Whitepaper
Measuring data quality for successful Master Data Management
Contents

Overview.................................................................................................................................................. 3
What is Master Data Management? ......................................................................................................... 3
  Master Data Modeling Approaches ...................................................................................................... 4
Three Styles of Enterprise Data ............................................................................................................... 4
  Transactional Data .................................................................................................................................. 4
  Analytical Data ....................................................................................................................................... 5
  Master Data .......................................................................................................................................... 5
The Data Quality Problem ....................................................................................................................... 5
Data Quality metrics are essential to successful MDM ............................................................................ 6
Yellowfin Business Intelligence & the MDM process .............................................................................. 7
Summary ............................................................................................................................................... 8

Yellowfin, and the Yellowfin logo are trademarks or registered trademarks of Yellowfin International Pty Ltd. All other names are trademarks or registered trademarks of their respective companies.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Yellowfin does not accept responsibility for any kind of loss resulting from the use of information contained in this document.

The information contained in this document is subject to change without notice. This text contains proprietary information, which is protected by copyright. All rights are reserved. No part of this document may be photocopied, reproduced, stored in a retrieval system, transmitted in any form or by any means, or translated into another language without the prior written consent of Yellowfin International Pty Ltd.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Yellowfin software product – as well as the timing of any such release or upgrade – is at the sole discretion of Yellowfin.

This document version published May 2010 Copyright © 2010 Yellowfin International Pty Ltd.
Overview

Business Intelligence systems are designed to help organisations understand their operations, customers, financial situation, product performance, trends and a host of key performance indicators.

This information is used to make both tactical as well as strategic decisions. Poor intelligence results in poor decision making. The costs can be enormous. Over the last few years, a serious effort has been made to understand the main cause of poor quality business analytics.

The majority of organizations and analysts now agree that one of the main reasons that BI projects fail to deliver is because the operational data feeding the analytical engines is filled with errors, duplications and inconsistencies. If the poor quality reporting is to be fixed, it has to be fixed at its source. This is achieved by addressing poor quality data in the applications that run the business via Master Data Management (MDM).

MDM is the glue that ties analytical systems to what is actually happening on the operational side of the business and Yellowfin BI is the perfect tool to analyse and report on that data.

This paper will discuss how Yellowfin assists in the MDM process; the importance of data quality, and how reporting on relevant metrics is vital for a successful project.

What is Master Data Management?

Master Data Management is all about fixing poor data quality at its source and managing its constant and inevitable change when shared across more than one transactional application. This data represents the business rules around which the transactions are executed. This data also represents the key dimensions around which analysis is done.

Maximum business value comes from managing both transactional and analytical master data. These solutions are called Enterprise MDM. Operational data cleansing improves the operational efficiencies of the applications themselves and the business processes that use these applications. The resultant dimensions for analytical analysis are true representations of how the business is actually running. Yellowfin provides the most comprehensive, integrated BI solution to support an enterprise MDM solution on the market today. The following sections will illustrate how this combination of operations and analytics solves key business problems.

Getting started with MDM is often difficult – does an organisation take a proactive or a reactive approach to MDM? Some organisations start applying MDM in isolated areas before moving on to other isolated areas, and maybe pull these together into an enterprise approach later on. Isolated starting points are more often problems than opportunities, with organisations reacting to problems that require immediate attention rather than taking a strategic approach. Ultimately, the organisation needs to get beyond the hectic fire drills of reactive MDM and also apply proactive MDM, which explores data and metadata to identify opportunities for master data improvement. Surviving MDM in the long run requires a mix of reactive and proactive practices and this is where Yellowfin can really assist in the reporting on the process.
Master Data Modeling Approaches

Data models for master data can be object-oriented, hierarchical, flat, relational, and so on:

- **Master data modeling should be object-oriented.** Recent years have seen vendors’ tools for databases, data modeling, and integration support a mix of object and relational data models. The rise of XML-described data has brought back hierarchical models, which artefacts can represent easily. MDM is ably served by this kind of object-oriented data modeling, given the hierarchical and multidimensional relationships found among most business entities.

- **Flat versus hierarchical models for master data.** There are two approaches to entity modeling worth noting here. At the low end, when MDM simply provides a system of record that lists a record for every instance of an entity (as most customer-data-oriented MDM does), the data model is simply a flat record (albeit a “wide” record with many fields) stored in a relational table or file. At the high end, the model can be a complex hierarchy, as when a large company defines financials with multiple dimensions per region, nation, and office, as well as per charted account and budget. Deciding how flat or how hierarchical the model should be is a basic design decision in MDM. Obviously; the latter takes more time and expertise.

- **Hierarchical models can increase the number of entities modeled.** As an extreme example, this is possible when a complex product is defined as a hierarchical collection of parts and subassemblies or when products have parent-child relationships within a product family.

Three Styles of Enterprise Data

An enterprise has three kinds of actual business data: Transactional, Analytical, and Master

**Transactional Data**

An organisation’s operations are supported by applications that automate key business processes. These include areas such as sales, service, order management, manufacturing, purchasing, billing, accounts receivable and accounts payable. These applications require significant amounts of data to function correctly. This includes data about the artefacts that are involved in transactions, as well as the transaction data itself. For example, when a customer buys a product, the transaction is managed by a sales application. The artefacts of the transaction are the Customer and the Product. The transactional data is the time, place, price, discount, payment methods, etc used at the point of sale. The transactional data is stored in Online Transaction Processing (OLTP) tables that are designed to support high volume low latency access and update.

Solutions that focus on managing the data artefacts under operational applications are called Operational MDM. They bring real value to the enterprise, but lack the ability to influence reporting and analytics.
Analytical Data

Analytical data is used to support the company’s decision making. Customer buying patterns are analysed to identify churn, profitability and marketing segmentation. Suppliers are categorised, based on performance characteristics over time, for better supply chain decisions. Product behaviour is scrutinised over long periods to identify failure patterns. This data is stored in large data warehouses and possibly smaller data marts with table structures designed to support heavy aggregation, ad hoc queries, and data mining. Typically the data is stored in large fact tables surrounded by key dimensions such as customer, product, account, location, and time.

Solutions that focus on managing dimensions data are called Analytical MDM. They master share entities such as financial data hierarchies and GL’s between multiple DW/BI systems across domains.

Analytical MDM products bring real value to the enterprise, but lack the ability to influence operational systems.

Master Data

Master Data represents the business artefacts that are shared across more than one transactional application. This data represents the business artefacts around which the transactions are executed and the key dimensions around which analytics are done.

Maximum business value comes from managing both transactional and analytical master data. These solutions are called Enterprise MDM. Operational data cleansing improves the operational efficiencies of the applications themselves and the business process that use these applications. The resultant dimensions for analytical analysis are true representations of how the business is actually running. The following sections will illustrate how this combination of operations and analytics solves key business problems.

The Data Quality Problem

On the operational side of the business, data is entered manually by thousands of employees across a large number of departments. This is error prone. Many poor data quality problems begin at this point. In addition, each department has its own rules. For example the Sales department rules for entering customer data into its sales automation application are quite different from the Accounting department rules for customer data entry into its Accounts Receivable application.

Another key characteristic of Master Data is that it is not static. It is in a constant state of change.

Changes to Master Data occur on a regular basis with customers, suppliers, contacts, locations, employees, citizens, competitors, distributors, partners, accounts, households, etc. Items like credit worthiness, vendor viability, and addresses for billing are always in a state of flux.

The operational side of the business must keep up with this constant change or business processes break down. If one application sees the change and another one doesn’t, the process step across these two applications will break down.
Data Quality metrics are essential to successful MDM

Once an organization has decided to go down the path of implementing a MDM project they need to institute a data quality report and decide what types of metrics will be used for data quality management.

Too often, data governance teams rely on existing measurements as the metrics used to populate a data quality report. But without a defined understanding of the relationship between specific measurement scores and the business’s success criteria, it is difficult to determine how to react to emergent data quality issues and determine whether fixing these problems has any measurable business value. When it comes to data governance, being able to differentiate between the “who cares” measurements and the “relevant metrics” becomes a critical success factor in managing the business’s expectations for data quality and how they will be able to use it.

The first step is for an organization to examine how data management practitioners can define metrics that are relevant to achieving their business objectives and to do this they must look at the characteristics of relevant data quality metrics, and then provide a process for characterising the business impacts in association with specific data quality issues.

The processes for analysing the raw data quality scores for base-level metrics can then feed into hierarchies of complex metrics, with a number of different views that address the reporting needs of different departments across the organisation.

Ultimately, this process drives the description, definition and management of base level and complex data quality metrics so that the organisations data quality reports reflect exactly what the business requires in terms of metrics as well as defining the metrics as separated from their contextual use, therefore allowing the same measurement to be used in different contexts, with different levels of acceptability thresholds and weights. It also allows the appropriate level of presentation to be generated based on the level of detail expected for the end users specific data governance role and level of accountability.

Then there is the “Who cares” metric – the fascination with metrics for data quality management is based on the expectation that the data is being continuously monitored to make sure an organisation’s knowledge meets the business process owner’s requirements. That the data managers are notified immediately when an issue is identified and remedial action is required. Interestingly, the desire to have a framework for reporting data quality metrics often outstrips the need to define relevant metrics for the business environment. Often analysts assemble data quality reports and then search the organisation for existing measurements from other auditing and oversight activities to populate them.

The challenge emerges when it becomes evident that existing measurements from other auditing or oversight processes may be insufficient for data quality management.

So how do you define relevant metrics? – Well, good data quality metrics exhibit certain characteristics, and defining those metrics to share these characteristics will lead to a data quality scorecard that avoids the dreaded “who cares” response.

• **Business Relevance**: The metric must be defined within a business context that explains how the metric score correlates to improved business performance
• **Measurability**: The metric must have a process that quantifies a measurement within a discrete range
• **Controllability**: The metric must reflect a controllable aspect of the business process; when the measurement is not in a desirable range, some action to improve the data needs to be triggered
• **Reportability**: The metric’s definition should provide the right level of information to the data manager when the measured value is not acceptable
• **Trackability:** Documenting a time series of reported measurements must provide insight into the result of improvement efforts over time as well as support statistical process control.

In addition, it is important to recognise that reporting a metric summarises its value, the reporting environment should also provide the ability to expose the underlying data that contributed to a particular metric score. Being able to review data quality measurements and evaluating the data instances that contributed to any low scores suggests the need to be able to drill down through performance metrics. This allows analysts to get a better understanding of any existing or emerging patterns that contributed to the low score and then assess the impact and help diagnose the root-cause analysis at the processing stage where any flaws were originally introduced.

Relating data quality metrics to business activities means aligning data quality rule conformance with the achievement of business objectives. The data quality rules must be defined to correspond to business impacts.

Applying these processes with the assistance of Yellowfin reporting will result in a set of metrics that can be combined into different reporting schemes that effectively address the data management responsibilities that rest with operational, data and senior managers to support the organisations data governance requirements.

---

**Yellowfin Business Intelligence & the MDM process**

Yellowfin’s Business Intelligence solution has been used in MDM projects to provide enterprises with visibility across the entire MDM project in terms of the quality of data being produced and reported on. For example, if an MDM repository is built, consisting of physical data and is loaded to a central repository, then Yellowfin is used to report on audit and reconciliation activities for Extract, Transform, and Load (ETL) between source systems and MDM hub (typically tabular reports) and report on defect management activities for the project.

Yellowfin has been described as the perfect tool to create a range of dashboard style reports to highlight the data quality issues identified in source systems and within the MDM hub.

Spatial dashboards, for example, provide an excellent way of displaying anomalies and invalid data for given locations. MDM provides a single source of the truth or a single view of the customer. Yellowfin’s Location Intelligence can present a 3D view of the customer by presenting associated customer entities and attributes in the one location as opposed to the 2D view of cross-tab or columnar reports. Additionally, Yellowfin Location Intelligence and 3D view of the customer after MDM implementation allows for enhanced modeling of business value propositions e.g. where can we place this branch, office or park that will provide the greatest benefit?

Yellowfin can retrieve data from multiple sources and report on this. Rather than having a customer repository with physical data, Yellowfin can link to alternate sources of data and show multiple views of the customer on one report i.e. a reference model. For example, a Yellowfin view can point to customer data in one part of a business and use virtual tables within the same view to point to one or more customer data sets in other parts of the business, link these together, and show common and mismatched customer attributes for further work and discussion.

Yellowfin Business Intelligence really assists organizations to achieve a clean, accurate set of data and allows them to manage that in relation to their data scorecard and relevant metrics for successful analysis and reporting.
Summary

Data quality is essential to the success of MDM projects and it’s important to institute data quality reporting at the outset and decide on what types of metrics will be used for data quality management in consultation with the whole organization. The metrics used for creating data quality reports must be relevant and aligned to the businesses overall objectives to gain the maximum results.

Business Intelligence must be used as part of the MDM process; it assists organisations by providing a mechanism to measure how clean, consolidated and accurate their master data is. This ensures quality controls for data reflects the actual operations of the organisation and BI insures that the data being reported on is accurate.

Yellowfin’s Business Intelligence solution has been used in MDM projects to provide enterprises with visibility across the entire MDM life cycle in terms of the quality of data being produced and reported on. This assists organizations to achieve a clean, accurate set of data to be used for critical operational and strategic analysis and reporting.

Find out more
Contact Yellowfin at www.yellowfin.bi
and ask for our proven roadmap to assist you to successfully implement Yellowfin Business Intelligence into your organization.