

Master Data Mix-Ups

Master Data is a key concept in Information Management, but one that can cause confusion.

Chris Daniels asks “What do you mean by Master Data?”



Introduction

Have you ever had a conversation where you are talking at cross purposes? I came across an amusing example when a British friend of mine married an American lady.

She: “I’d really like you to wear suspenders on our wedding day.”

He: “What do you mean? Suspenders are for women! Will you be wearing some?”

She: “Of course not. I won’t be wearing any pants.”

He: “What kind of a woman am I marrying?!”

What this example illustrates is that effective communication between people is only possible if we can be sure we’re talking about the same thing. Because people have different backgrounds and experiences, it’s always possible that the message we’re intending to send is different to the message the other person receives. In the example, the lady thinks the conversation is about braces and trousers, while the man thinks they are talking about garters and underpants.

The need to agree on the meaning of terms is one of the basic drivers behind Master Data Management initiatives. But what do we actually mean when we use the term “Master Data”? And more importantly, do we all mean the same thing? It matters because if we have different interpretations, any conversations we have about Master Data and Master Data Management could be at cross purposes.

Master Data forms an inventory of the “things” and “concepts” that are of interest

Interpretation 1: Master Data is data about things, not events.

In this interpretation, data is divided into two broad categories known as Master Data and Event Data.

Master Data is data describing the individual items about which events, figures and observations can be recorded. Such items include customers, products, locations and periods of time. Master Data forms an inventory of the “things” and “concepts” that are of interest.

Meanwhile, Event Data is data describing events that occur. Event Data can include measured figures, textual observations and information to identify the “things” that participate in the event.

The withdrawal of money from a bank account is an example of an event. In this example, the Event Data to be recorded includes the amount withdrawn, the account number and the location, date and time of the withdrawal. The “things” that participate in the event are the account, the location and the date. It is important to note that while the Event Data includes enough data to identify the “things”, the definitions of those things is described via Master Data. For instance, the Event Data above includes an account number, but information about the account holder and the

account type associated with that number is stored in Master Data.

Published sources seem to be in agreement that this interpretation of Master Data is “correct”, with definitions including the following:

“Master data is the consistent and uniform set of identifiers and extended attributes that describe the core entities of the enterprise” (Gartner)

“Master data is that persistent, non-transactional data that defines a business entity for which there is, or should be, an agreed upon view across the organisation” (Wikipedia)

This interpretation of Master Data is the one that I’ve used throughout my career and there seems to be broad agreement. So why would people think anything different? Actually, it turns out that there are good reasons...

Interpretation 2: Master Data is Reference Data with consensus

The two broad categories of Master Data and Event Data in Interpretation 1 are often known by the alternative names “Reference Data” and “Transactional Data”. In other words data describing things is Reference Data, and data describing events is Transactional Data. Within this section I’ll use those alternative names.

One problem with Reference Data is that multiple inconsistent versions may exist throughout the business. Data about each particular “thing” may be stored in multiple databases, each of which stores different attributes, uses different identifiers and is maintained by different people. Master Data Management initiatives set out to solve these problems by providing a single maintenance structure and a single place to make changes. Changes are synchronised between the Master Data Management tool and each consuming system so that the Master Data in all systems is consistent.

In practice, Master Data Management tools are hardly ever introduced in a single big bang and are rarely used to maintain all Reference Data. Instead, analysts and developers will begin by tackling a single type of Reference Data such as customer. As a consequence, at any given time there may be plenty of Reference Data that is not centrally maintained and for which inconsistent versions exist in different systems.

This situation introduces the need to distinguish between those items of Reference Data that are subject to Master Data Management and those that are not. To vendors of Master Data Management tools, only Reference Data that is stored within their tools is worthy of the title Master Data. In this interpretation, Master Data can be defined as “Reference Data with consensus”.

Consider for a minute, what term would you choose for the data that is the subject of Master Data Management? In the absence of a better suggestion I think I’d call it Master Data too!

MORAL: When discussing Master Data, agree whether you’re referring to all Reference Data or just the data within a Master Data Management system.

(Mis)interpretation 3: Master Data is data that needs to be highly shared

An aim of Master Data Management is to ensure that tools and processes are in place to maintain and share common definitions. The sharing of data is an important aspect of Master Data Management. After all, it is important that all the systems within a business have a consistent view of customers, products, locations, etc.

However it’s important to note that the need to share data is not limited to Master Data. For instance, any numbers that are published by an organisation in a written document or on a web page can certainly be said to be “highly shared”, but that doesn’t mean that they are Master Data – instead they are measures derived from event data.

Similarly, metadata that describes data structures and business rules may need to be highly shared. Again, this doesn’t mean they are Master Data.

MORAL: Master Data typically does need to be shared, but this isn’t the critical factor in determining whether data should be classified as Master Data.

(Mis)interpretation 4: Master Data is a term referring to the master copy of the data

It is quite common for people to believe that Master Data refers to the master copy of data (as in “master tape”).

The reasons why people can come to this conclusion are obvious. This view can be further reinforced by the fact that discussions about Master Data Management almost inevitably include phrases such as “golden copy of data” and “single version of the truth”.

This interpretation can cause particular confusion. It places emphasis on the place where data is stored rather than on the characteristics of data. Therefore it can lead people to believe that Master Data Management tools should be used to manage all data, rather than just the “critical nouns” of an organisation.

MORAL: Make sure you dispel the myth that Master Data refers to the master copy of data.

Conclusion

It is easy to find yourself talking at cross purposes when discussing Master Data. The people you talk to could have any one of several interpretations in their mind.

To avoid any confusion it’s important to make sure all parties are using the same interpretation. To summarise, here are some key points to re-emphasise my own interpretation of Master Data as set out in Interpretation 1:

- *Master Data is data describing the individual items about which events, figures and observations can be recorded. Such items include customers, products, locations and periods of time. Master Data forms an inventory of the “things” and “concepts” that are of interest.*
- *Master Data is distinct from Event Data and Metadata.*
- *The need to share data doesn’t make it Master Data.*
- *Master Data is not simply a term referring to the master copy of data.*

Chris Daniels is an Information Management Consultant at IPL, a leading UK IT services company specialising in the delivery of intelligent business solutions. He has over 10 years of experience in helping a wide range of high profile clients exploit the full potential of their information.

Chris has significant information management expertise gained across a variety of sectors, including manufacturing, finance, government, and telecoms. Chris has ongoing engagements in business intelligence, undertaking business analysis and the specification of information systems. He can be reached at chris.daniels@ipl.com.