Is all E&P Data Master Data?

Nigel Corbin, Schlumberger Information Solutions

*15th Petroleum Data Integration, Information & Data Management Conference*

Introduction

The Data Management Association (DAMA) Data Management Body of Knowledge (DMBoK) (DAMA, 2009) generally provides an excellent framework for data management, but has been criticized for lacking concrete examples. In some areas, such as Master Data Management, it seems to be self contradictory or incomplete. DMBoK claims that the role of Master Data is to provide context for transactional data. However, some E&P business functions do not work that way. For example, petrophysicists would regard well logs as part of their working data, but other specialists might consider them master data.

The E&P business depends on engineering or measured data. An interpretation process builds a living model that changes over time, and is very much a scientific hypothesis. These models are used to predict, to plan and assess activities, and are then modified by the new data acquired through those activities. This is a different approach to many businesses where adding one customer or product has no impact on others.

Overview – Data according to DAMA DMBoK

DAMA defines 10 functions, central to which is Data Governance.
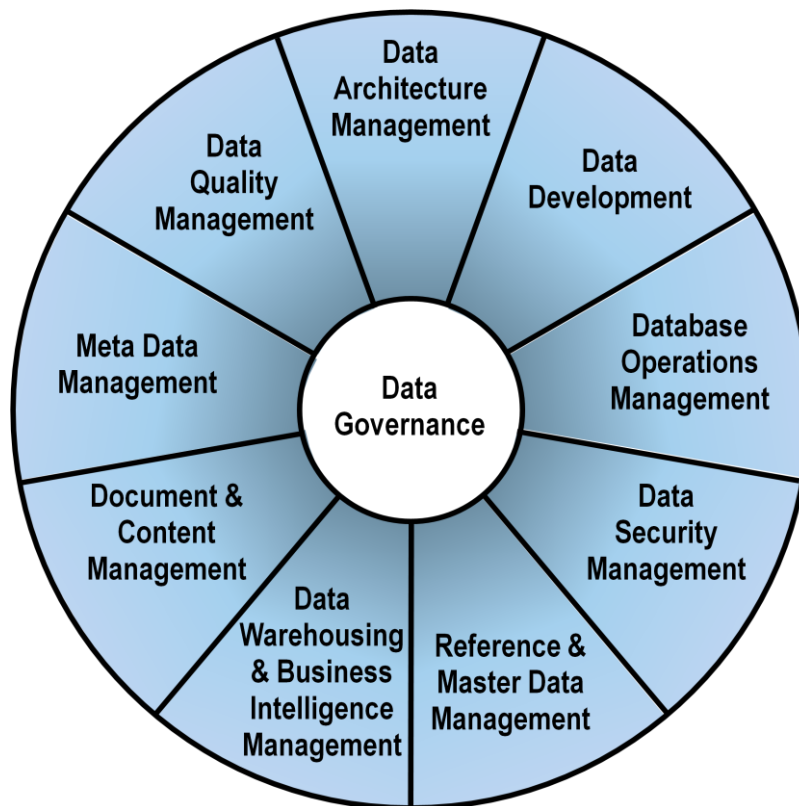
Some of these functions relate to specific kinds of data, and some (such as Data Security) relate generically to all data. This paper will focus on Master Data, and the definitions of it given by DMBoK and others.

Schlumberger

We will review several ways to classify data, and see how these align with DAMA and with E&P industry practice.

All ways of classifying data depend on use of the data, but also on attributes such as:

- Frequency of changes
- Volume
- Life cycle time, and/or relevance, timeliness.
- Data Quality

The other approach, which DAMA follows, can be identified from the functions defined above.

The cost of managing the data within this framework must be balanced with the business value of the data. Data required locally within a process is relatively uncontrolled, and provides value within that process. As data is made available for a wider audience or longer time-frame, the importance of data management increases.

Metadata

Data about data –here we include most business process definitions, requirements, standards and logical or physical data models. The metadata will tend to be proprietary and stored in a repository that provides context for all other data.

Reference Data

Value domain definitions, typically defined by multiple standards and required for interchange.

DAMA Definition – Reference Data Management

Control over defined domain values (also known as vocabularies) including control over standardized terms, code values, and other unique identifiers, business definitions for each value, business relationships within and across domain value lists, and the consistent, shared use of accurate, timely and relevant reference data values to classify and categorize data.

There tend to be many competing standards, and any business must maintain either one common internal standard, or a large number of different contradictory standards with mappings between them.

These typically are externally defined by bodies such as the International Organization for Standards (ISO) and have a very slow change cycle. The volume of data, if any, is small. (The standard may of course be stated in terms of a reference rule set, rather than a list of values – date formats being one such). There may be several applicable standards within any reference domain.

Country codes provide a familiar example. There are at least two standard code sets, ISO 3166 being one. (It even defines more than one code set – 2 or 3 character, or numeric).

Example:
- British road vehicles MUST display **"GB"** code abroad
 - See http://www.direct.gov.**uk/**

Master Data
DAMA definition – not very controversial, is:

> **DAMA Definition - Master Data Management**
>
> Control over master data values to enable consistent, shared, contextual use across systems, of the most accurate, timely, and relevant version of the truth about essential business entities.

More commonly, DMBoK refers to Master Data as providing the context for "transactional data". That term is both obvious and yet undefined in either the DMBoK or DAMA Dictionary (DAMA, 2008).

The implication in DMBoK is roughly:
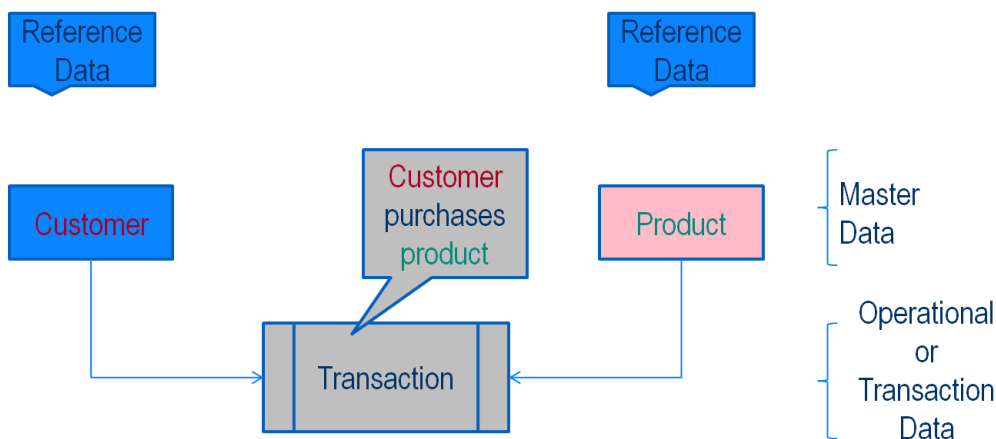
This works really well in the context of a business that is selling goods and services. Reference data consists of the rules defining credit card numbers, ZIP codes etc, and product descriptions. The transactions comprise the bulk of the data being held, and is held in what might be called an operational datastore. Information extracted from that store is used – via data warehousing, analytics, - to extract business intelligence that is fed back to business decisions about suppliers and products.

Amazon, for example, is one of many online retailers that allows customers to provide feedback and adds it into the product descriptions as "Reviews" or "Ratings".

The illustration above can be generalized –
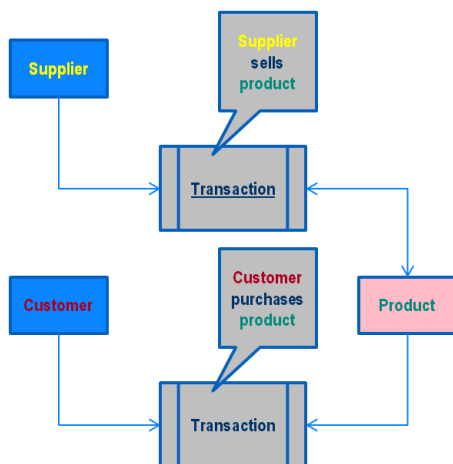
A typical business sells goods and services (in some combination) but also purchases them. We see product information on Amazon – clearly, this must match the information provided by Amazons' suppliers.

A manufacturer would be providing Amazon with product information – they have much more than that themselves. The original product database in the manufacturer contains details of their suppliers, recipes, designs, bills of material and so forth. Much is commercially sensitive.

For some companies, the supplier side transactions will be far more voluminous than the customer side. (Consider Boeing, as an example…. How many components does a 747 have? How much quality control information is attached to each?) We could discuss supply chain management or customer relationship management, and both are very important in context. In all cases the customer side data feeds back to changes in supply side management.

Data Warehouse + Business Intelligence
In fact, we have a loop of "Continuous Improvement".
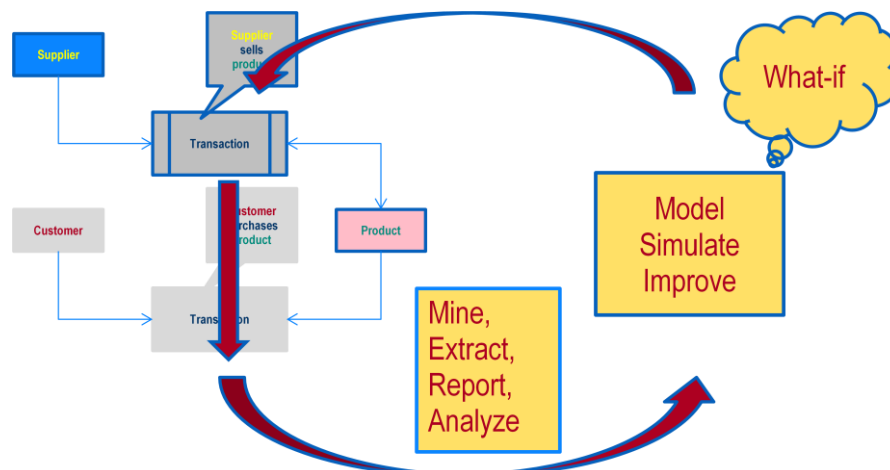
This is (actually the next function in their list) Data Warehousing in DAMA terms, the what-if process that improves products and also plans the production levels, generally optimizing business performance by prediction. DAMA describes a very generic data warehouse in transactional terms, in considerable detail.

This what-if data is not Master Data – we have no "one version of the truth" here. Everything here is sourced from the Master Data, AND newly acquired data, PLUS the results of predictive efforts by expert "interpreters" or equivalent, such as software modeling and simulation. Geologically, this is our model evolving in time. In scientific terms, this model represents a hypotheses rather than proven facts and theorems. Each business decision results in more data acquisition, which is definitely to be classed as Master Data. Feeding that data into the model changes it, and simultaneously feeds back data quality issues. Most of the transactions are with suppliers – not customers.

The downstream oil business clearly is a product/customer model, exactly as DAMA DMBoK discusses. It will have its own issues for product quality, but is at the downstream end a relatively simple sales business. But what about the upstream E&P business? It produces product, or at least raw material, from facilities in many and varied locations using complex plant and equipment, all of which have complex measurement systems in place. We have no great volume of transactions, all nearly identical. All of the E&P technical data is Master Data, perhaps? Certainly the data about wells, facilities, activities, and all other key entities is, by most definitions of Master Data.

DAMA DMBoK probably over-emphasizes "transactional" when another word, such as "operational" would suit better. The definition given above is excellent, but the surrounding discussion is flawed and incomplete.

Operational Data

I will define Operational Data:

Definition – Operational Data
Data acquired or generated in past and present operations. Master Data provides the context within which it was acquired or generated.

This is not a DAMA DMBoK definition. This data has a number of interesting characteristics:

- It is often measured in some activity. As such, it contains measurement errors and inaccuracies inherently.
- It is definitely *not* required to be transactional.
- It has many kinds – corresponding to different activities that acquire it. Planned well paths, real-time telemetry data from all kinds of equipment, well logs, seismic surveys, satellite imagery, simulated reservoir properties and more.

Example 1:

A well is instrumented with a sensor that measures and records temperature and feeds it via some technology such as a SCADA system. We probably have a tuple (timestamp, tag, value) like (20101020 08:45, 56873, 156.1). We need to check our master data, metadata, and reference data to translate this to something useful:

Sensor 148/T recorded 156.1F on 2010/02/02 08:45

"Tag ID 56873, or 148/T, is the downhole temperature sensor, placed1500m MD in the A8 well, recorded every minute in degrees F".

Policy and procedure determine whether this is correct, accepted, and thus itself becomes Master Data. Data conditioning may be required and analytic tools used to identify trends and problems. We are in this case more concerned with changes and trends.

Example 2:

Simulation techniques are used to model and predict the behavior of a complex system in many technical domains. Weather forecasting is one example very widely known. In the E&P domain, reservoir simulation is another example.

- Build a model using all currently available data. (best, current => Master Data)
- Add possible future data (plans, test cases, scenarios + parameters)
  - Simulate to see what outcome is attained.
- There may be many parameter values to choose, and the simulation is repeated over a range of conditions. Each simulation generates a large amount of data.

Again, policy and procedure determine the extent to which the simulation results are stored as Master Data; typically the optimal cases are elaborated as plans for new activities and facilities.

The activities that result from the plans themselves generate measured data that adds to the overall model and in turn becomes master data. Everything that DAMA has to say about Master Data Management is true, insofar as quality control is rigorously enforced as data crosses from operational to master store.
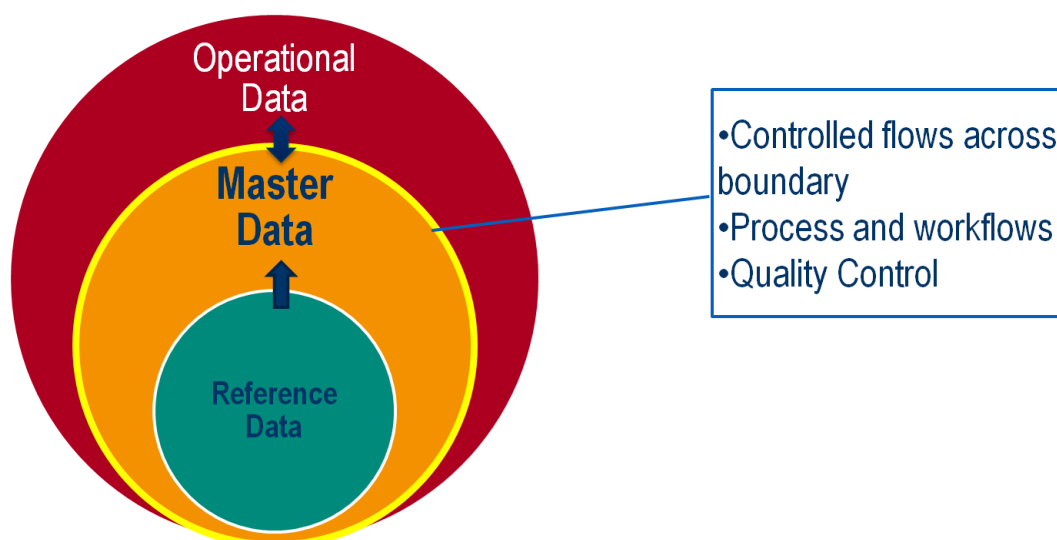
Conclusion

Most of what DAMA says about Data Operations is true, with the proviso that the E&P data may not be transactional. All measured data is operational data, until it passes the required quality/consistency checks and becomes Master data. This process may be automated, or require expert intervention. Most measured data either itself becomes master, or leads to changes in master data via some defined process. At least some operational data is *automatically* promoted to Master, even when raw and unchecked. This may be the case for one-off measurements that, for either economic or technical reasons, cannot be repeated. (Well logs might be an example, in the E&P arena). The extreme example would be data required for audit, regulatory or forensic purposes – aircraft flight-data recorders as an example. The vast bulk of operational data is either "what-if" scenarios, or raw data that is not "one-off" or has a very short operational life cycle. This data is used in processes that lead to operational decisions that themselves update Master data – produce plans for new activities going forward.

Bibliography

DAMA. (2008). *DAMA Dictionary of Data Management.* (M. Mosley, Ed.) New Jersey: Technics Publications LLC.

DAMA. (2009). *The DAMA Guide to the Data Management Body of Knowledge.* (M. Mosley, Ed.) Technics Publications LLC.