and domain details of data content. An ETL programmer needs to understand the validity and accuracy of a given data element.

Moreover, a CDI hub can support the addition of heterogeneous data sources without incremental complexity. With ETL, because of the human involvement, the complexity is exponential and not linear. ETL needs to be customized each time a new source is added, and relies heavily on developer expertise.

Where ETL is broad, CDI is narrow and deep. ETL is best used when loading data. Because its strength is data movement, ETL has more flexibility on the type of data it supports. Consequently, it can't have subject-area-specific logic built in. This means that ETL can leverage CDI and MDM to load subject-specific master data into another application or data platform, easing the burden on the programmer for developing application logic.

# 6

### SEVEN

#### *Assuming Your ERP Package Can Handle MDM*

“The ERP system is our customer data repository,” a development manager told us recently. “All of our application systems get their customer data from our ERP system. So we already have our customer integration hub.”

Not necessarily.

An enterprise resource planning system is an application focused on supporting the management of a company’s resources efficiently and correctly. This can include invoice/bill processing, staff and HR management, and inventory control, among other things. The goal is to ensure that all processing associated with a company’s resources is managed from a central point. An ERP system is not designed to maintain an integrated, de-duplicated, accurate view of customers, let alone to support other operational systems accessing customer data on a transactional basis.

This difference is why many ERP vendors have chosen to augment their products with either internally developed or partner-furnished software that supports MDM functionality. ERP systems have many strengths, but interfacing to disparate operational applications is not one of them. The processing associated with a CDI or MDM hub requires a different set of data structures and processing functions than an ERP system typically uses. The application interfaces are different. It’s basically an apples-and-oranges comparison.

And then there’s the data quality factor. ERP systems typically address data quality processing through external software or third-party functions. With CDI and MDM, data quality is a core functional offering, and a benefit to the range of systems served by the hub.

## EIGHT

### *Putting Too Much Faith in the "Golden Record"*

Customer master data is often seen as the reconciliation of disparate data into a “golden record,” implying a single, authoritative record about each individual customer. While many companies are able to represent their customer relationship with a single entry, this might not be practical in more complex businesses.

There are circumstances in which a company might have multiple relationships with its constituents, often represented by multiple locations or human representatives. One of our clients, a major auto parts manufacturer, has multiple contacts at individual customers within a geographical location. If this company isn’t careful, it could merge these multiple relationships into a single customer record, thereby invalidating the data and losing important and distinguishing detail.

This is an issue in identity fraud, in which someone uses a different name or co-opts identifying attributes of another individual with the goal of passing as someone else. In this instance, a single golden record might be risky, since it conflicts with the objective of tracking multiple, related identities. If an individual is using multiple addresses, it’s important to retain historical or related details. There just isn’t a single, specific set of identifying attributes.

However, by having a “best record,” the organization can recognize and maintain the identity of the individual based on the most current details. With the best-record approach, the CDI hub won’t discard or ignore other identifying attributes. The choice of “best” versus “golden” record isn’t a technology choice, but one associated with the business goals of CDI and MDM.



### NINE

#### *Pitching MDM as an Application, Not an Enterprise Resource*

Too many managers launch their CDI and MDM initiatives with the vision of the hub as a “server” to support the needs of a specific application. While we applaud starting an MDM initiative with a specific business need or technology benefit, the hub should ultimately be an infrastructure solution to facilitate the sharing of data across multiple applications and systems.

The key is remaining mindful of the data needs of other applications and systems. As more application teams become aware of the CDI hub, the actual attributes associated with customer identification are likely to change and evolve. It’s also important to realize that there are circumstances in which a customer might have multiple entries within an operational application to support specific processing needs, so de-duplication and record consolidation should be reviewed carefully within each successive system.

This implies the coordination of different application development teams, as well as data management and stewardship functions. Data representation, quality, and correction activities should be implemented in a way that enlists the different operational system stakeholders. The CDI project manager should have one eye on the future of customer data requirements and processing, which means understanding proposed development initiatives and business cases that might be in the pipeline.

This ultimately means pitching MDM and CDI hub solutions as infrastructure: to centralize and consolidate data reconciliation and access for the enterprise at large. The old adage of “start small, think big” is an apt one for the management of master data.



## TEN

### *Treating Data Governance as Luxury, Not Necessity*

Data warehousing and business intelligence professionals are often at the forefront of the data governance debate. Indeed, the data warehouse is often the only platform that intermingles heterogeneous data from across different organizations and systems, creating the need for tiebreaking and policy-making.

We define data governance as the framework, processes, and oversight for establishing policies and definitions for corporate data. Data governance puts structure around the decision-making process that prioritizes investments, allocates resources, and monitors results to ensure that the data being managed and deployed on projects is aligned with corporate objectives, supports desirable business actions and behaviors, and creates value. While this is great for a BI program, it's absolutely critical for master data management.

Data governance involves management in IT as well as the lines of business coming together on a consistent basis to establish the policies, guidelines, and definitions for master data. Such decisions ultimately inform what the customer master record contains, and how reliable and useful it is to the various business applications that access it. (We distinguish data governance and data management, the latter being the day-to-day tactics of ensuring integrated, accurate, and defined enterprise data.)

For instance, at a large phone company we work with, the operations group focuses on the customer's phone number. When the customer needs service, operations focuses on the customer's physical address. Conversely, the finance organization wants to know who's responsible for paying the bill, regardless of where the phone line goes. Finance doesn't care about who uses the phone or data line. At this company, the "user" and the "payer" are frequently different people.

Both operations and finance use the term customer to mean something different. Resolving the definition of customer is not the responsibility of the CDI hub—it's a business decision. The data governance committee arbitrates the differences between these two "customers," and determines how they should be supported, ensuring that both organizations' needs are met.

A key point here is that data governance shouldn't concern itself with data storage, but with how the data is processed and used. Data governance is important because it focuses on the usage of data in the business—ensuring that sustained, meaningful, and accurate data isn't just a goal, but a practice. Master data initiatives like CDI serve as just the pretext for launching a formal and sustained data governance program on behalf of the entire enterprise.



### About The Authors

---

Jill Dyche and Evan Levy are partners with Baseline Consulting, a business analytics and data integration services firm. Both are recognized industry experts on the topics of business analytics, data warehousing, data management, and data integration. Together they have written a new book on selling, launching, and sustaining initiatives for managing customer master data: *Customer Data Integration: Reaching a Single Version of the Truth* (John Wiley & Sons, 2006).

*This article originally appeared in TDWI Quarterly, Q3 2006, and is reprinted with permission.*





# baseline

## C O N S U L T I N G

**Data Mastered. Value Unleashed.**

Baseline Consulting is a management and technology consulting firm specializing in data integration and business analytic services to help companies enhance the value of enterprise data and improve the performance of their business. Baseline's proven, structured approaches uniquely position us to help clients achieve self-sufficiency in designing, delivering, and managing data as a corporate asset.

### **Baseline Consulting Group**

15300 Ventura Blvd., Suite 523  
Sherman Oaks, CA 91403

1-818-906-7638

[www.baseline-consulting.com](http://www.baseline-consulting.com)