Reference Data vs. Master Data Management

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'
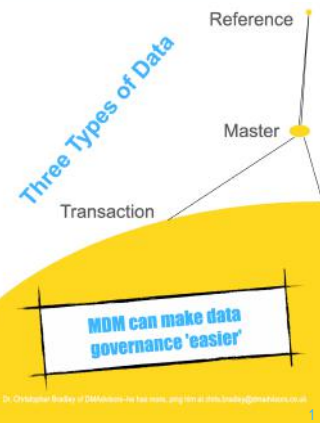
Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Institute for Defense Analyses (ida.org)
- DAMA International (dama.org)
- MIT CDO Society (iscdo.org)
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices worldwide
- 13 books and dozens of articles
- Multi-year immersions
    – US DoD (DISA/Army/Marines/DLA)
    – Nokia
    – Deutsche Bank
    – Wells Fargo
    – Walmart
    – HUD …
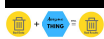
$1,500,000,000.00 USD

Program Overview

- Data Management Overview

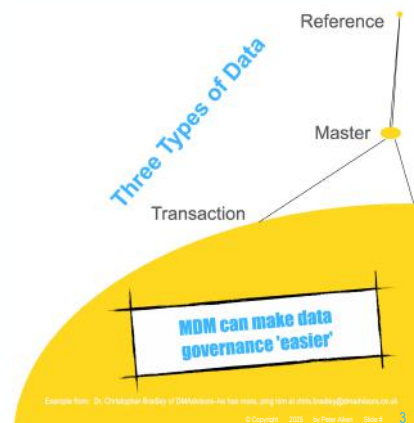- What is Reference and MDM?

- Why is Reference and MDM important?

- Reference & MDM Building Blocks
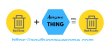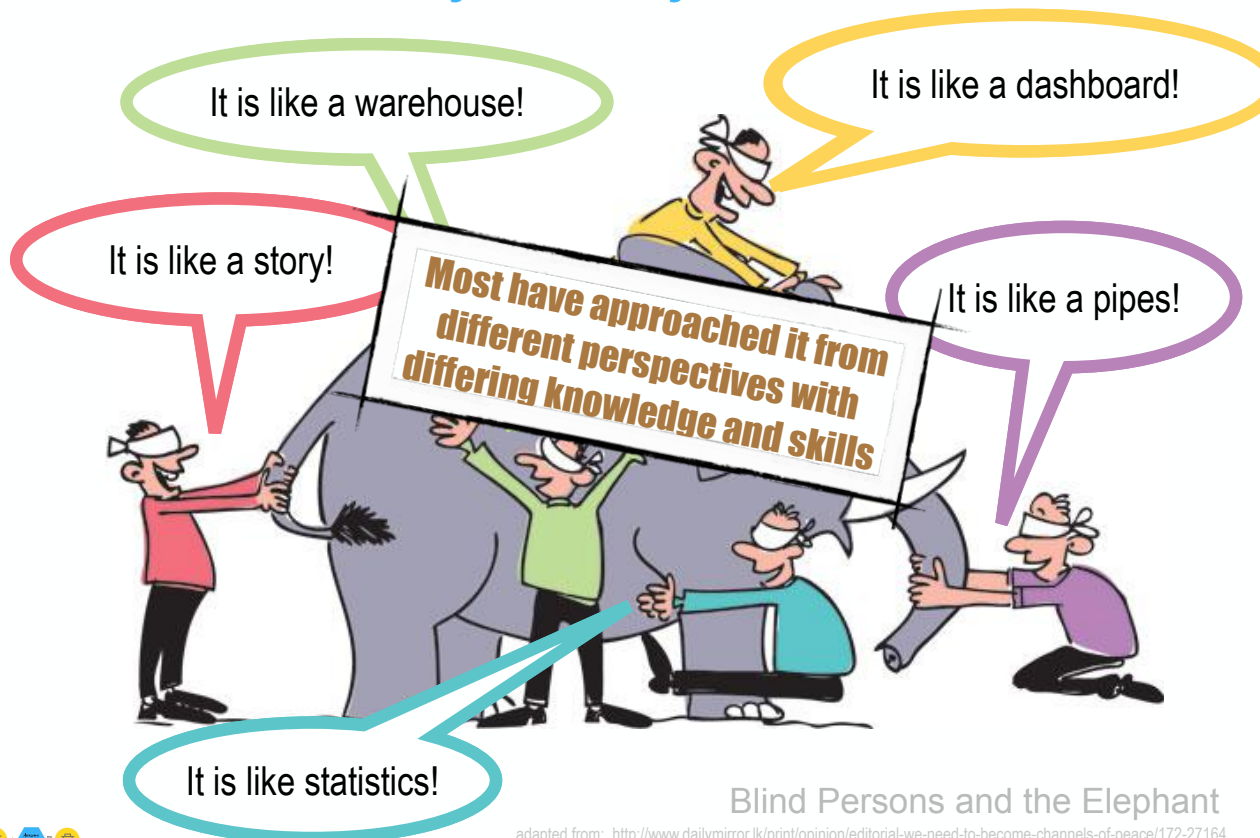
- Guiding Principles & Best Practices

- Take Aways, References & Q&A

Transforming Data with AI: Essential Reference and Master Data Management Strategies

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'

Data is not broadly or widely understood

It is like a warehouse!

It is like a dashboard!

It is like a story!

Most have approached it from different perspectives with differing knowledge and skills

It is like a pipes!

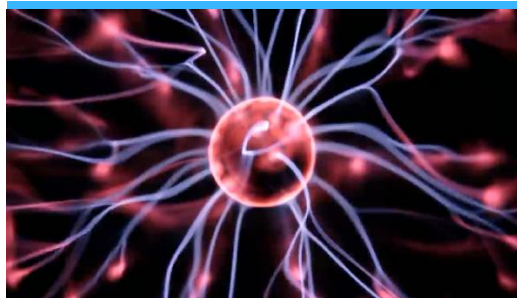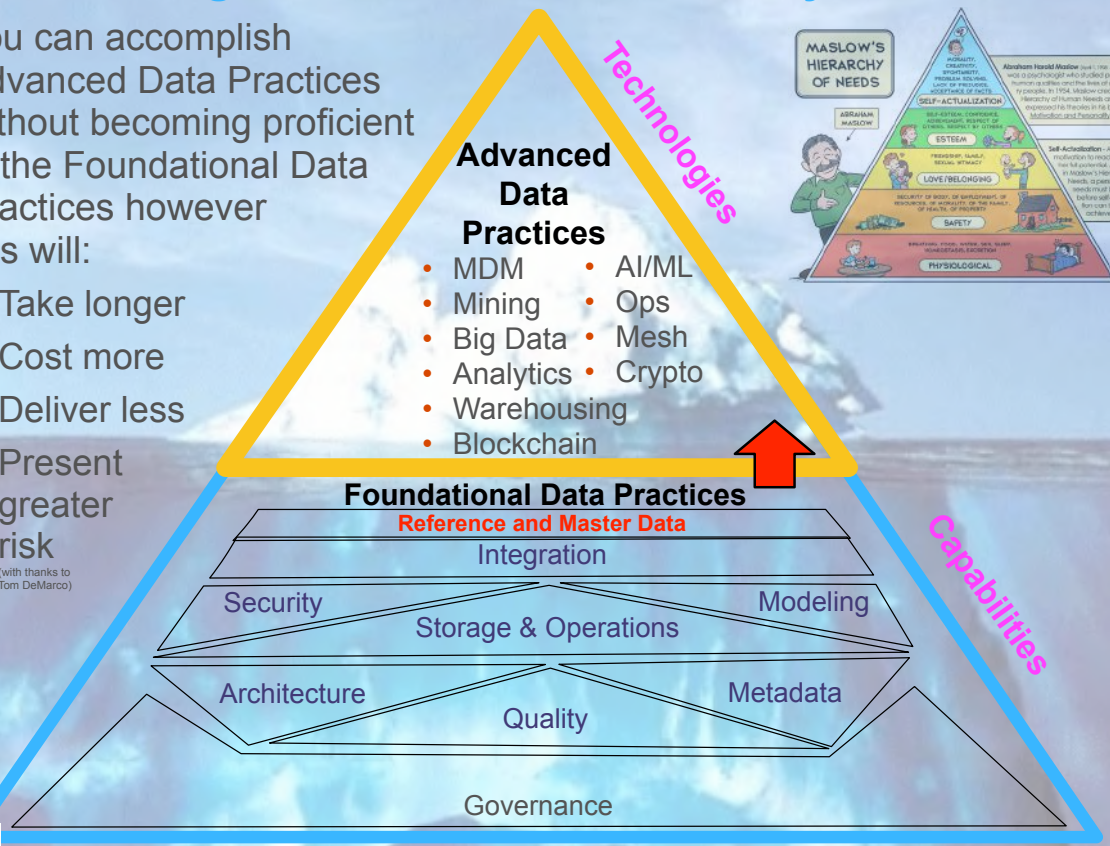It is like statistics!

Blind Persons and the Elephant

adapted from: http://www.dailymirror.lk/print/opinion/editorial-we-need-to-become-channels-of-peace/172-27164

Data Management Practices Hierarchy

You can accomplish Advanced Data Practices without becoming proficient in the Foundational Data Practices however this will:

- Take longer
- Cost more
- Deliver less
- Present greater risk

(with thanks to Tom DeMarco)

Technologies

Advanced Data Practices
- MDM
- Mining
- Big Data
- Analytics
- Warehousing
- Blockchain
- AI/ML
- Ops
- Mesh
- Crypto

Foundational Data Practices

Reference and Master Data

Integration

Security — Storage & Operations — Modeling

Architecture — Quality — Metadata

Governance

Capabilities

Unrefined data management definition

Sources

Data Management

Uses

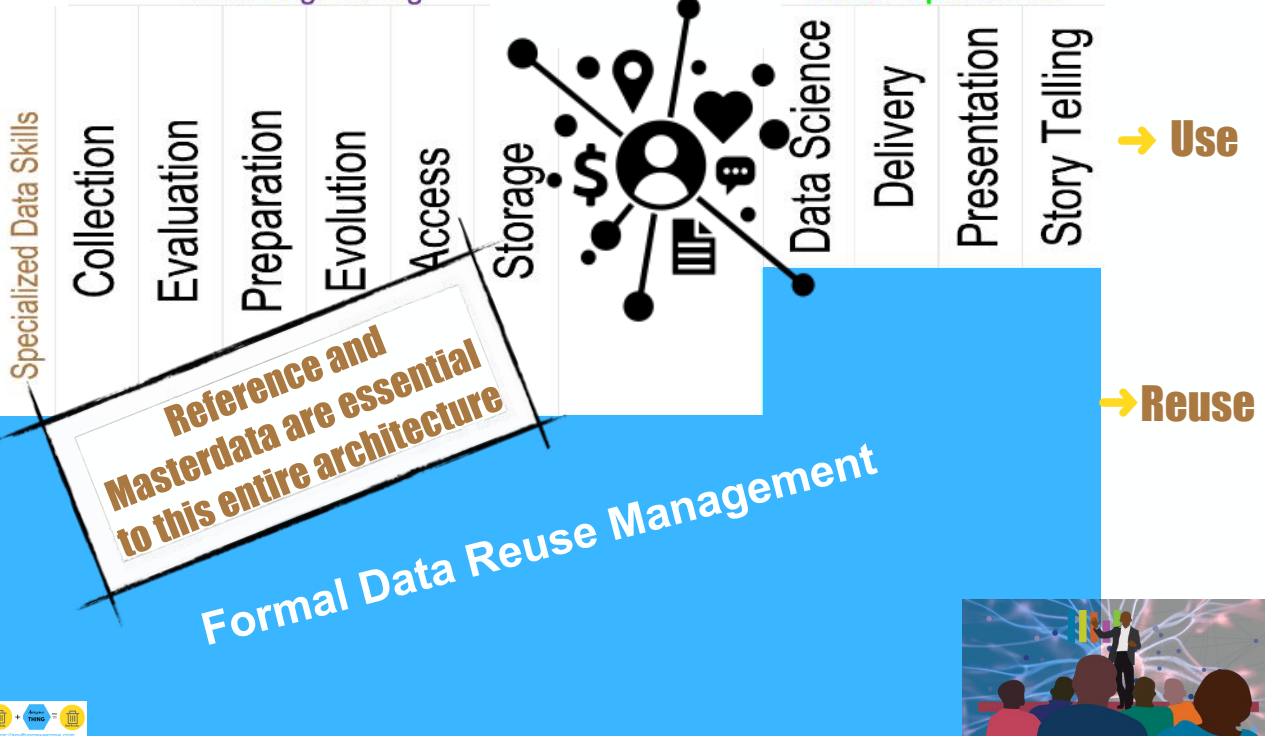## More refined data management definition

Sources ➡ Data Management ➡

Reuse

Better still data management definition

Sources

Data Engineering

Data Exploitation

Specialized Data Skills · Collection · Evaluation · Preparation · Evolution · Access · Storage

Data Science · Delivery · Presentation · Story Telling

➡ Use

➡ Reuse

Reference and Masterdata are essential to this entire architecture

Formal Data Reuse Management

Program Overview

- Data Management Overview
- What is Reference and MDM?
- Why is Reference and MDM important?
- Reference & MDM Building Blocks
- Guiding Principles & Best Practices
- Take Aways, References & Q&A

Transforming Data with AI: Essential Reference and Master Data Management Strategies

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'

Program Overview

- Data Management Overview
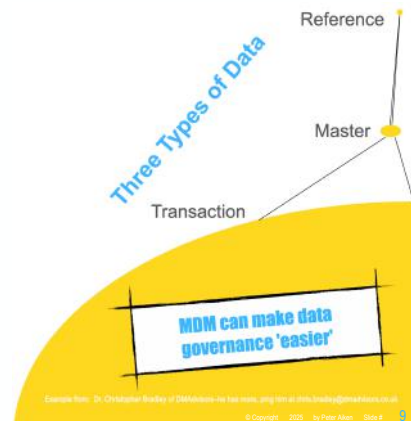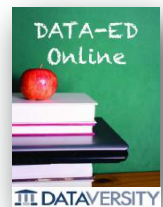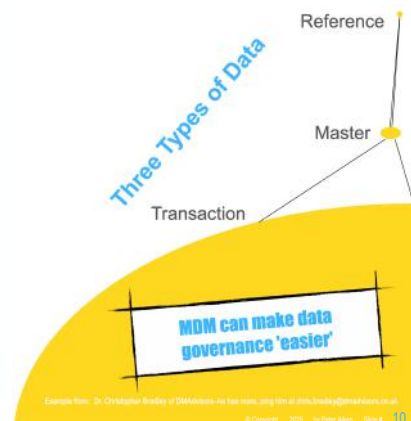- What is Reference and MDM?
- Why is Reference and MDM important?
- Reference & MDM Building Blocks
- Guiding Principles & Best Practices
- Take Aways, References & Q&A

Transforming Data with AI: Essential Reference and Master Data Management Strategies

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'

Data Preparation Tools & Vendor Hype

- CIOs/CDOs feel pressure

- Vendor/project promise auditing

- No understanding of hype cycle

Audit

HYPE

Who wrote this … ?
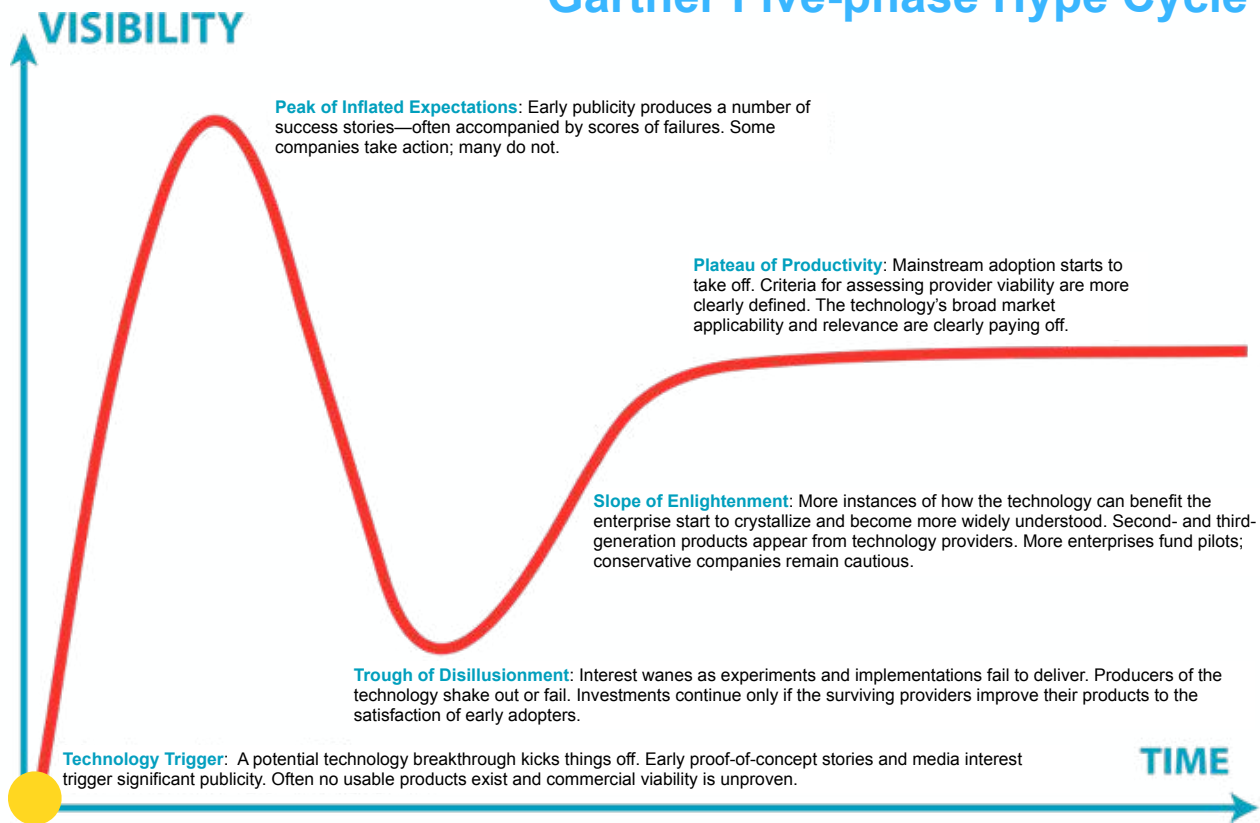
- *In considering any new subject,*

- *there is frequently a tendency first to overrate what we find to be already interesting or remarkable, and*

- *secondly - by a sort of natural reaction - to undervalue the true state of the case.*

  – Lady Augusta Ada King, (1815 – 1852) Countess of Lovelace

  – (aka) Ada Lovelace, daughter of Lord Byron

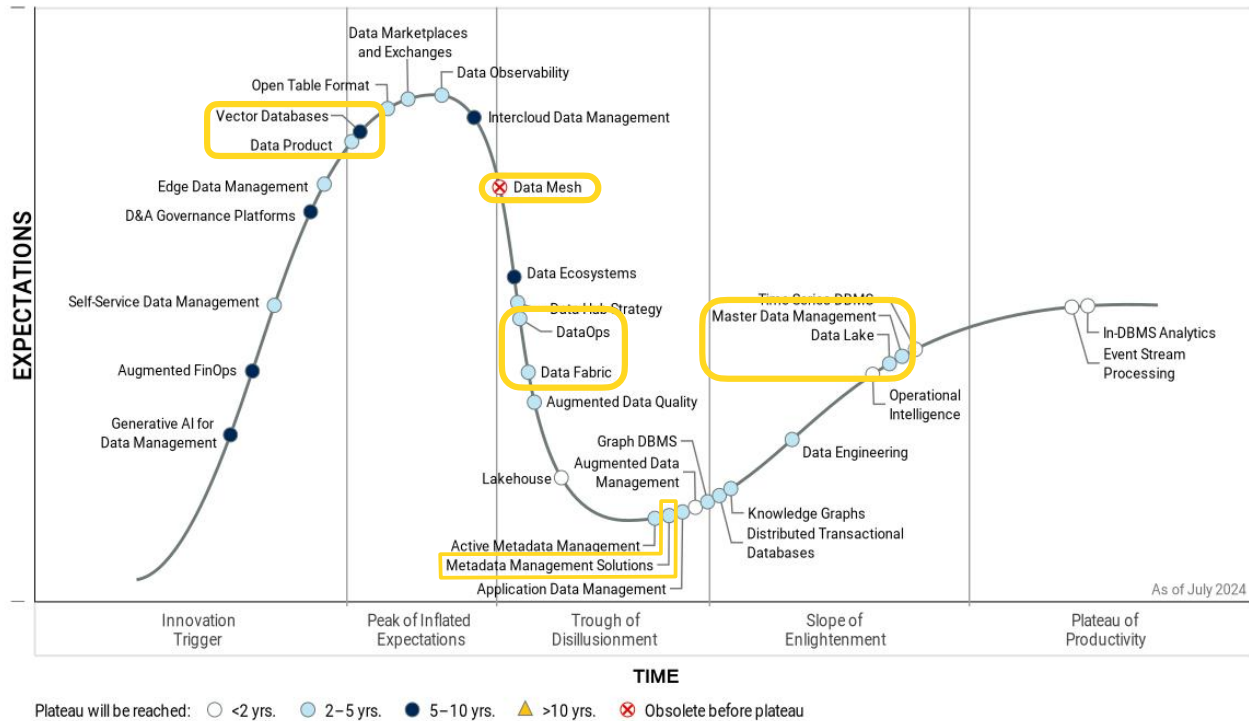 – Publisher of the first computing program

Gartner Five-phase Hype Cycle

VISIBILITY

Peak of Inflated Expectations: Early publicity produces a number of success stories—often accompanied by scores of failures. Some companies take action; many do not.

Plateau of Productivity: Mainstream adoption starts to take off. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying off.

Slope of Enlightenment: More instances of how the technology can benefit the enterprise start to crystallize and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises fund pilots; conservative companies remain cautious.

Trough of Disillusionment: Interest wanes as experiments and implementations fail to deliver. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters.

Technology Trigger: A potential technology breakthrough kicks things off. Early proof-of-concept stories and media interest trigger significant publicity. Often no usable products exist and commercial viability is unproven.

TIME

http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

---

Gartner Hype Cycle for Data Management 2024

EXPECTATIONS

- Data Marketplaces and Exchanges
- Data Observability
- Open Table Format
- Vector Databases
- Intercloud Data Management
- Data Product
- Edge Data Management
- D&A Governance Platforms
- Data Mesh (Obsolete before plateau)
- Self-Service Data Management
- Data Ecosystems
- Data Hub Strategy
- Time Series DBMS
- Master Data Management
- DataOps
- Data Lake
- In-DBMS Analytics
- Event Stream Processing
- Augmented FinOps
- Data Fabric
- Augmented Data Quality
- Operational Intelligence
- Generative AI for Data Management
- Graph DBMS
- Data Engineering
- Augmented Data Management
- Lakehouse
- Knowledge Graphs
- Distributed Transactional Databases
- Active Metadata Management
- Metadata Management Solutions
- Application Data Management

As of July 2024

| Innovation Trigger | Peak of Inflated Expectations | Trough of Disillusionment | Slope of Enlightenment | Plateau of Productivity |

TIME

Plateau will be reached: ○ <2 yrs. ◔ 2–5 yrs. ● 5–10 yrs. △ >10 yrs. ⊗ Obsolete before plateau

Gartner

from *The DAMA Guide to the Data Management Body of Knowledge* 2E © 2017 by DAMA International

---

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.
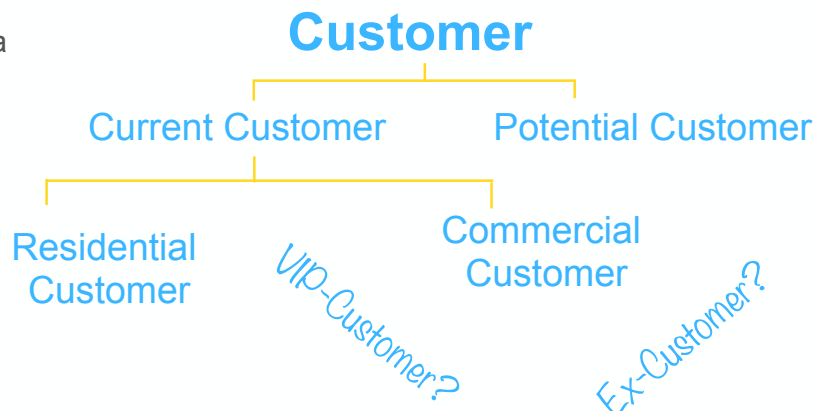
Reference Data

- Data used to classify to categorize other data, the value domain

- Order status: new, in progress, closed, cancelled

- Two-letter USPS state code abbreviations (VA)

- Reference Data Sets

US	United States
GB (not UK)	United Kingdom

from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International

---

Master Data

- Data about business entities providing context for transactions but not limited to pre-defined values

- Business rules dictate format and allowable ranges

  – Parties (individuals, organizations, customers, citizens, patients, vendors, supplies, business partners, competitors, employees, students)

  – Locations, products, financial structures

- Provide context for transactions

- From the term "Master File"
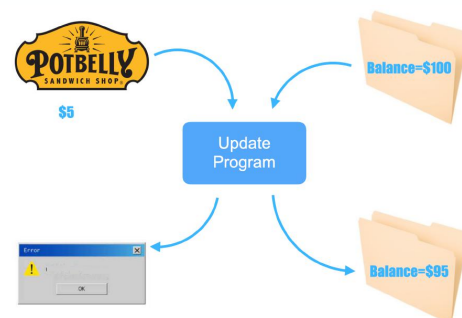
from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International

Example Transaction Processing System

$5

Balance=$100

Update Program

Error — ⚠ Something is wrong! [OK]

Balance=$95
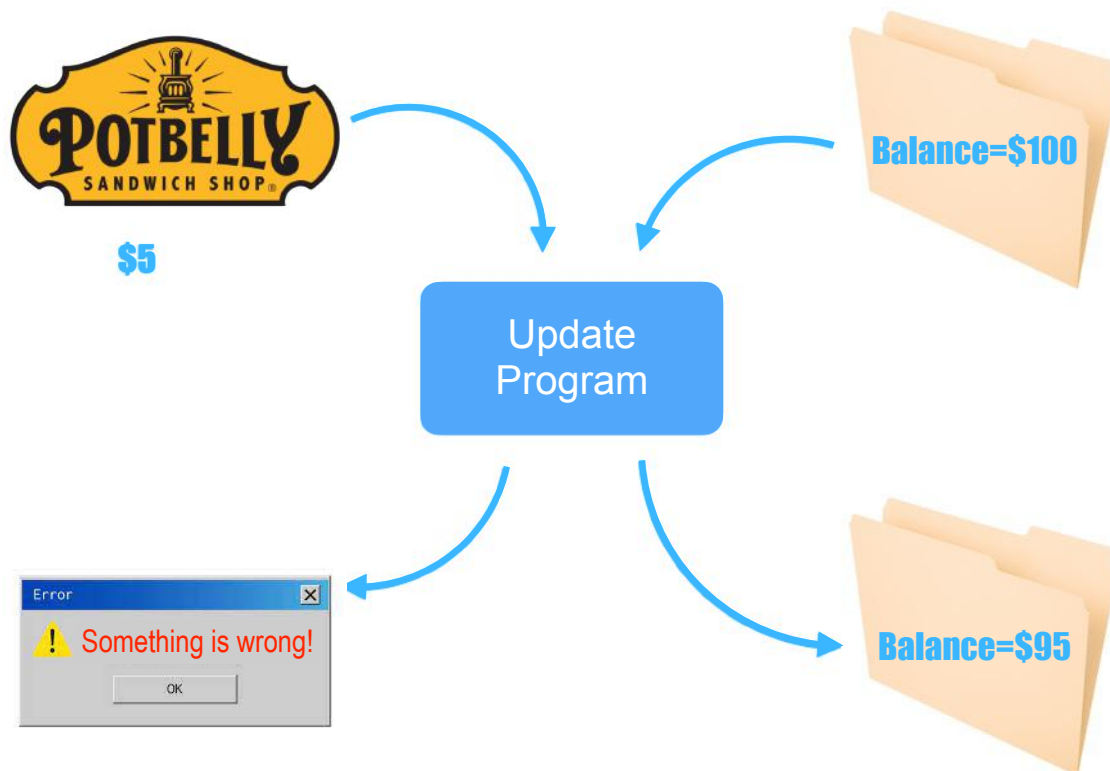
Reference Data versus Master Data

- Reference Data:
  - Control over defined domain values (vocabularies) for standardized terms, code values, and other unique identifiers
  - The fact that we maintain these 9 specific gender codes

- Master Data:
  - Control over master data values to enable consistent, shared, contextual use across systems
  - The "golden" source of the gender of your customer "Pat"

FBI & Canadian Social Security Gender Codes

1. Male
2. Female
3. Formerly male now female
4. Formerly female now male
5. Uncertain
6. Won't tell
7. Doesn't know
8. Male soon to be female
9. Female soon to be male

Both provide the context for transaction data

Definitions

- Planning, implementation and control activities to ensure consistency with a "golden" version of contextual data values

- … as opposed to **m**obile **d**evice **m**anagement

- Gartner holds that MDM is a **discipline** or **strategy**

  – "… where the business and the IT organization
  work together to ensure the uniformity, accuracy,
  semantic persistence, stewardship and accountability
  of the enterprise's official, shared master data."

- Sold as technology-based solution

- Official, consistent set of identifiers - examples of these core entities include:

  – **Parties** (customers, prospects, people, citizens, employees, vendors, suppliers, trading partners, individuals, organizations, citizens, patients, vendors, supplies, business partners, competitors, students, products, financial structures *LEI*)

  – **Places** (locations, offices, regional alignments, geographies)

  – **Things** (accounts, assets, policies, products, services)

Hannah Fry - Why you should be polite to your AI
(in 2 minutes)

Why you should be polite to AI
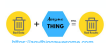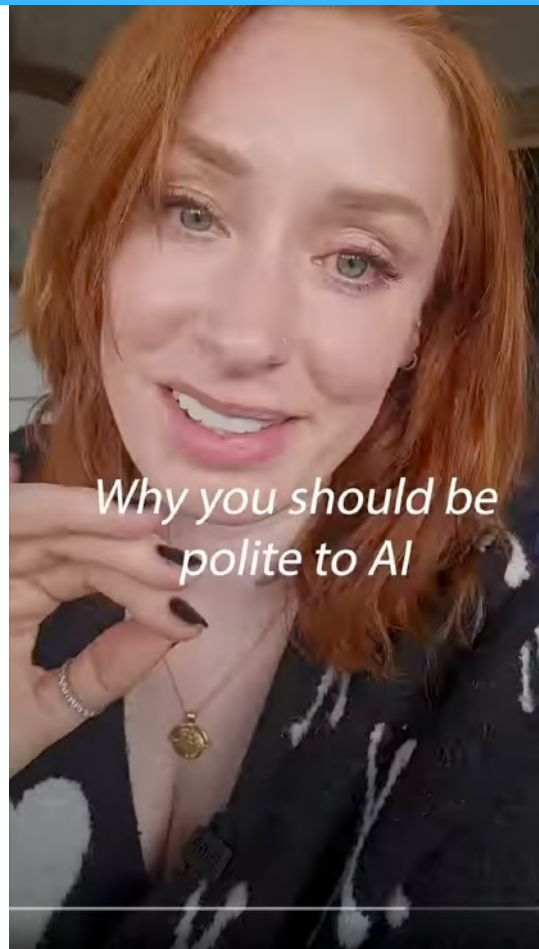
How to tell what is Reference/Master Data?

Determining whether a specific data item is best managed as a reference or master data item is a crucial decision in data management. It impacts data quality, consistency, and how information is used across an organization. Here's a breakdown of how to approach this:

1. Understanding the Core Concepts

Master Data:

- Master data represents the core, consistent, and authoritative data entities that are used across multiple systems and business processes.
- It typically includes key business entities like customers, products, locations, suppliers, etc.
- **Master** data aims to provide a single source of truth for these critical entities.
- **Reference** Data:
- Reference data is used to classify or categorize other data.
- It's often used for lookups, validation, or to control the values that are allowed for a data field.
- Examples include:
- Codes (e.g., country codes, currency codes)
- Categories (e.g., product categories)
- Units of measure
- Status values (e.g., order status)

2. Key Questions to Guide Your Decision

- To determine whether a data item should be managed as reference or master data, ask these key questions:

How widely is this data used across the organization?

- **Master Data**: If the data is used by many different systems, departments, and business processes, and it's critical for consistent reporting and decision-making, it's likely master data.
- **Reference Data**: If the data is primarily used for classification, categorization, or validation within specific contexts, it might be reference data.

How frequently does the data change, and how important is it to track those changes?

- **Master Data**: Master data changes relatively infrequently, and when it does, it's essential to manage and propagate those changes consistently. There's often a need for audit trails and versioning.
- **Reference Data**: Reference data may also change, but the changes may be less frequent or less critical to track at a granular level.
- What is the level of granularity and complexity of the data?
- **Master Data**: Master data often involves complex relationships and attributes. For example, a "Customer" master data record might include contact information, billing addresses, shipping addresses, account details, etc.

- **Reference Data**: Reference data is often simpler and more atomic, such as a single code or a category name.

Is this data a key identifier or a descriptive attribute?

- **Master Data**: Master data often includes key identifiers (e.g., customer ID, product ID) that are used to link data across systems.
- **Reference** Data: Reference data is often used as a descriptive attribute to provide context or categorization (e.g., a product category is a descriptive attribute of a product).

What are the data quality requirements?

- **Master Data**: Master data requires high levels of data quality, accuracy, and consistency. It's essential to have processes for data cleansing, validation, and deduplication.
- **Reference Data**: Reference data also needs to be accurate, but the quality requirements might be less stringent than for master data in some cases.

3. Decision-Making Framework

- Here's a simplified framework to help guide your decision:
- **Identify the Data Item**: Clearly define the specific data item you're evaluating (e.g., "Customer," "Product Category," "Shipping Code").
- **Answer the Key Questions**: Carefully consider the questions outlined above in the context of your organization's needs.

Apply the Criteria:

- If the data item is widely used, critical for consistency, complex, and requires high data quality, classify it as Master Data.
- If the data item is primarily used for classification, categorization, or validation, and is relatively simple, classify it as Reference Data.
- **Document Your Decision**: Document your decision and the reasons behind it. This will help ensure consistency and provide context for future data management efforts.

Examples

- **Customer**: This is typically master data. It's used across sales, marketing, service, and other departments. It's complex, and data quality is critical.
- **Product Category**: This is typically reference data. It's used to classify products. It's relatively simple, and while accuracy is important, the complexity and usage are different from "Product" master data.
- **Country Code**: This is clearly reference data. It's a simple code used for validation and standardization of country information.

By carefully considering these factors, you can make informed decisions about how to manage your data effectively, ensuring data quality, consistency, and value across your organization.
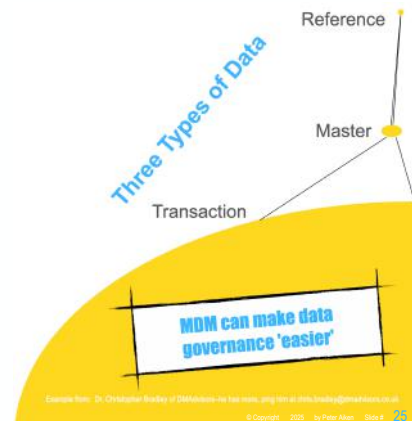
Program Overview

- Data Management Overview

- What is Reference and MDM?

- Why is Reference and MDM important?

- Reference & MDM Building Blocks

- Guiding Principles & Best Practices

- Take Aways, References & Q&A

Transforming Data with AI: Essential Reference and Master Data Management Strategies

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'

Example from: Dr. Christopher Bradley of DMAdvisors–he had more, ping him at chris.bradley@dmadvisors.co.uk

© Copyright 2020 by Peter Aiken Slide 8   25

Three Types of Data

- Reference
  » Countries where we do business?
  » Types of accounts available?
  » Controlled vocabulary items

  – Controls accessible data values

- Master
  » Are you a member of our premium club?
  » Authorizing uses/users?
  » Common/standard data structures

  – Controls access to system capabilities

- Transaction
  » $5
  » Authorized
  » Like

  – Instances of values

MDM can make data governance 'easier'

Example from: Dr. Christopher Bradley of DMAdvisors–he has more, ping him at chris.bradley@dmadvisors.co.uk

26

+ 1 Year

- Confusion as to the system's value
  - Users lack confidence
  - Business did not know how to use "*the MDM*"
- General agreement
  - Restart the effort
- "Root cause" analysis
  - Consensus
  - Poor quality data
  - Inadequate training
- Response
  - *Get data quality-ing!*
- Inexperienced
  - Immature data quality practices
  - Tool/technological focus
  - Purchased a data quality tool

Cleans *AND* Integrates

Interdependencies

Data Governance

makes the case and is responsible for

MD capabilities constrain governance effectiveness

Data Quality

Master Data
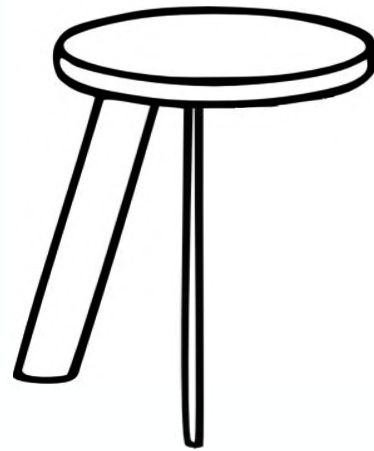
is a necessary but insufficient prerequisite to success
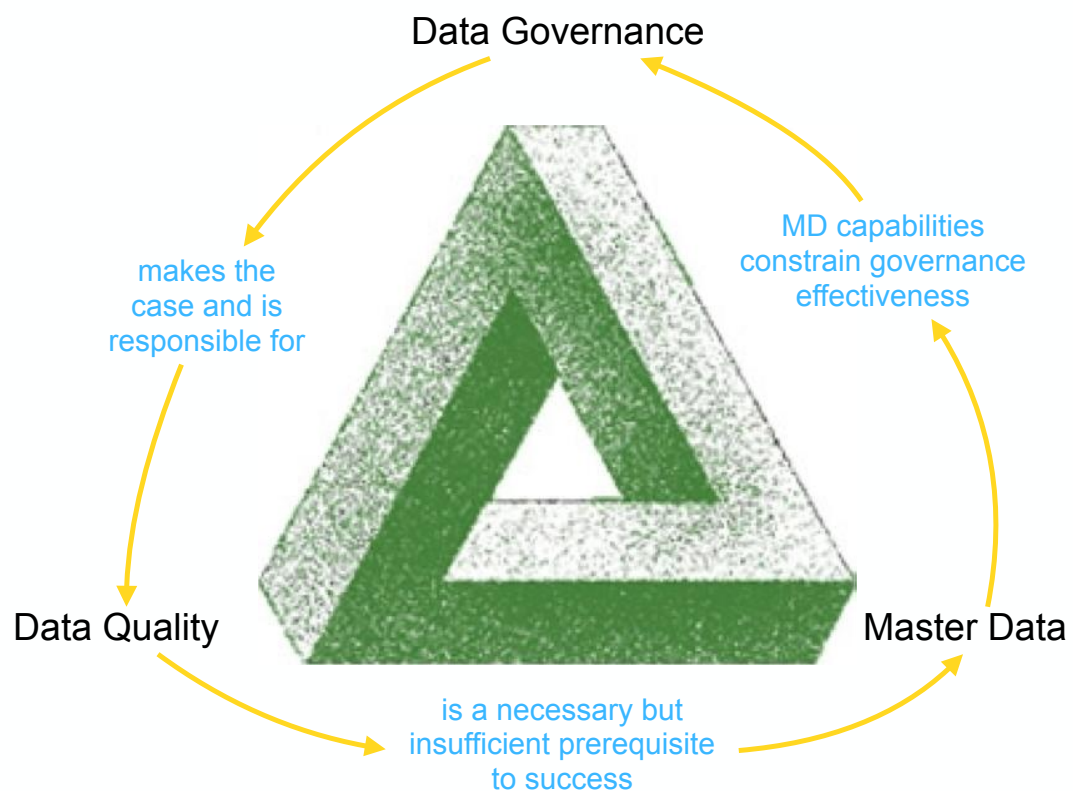
A realistic way to begin practicing MDM

- Select 3 data management practice areas (for example)

  – Reference and Master Data Management

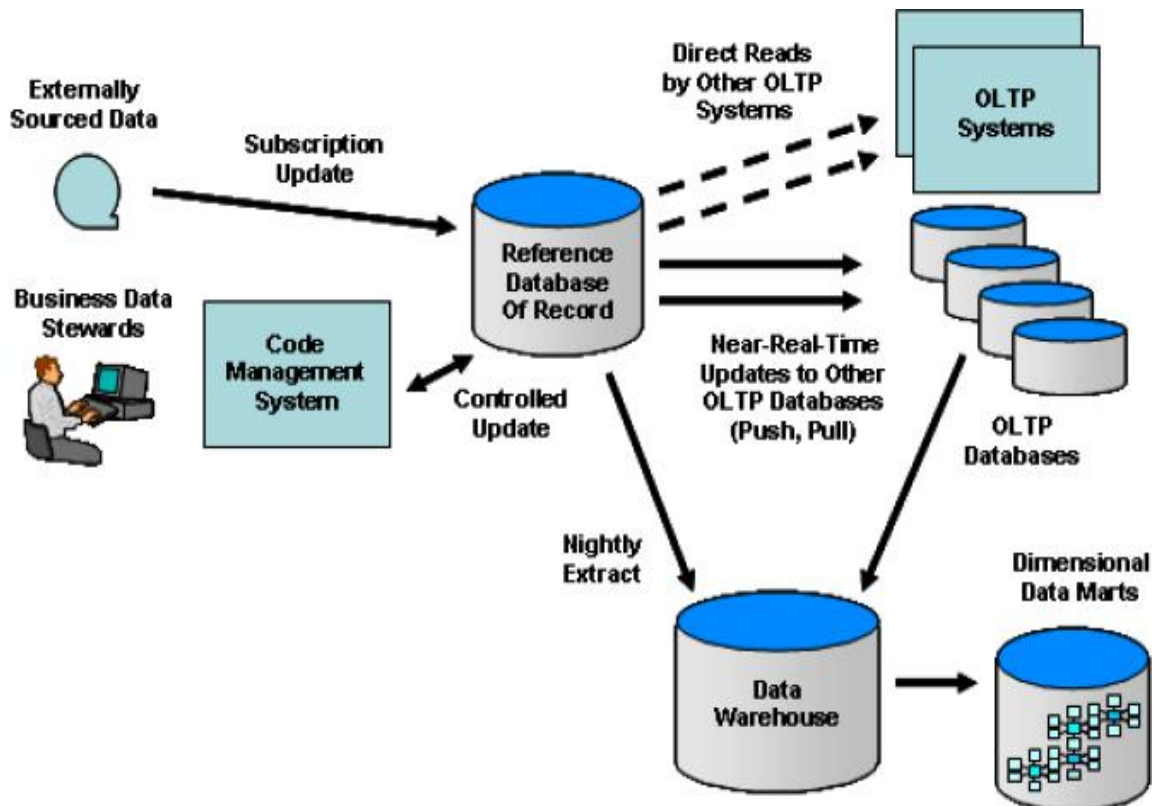 – Data Quality Management

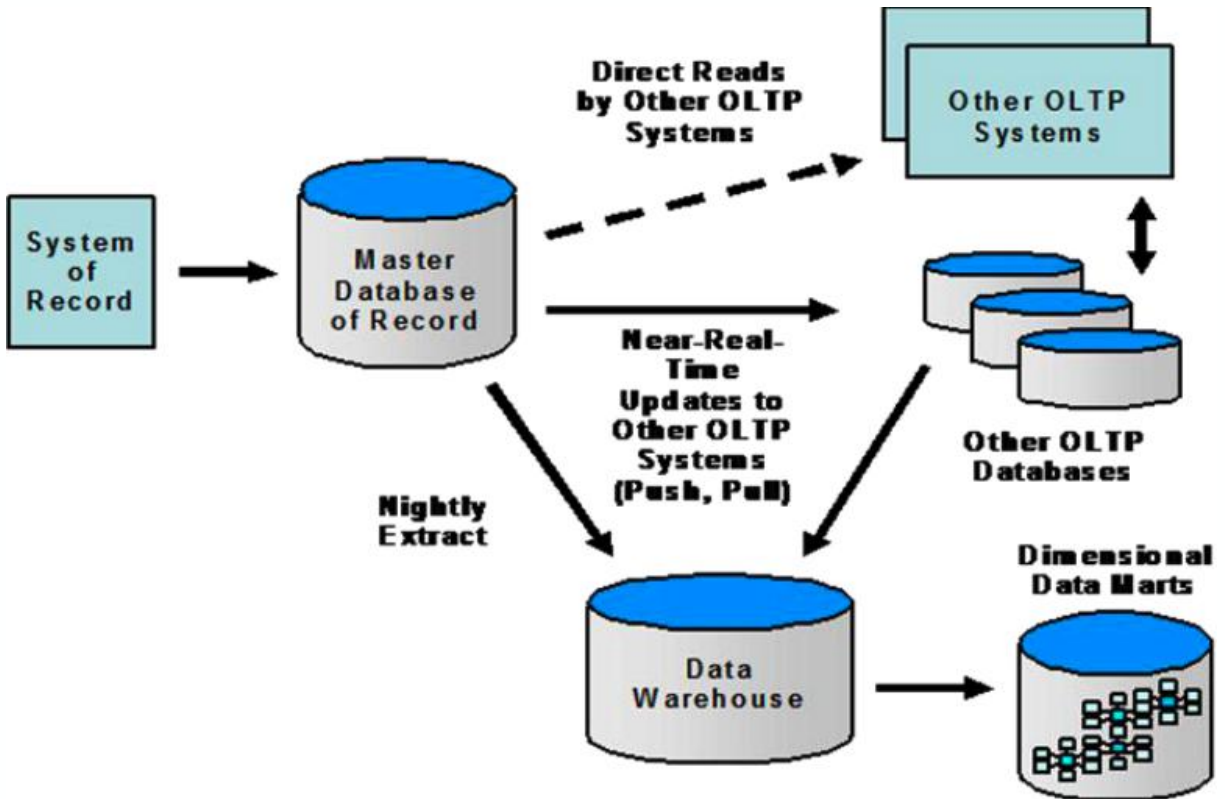 – Data Governance
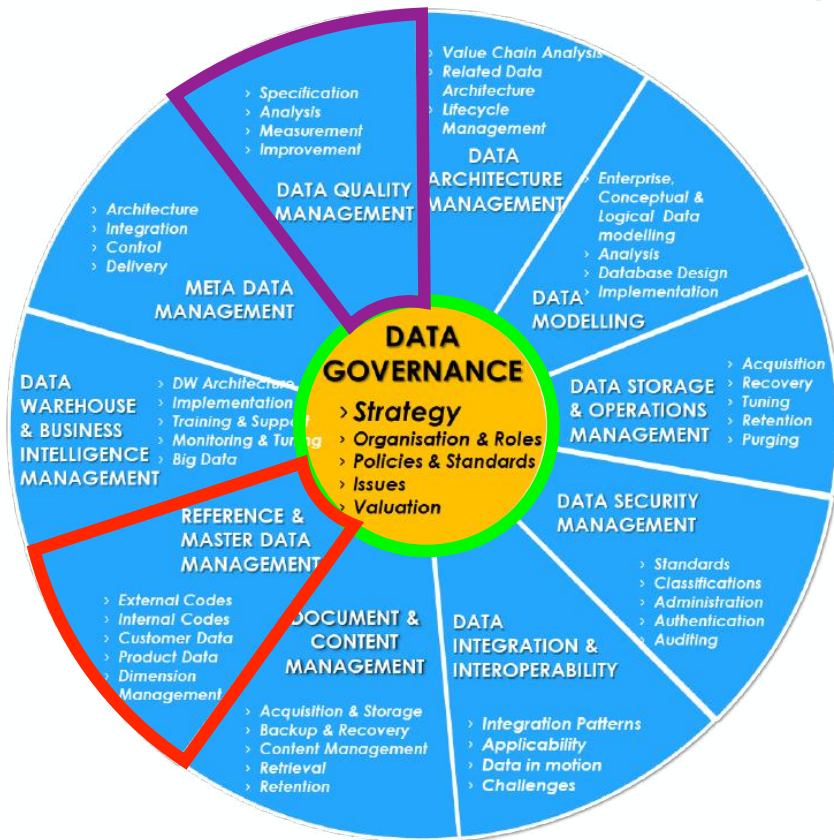
Vocabulary is Important-Tank, Tanks, Tankers
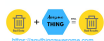
Multiple Sources of Master/Reference Data

Payroll Data (database)

Payroll Application (3rd GL)

Finance Application (3rd GL, batch system, no source)

Finance Data (indexed)

Marketing Data (external database)

Marketing Application (4rd GL, query facilities, no reporting, very large)

Personnel Data (database)

Personnel App. (20 years old, un-normalized data)

R & D Data (raw)

R& D Applications (researcher supported, no documentation)

Mfg. Data (home grown database)

Mfg. Applications (contractor supported)

Reference Data Architecture

Externally Sourced Data

Subscription Update

Direct Reads by Other OLTP Systems

OLTP Systems

Business Data Stewards

Code Management System

Reference Database Of Record

Controlled Update

Near-Real-Time Updates to Other OLTP Databases (Push, Pull)

OLTP Databases

Nightly Extract

Data Warehouse

Dimensional Data Marts

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Master Data Architecture

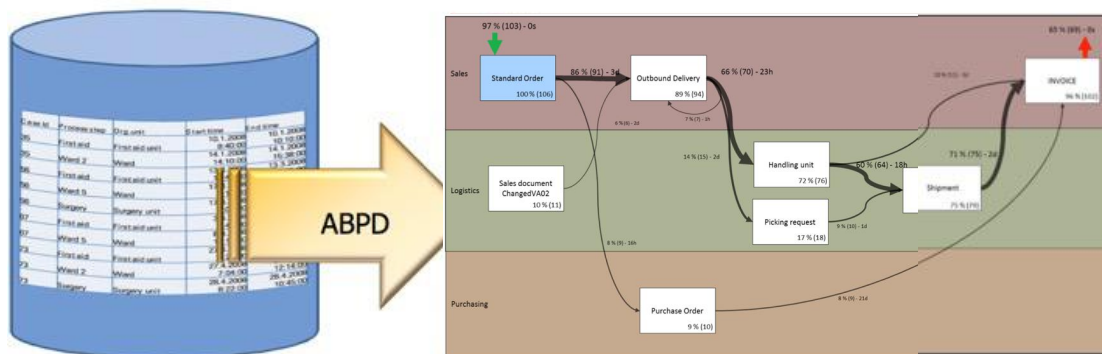# Combined R/M Data Architecture
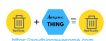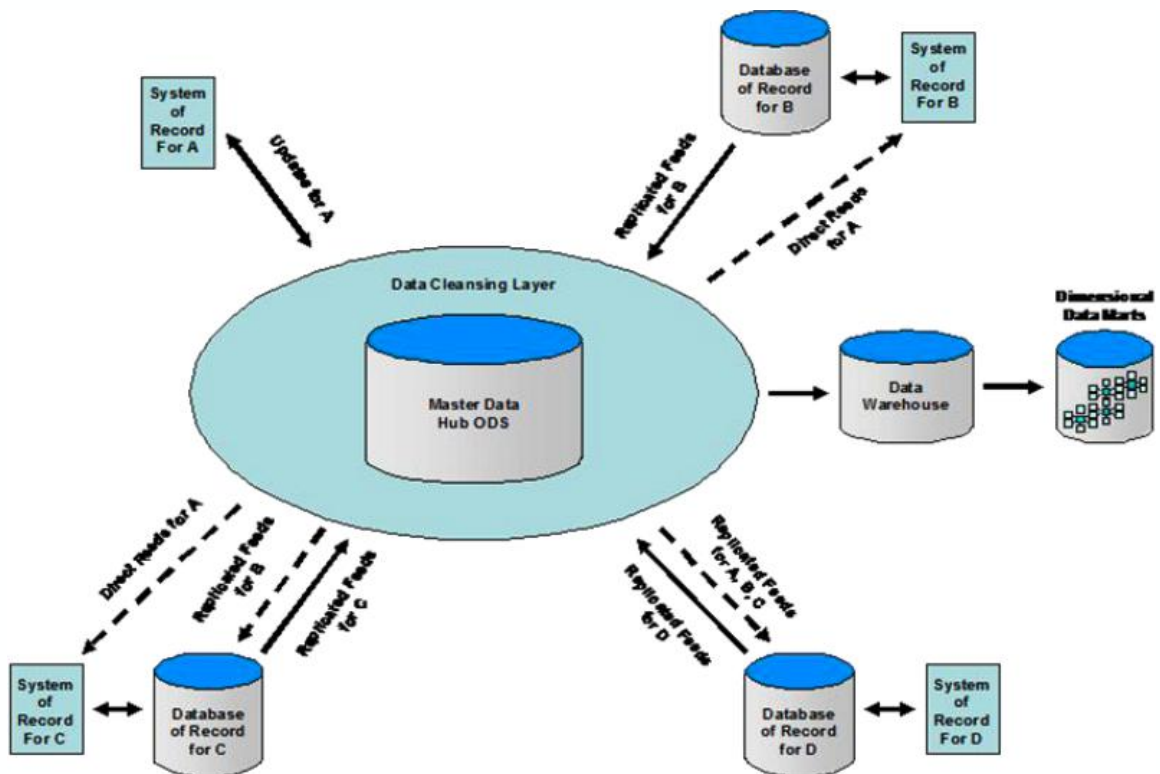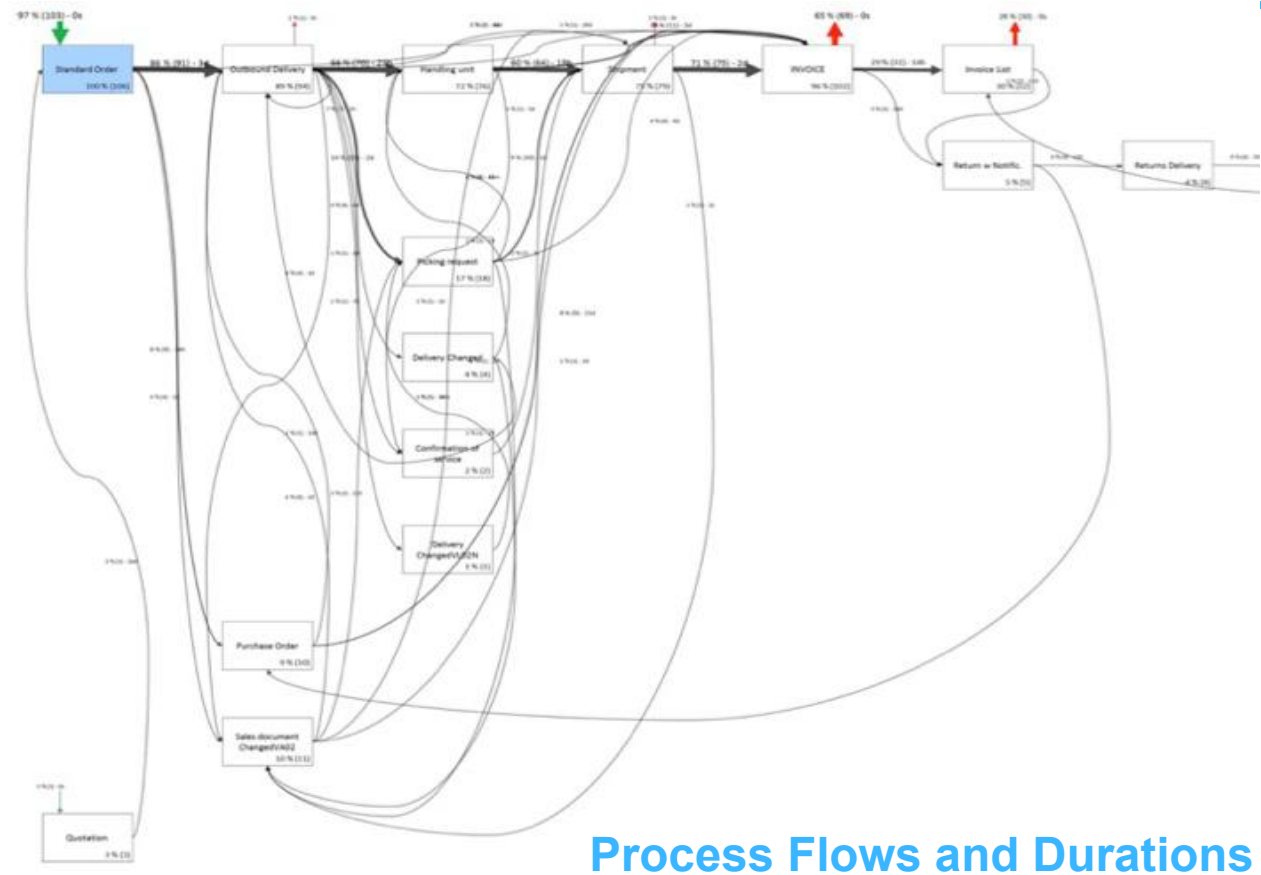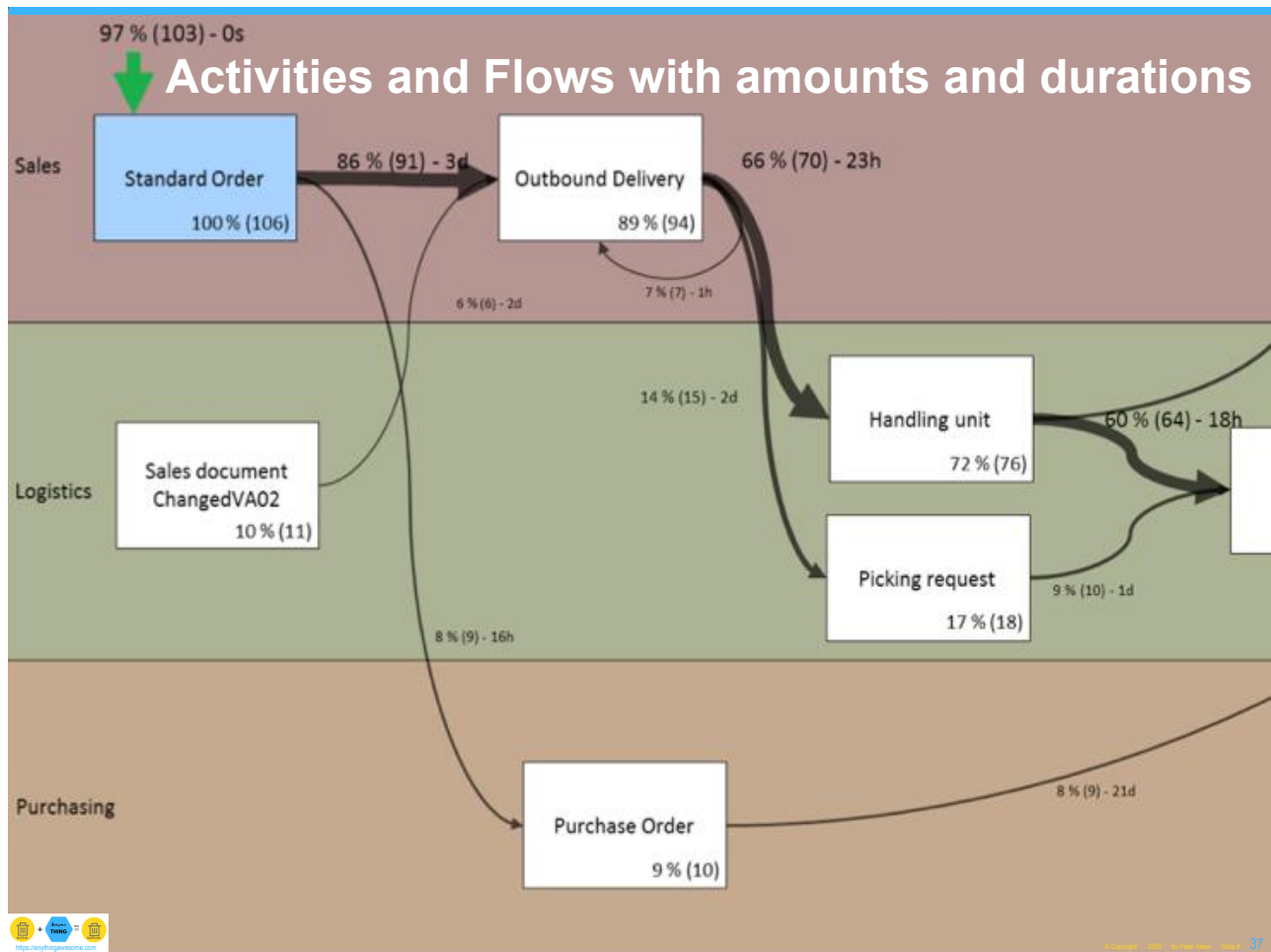
Task vs. Process Orientation

- **What is meant by a task orientation?**

  – Industrial work should be broken down into its simplest and most basic tasks

- **What is meant by a process orientation?**

  – Reunifying tasks into coherent business processes

- **What else must be part of the analysis?**

  – Identify and abandon outdated rules and assumptions that underlie current business operations

Task 1	Task 5	Task 9
Task 2	Task 6	Task 10
Task 3	Task 7	Task 11
Task 4	Task 8	Task 12

Source System: all info on 1 screen
Target system: same info spread across 23 screens
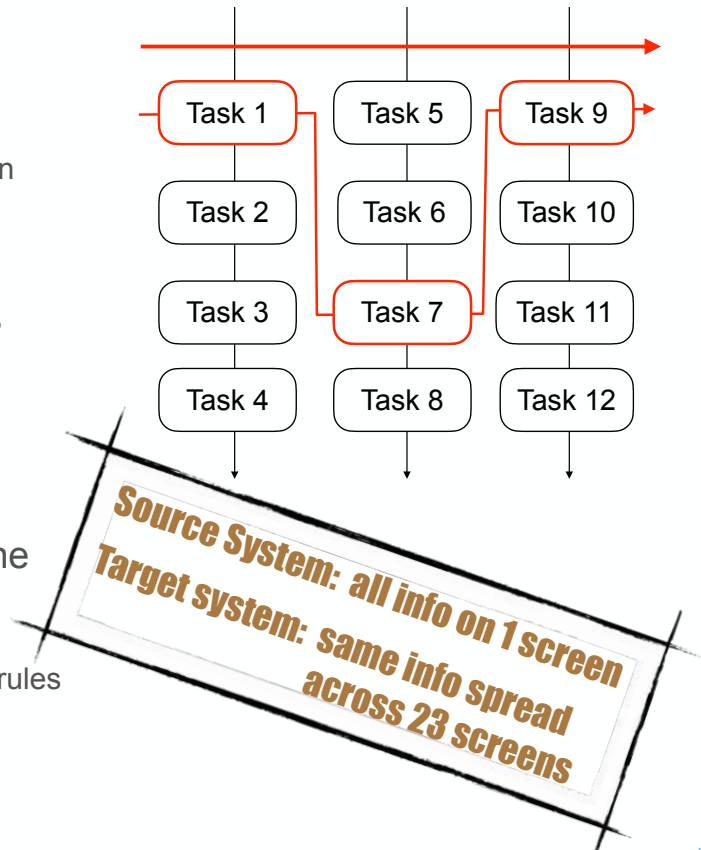
Automating Business Process Discovery (qpr.com)
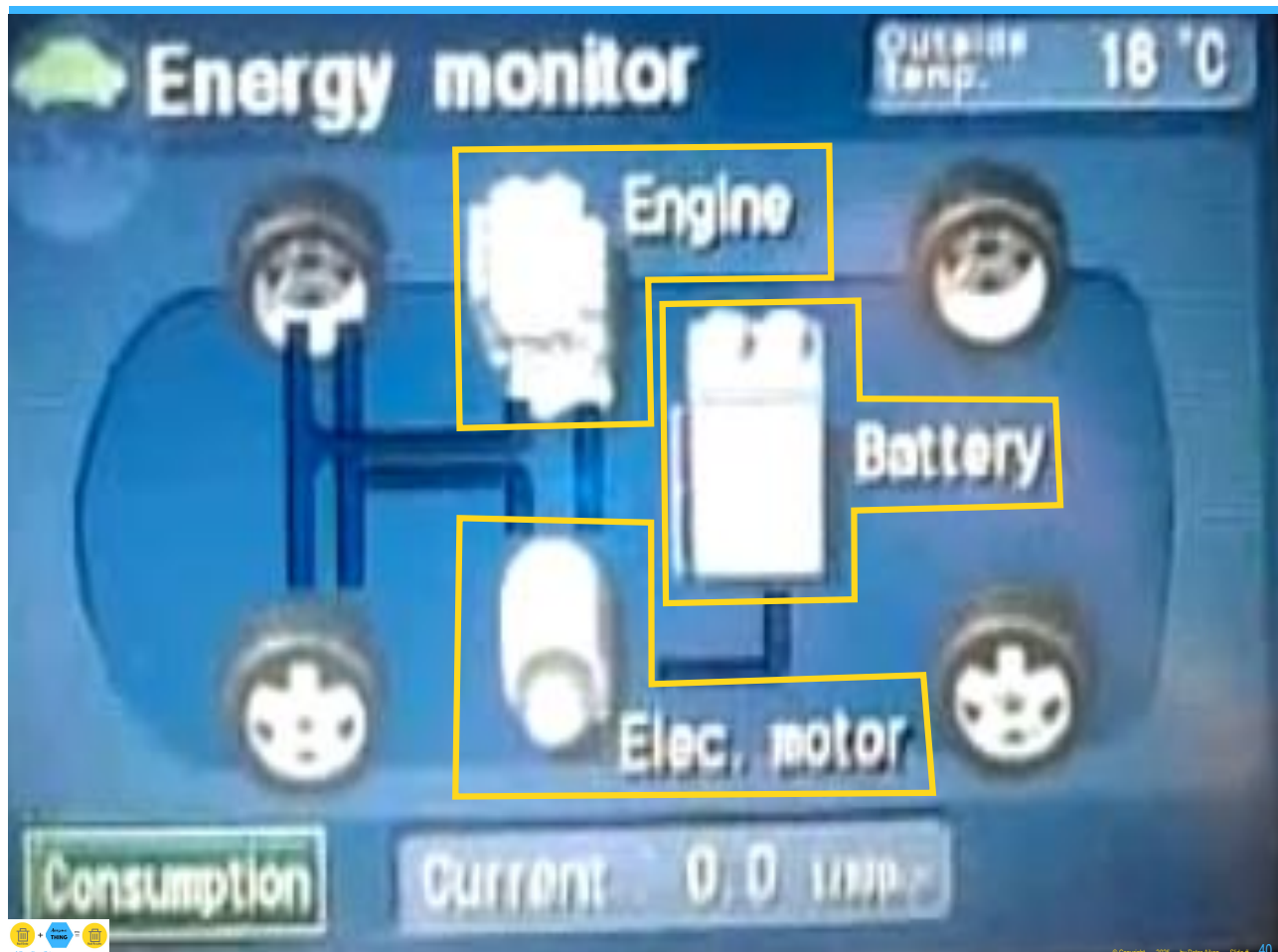
ABPD

- **Benefits**

  – Obtain holistic perspective on roles and value creation

  – Customers understand and value outputs

  – All develop better shared understanding

- **Results**

  – Speed up process

  – Cost savings

  – Increased compliance

  – Increased output

  – IT systems documentation

Activities and Flows with amounts and durations

97 % (103) - 0s

Sales

Standard Order
100 % (106)

86 % (91) - 3d

Outbound Delivery
89 % (94)

66 % (70) - 23h

7 % (7) - 1h

6 % (6) - 2d

14 % (15) - 2d

Logistics

Sales document ChangedVA02
10 % (11)

Handling unit
72 % (76)

60 % (64) - 18h

Picking request
17 % (18)

9 % (10) - 1d

8 % (9) - 16h

Purchasing

8 % (9) - 21d

Purchase Order
9 % (10)

Process Flows and Durations

Traditional Engine

Energy monitor — Outside Temp. 18 °C

Engine

Battery

Elec. motor

Consumption — Current 0.0 amp

Customer Data Management in a Retail Setting

Scenario: Reducing Wasteful Marketing Spend

Imagine a large retail company with multiple sales channels (online, brick-and-mortar stores, call centers). Over time, their customer data becomes fragmented and inconsistent:

- **Duplicate Records:** The same customer might be entered multiple times with slight variations in their name, address, or email.
- **Inconsistent Data:** Different systems might store customer information in different formats or with conflicting details.
- **Outdated Information:** Customer contact details might be outdated, leading to failed deliveries or communication attempts.

How Reference and Master Data Management Solve This and Save Money:

1. **Master Data Management (MDM):**
    - MDM initiatives focus on creating a "single version of the truth" for customer data.
    - This involves:
        - **Data Cleansing:** Identifying and correcting errors in customer data.
        - **Data Deduplication:** Merging or linking duplicate customer records to create a single, accurate view of each customer.
        - **Data Standardization:** Ensuring that customer data is stored in a consistent format across all systems.
    - By implementing MDM, the retailer creates a reliable, accurate, and up-to-date repository of customer information.
2. **Reference Data Management:**
    - Reference data plays a supporting role by ensuring consistency in how customer attributes are classified.
    - For example:
        - A "Customer Type" reference data set might include values like "Loyalty Member," "New Customer," or "VIP Customer."
        - This ensures that customer segmentation and targeting are based on consistent criteria.

Cost Savings Achieved:

- **Reduced Marketing Waste:**
    - With clean, deduplicated data, the retailer avoids sending the same marketing materials to the same customer multiple times.
    - This directly reduces printing, postage, and email marketing costs.
    - Targeted campaigns become more effective because they reach the correct customers with the right message, increasing conversion rates and reducing wasted effort.
- **Improved Operational Efficiency:**
    - Accurate customer data improves order processing, shipping, and customer service.
    - Fewer resources are spent on correcting errors or resolving issues caused by inaccurate data.
    - For example, reducing address errors minimizes failed deliveries, saving shipping costs and improving customer satisfaction.
- **Enhanced Decision-Making:**
    - With a single, accurate view of the customer, the retailer can gain better insights into customer behavior and preferences.
    - This leads to more informed business decisions, such as optimizing product offerings, personalizing marketing campaigns, and improving customer retention strategies.
    - Ultimately, better decisions drive increased revenue and profitability.

In Summary:

In this retail scenario, MDM and reference data management work together to improve the accuracy, consistency, and completeness of customer data. This leads to tangible cost savings by reducing marketing waste, improving operational efficiency, and enabling better decision-making. While the exact figures will vary, the principle remains consistent: **effective data management directly contributes to significant financial benefits for organizations.**

Program Overview

- Data Management Overview

- What is Reference and MDM?

- Why is Reference and MDM important?

- Reference & MDM Building Blocks

- Guiding Principles & Best Practices

- Take Aways, References & Q&A

Transforming Data with AI: Essential Reference and Master Data Management Strategies

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'
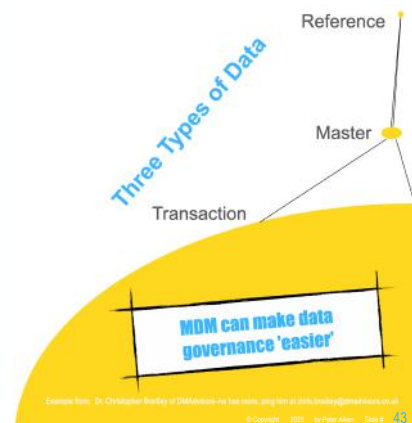
Goals and Principles

1. Provide authoritative source of reconciled, high-quality master and reference data.

2. Lower cost and complexity through reuse and leverage of standards.

3. Support business intelligence and information integration efforts

GOAL SETTING
- S SPECIFIC
- M MEASURABLE
- A ATTAINABLE
- R RELEVANT
- T TIME-BOUND
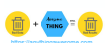
Reference & MDM Activities

- Understand reference and master data integration needs

- Identify master and reference data sources and contributors

- Define and maintain the data integration architecture

- Implement reference and master data management solutions

- Define and maintain match rules

- Establish "golden" records

- Define and maintain hierarchies and affiliations

- Plan and implement integration of new data sources

- Replicate and distribute reference and master data

- Manage changes to reference and master data
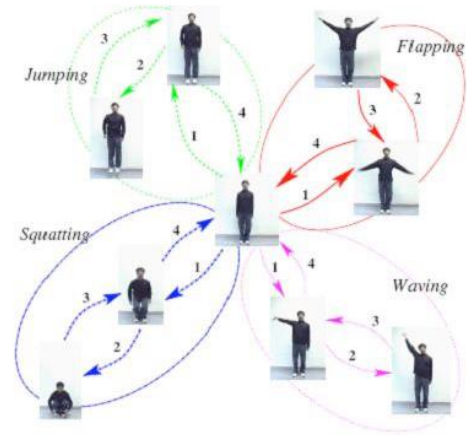
Specific Reference and MDM Investigations

- Who needs what information?

- What data is available from different sources?

- How does data from different sources differ?

- How can inconsistencies be reconciled?

- How should valid values be shared?

BE SPECIFIC
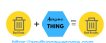
Primary Deliverables

- Data Cleansing Services

- Master and Reference Data Requirements

- Data Models and Documentation

- Reliable Reference and Master Data

- "Golden Record" Data Lineage

- Data Quality Metrics and Reports

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

---

Roles and Responsibilities

- Suppliers:
  - Steering Committees
  - Business Data Stewards
  - Subject Matter Experts
  - Data Consumers
  - Standards Organizations
  - Data Providers
  - ...

- Consumers:
  - Application Users
  - BI and Reporting Users
  - Application Developers and Architects
  - Data integration Developers and Architects
  - BI Vendors and Architects
  - Vendors, Customers and Partners
  - ...

- Participants:
  - Data Stewards
  - Subject Matter Experts
  - Data Architects
  - Data Analysts
  - Application Architects
  - Data Governance Council
  - Data Providers
  - Other IT Professionals
  - ...

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Technology

- ETL
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Metadata Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
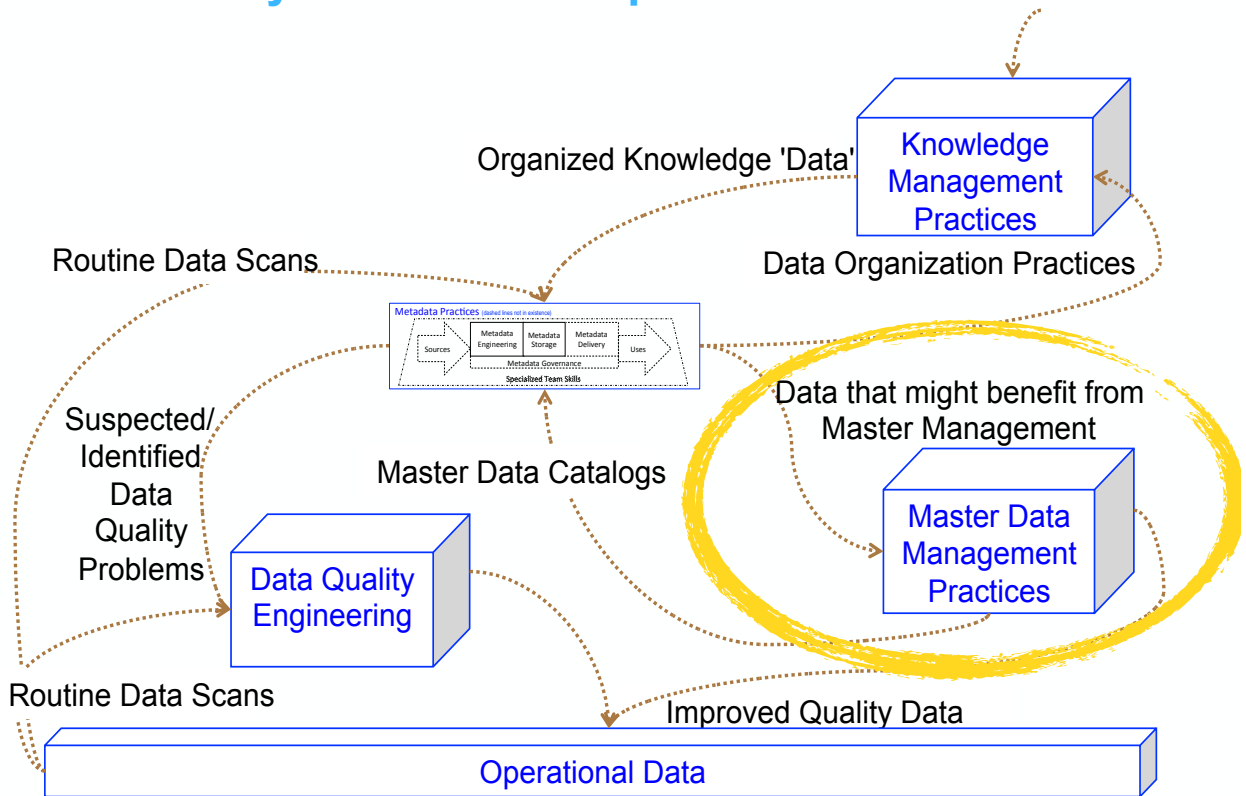- Business Process and Rule Engines
- Change Management Tools

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

---

General MDM Leaders and Well-Regarded Vendors

Based on evaluations from firms like Gartner and Forrester, as well as customer reviews, some of the MDM vendors that are frequently recognized as leaders and have a strong presence in the market include:

- **Informatica**: A broad data management platform with strong MDM capabilities.
- **Profisee**: Focuses specifically on MDM solutions and is known for its ease of use and implementation speed.
- **Stibo Systems**: A robust MDM platform with a focus on product MDM and multi-domain MDM.
- **Reltio**: A cloud-native MDM platform known for its data flexibility and graph-based approach.
- **SAP**: Offers MDM solutions as part of its broader enterprise software suite.
- **Oracle**: Also provides MDM solutions within its comprehensive data management offerings.

BYF
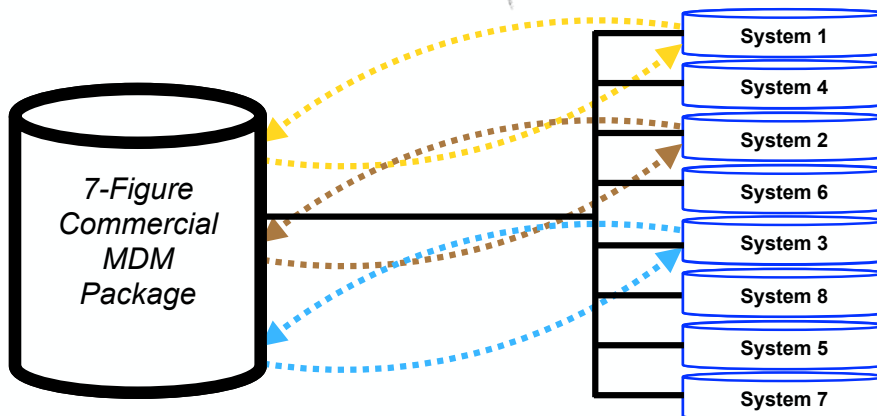
Inextricably intertwined implementations and …

Knowledge Management Practices

Organized Knowledge 'Data'

Data Organization Practices

Routine Data Scans

Metadata Practices (dotted lines not in existence)

Sources → Metadata Engineering → Metadata Storage → Metadata Delivery → Uses

Metadata Governance

Specialized Team Skills

Data that might benefit from Master Management

Master Data Catalogs

Suspected/ Identified Data Quality Problems

Master Data Management Practices

Data Quality Engineering

Routine Data Scans

Improved Quality Data

Operational Data

Interactions

Data Harvesting

Governance Violations Monitoring

Data Quality Monitoring

Master Data Monitoring

Routine Data Scans

Governance Rules

Monitoring Rules

Monitoring Results: Suspected/ Identified Data Quality Problems

Monitoring Results: Suspected/ Master Data & Characteristics

Master Data Catalogs

Data Governance Practices

Data Quality Rules

Quality Rules

Data Quality Engineering Practices

Master Data Management Practices

Routine Data Scans

Focused Data Scans

Improved Quality Data

Operational Data

Important Considerations

- **No Guarantee of Success**: Even the best MDM product won't guarantee success if the implementation is poorly planned or if there's a lack of data governance.
- **Focus on Your Needs**: The "best" MDM product for your organization will depend on your specific requirements, such as:

The domains of master data you need to manage (e.g., customer, product, location)

- Your industry-specific needs
- Your budget and technical capabilities
- Your preferred deployment model (e.g., cloud, on-premises)
- To find information on customer success rates:
  - **Vendor Case Studies**: MDM vendors often publish case studies on their websites that highlight successful customer implementations.
  - **Customer Reviews**: As mentioned earlier, websites like G2, TrustRadius, and PeerSpot provide customer reviews and ratings, which can give you a sense of customer satisfaction.
  - **Industry Reports**: Reports from Gartner, Forrester, and other analyst firms may include data or analysis on customer satisfaction with different MDM vendors.

Remember to conduct thorough research and consider your organization's specific needs before choosing an MDM product.

Program Overview

- Data Management Overview

- What is Reference and MDM?

- Why is Reference and MDM important?

- Reference & MDM Building Blocks

- Guiding Principles & Best Practices

- Take Aways, References & Q&A

DATA-ED Online

Transforming Data with AI: Essential Reference and Master Data Management Strategies

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'

Program Overview

- Data Management Overview

- What is Reference and MDM?

- Why is Reference and MDM important?

- Reference & MDM Building Blocks

- Guiding Principles & Best Practices

- Take Aways, References & Q&A

Transforming Data with AI: Essential Reference and Master Data Management Strategies

Three Types of Data

Reference

Master

Transaction

MDM can make data governance 'easier'

Example from: Dr. Christopher Bradley of DMMatters He had more, png him at chris.bradley@dmbloom.co.uk

© Copyright 2020 by Peter Aiken Slide # 57
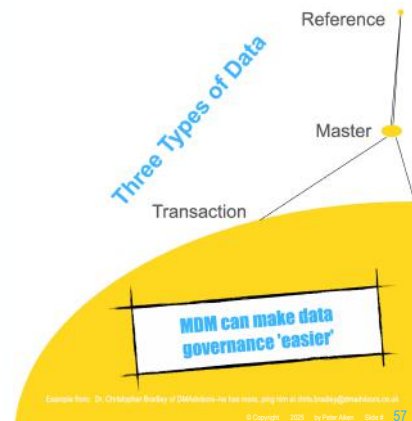
"180% Failure Rate" Fred Cohen, Patni

- 99% were less than fully satisfied with their Data Programs with a full 70% even less than just satisfied
  They were more comfortable with the quality of their pricing data but much less about Market, Account and Customer data.

- 76% cannot track and have no consolidated governance of their market/data spend and distribution. This also means they have limited control and understanding on how the consuming applications use the data thus any rationalization efforts are inhibited.

- 25% of a clients Reference Data management spend is wasted on duplicated data, 74% have silo data, expensive cleansing management and inefficient distribution
  Reference Data Management programs will be most active in Centralized Security masters and STP programs as well as better corporate actions processing and data scrubbing.

- Top motivators are by a far margin Risk reduction, followed by customer satisfaction, efficiency and then only decreased cost or increased revenue.

- Over 64% are planning to re-architect their reference data in the near future

- Over half spend more than four million dollars a year on Reference Data. By applying the estimated 25% inefficiencies reference Data management programs can clearly self fund initiatives.

Company Confidential   4   http://www.igatepatni.com/bfs/solutions/payments.aspx   patni

© Copyright 2020 by Peter Aiken Slide # 58

MDM Failure Root-Causes

- 30% of MDM programs are regarded as failures

- 70% of SOA projects in complex, heterogeneous environments had failed to yield the expected business benefits unless MDM is included

- Root-causes of failures:

  – 80% percent of MDM initiatives fail because of ineffective leadership, underestimated magnitudes or an inability to deal with the cultural impact of the change

  – MDM was implemented as a technology or as a project

  – MDM was an Enterprise Data Warehouse (EDW) or an ERP

  – MDM was an IT Effort

  – MDM is separate to data governance and data quality

  – MDM initiatives are implemented with inappropriate technology

  – Internal politics and the silo mentality impede the MDM initiatives

---
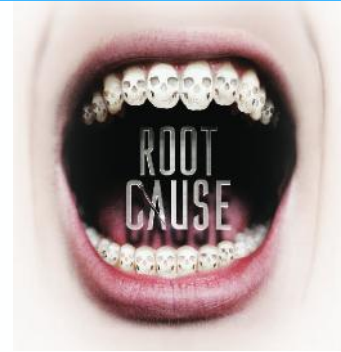
15 MDM Success Factors

1. Success is more likely and when users and prospects understand MDM limitations/strengths.
2. Taking small steps and remaining educated will increase longer-term success with MDM.
3. Set the right expectations.
4. Long-term MDM success requires information architecture.
5. Create incentives to ensure that manage master data is desirable.
6. Strong alignment with the organization's business vision, will underpin MDM success.
7. Use a framework through all stages of the MDM program — strategize, evaluate, execute and review.
8. Gain high-level business sponsorship and build strong stakeholder support.
9. Creating an MDM vision and a strategy aligned to the organization's business vision.
10. Use MDM metrics to communicate success and measure progress.
11. Use a business case to increase business engagement.
12. Get the business to propose and own the KPIs.
13. Measure the situation before and after.
14. Translate the change in metrics into financial results.
15. Achieve a single view of master data

[Source: unknown]

10 Best Practices for MDM

- Active, involved executive sponsorship
- The business should own the data governance process and the MDM or CDI project
- Strong project management and organizational change management
- Use a holistic approach - people, process, technology and information
- Build your processes to be ongoing and repeatable, supporting continuous improvement
- Management needs to recognize the importance of a dedicated team of data stewards
- Understand your MDM hub's data model and how it integrates with your internal source systems and external content providers
- Resist the urge to customize
- Stay current with vendor-provided patches
- Test, test, test and then test again.

GUIDING PRINCIPLES

1. Shared R/M data belong to the organization
2. R/M data management is an on-going data quality improvement program – goals cannot be achieved by 1 project alone.
3. Business data stewards are the authorities accountable at determining the golden values.
4. Golden values represent the "best" sources.
5. Replicate master data values only from golden sources.
6. Reference data changes require formal change management

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

a great customer experience
starts with excellent data

previously information was stored
in hundreds of databases across BT

Seven Sisters (from British Telecom)

now we are sorting
all of our data

Which Approach Is Better?

There's no universally "better" approach. The best choice depends on:

- **Complexity of Requirements**: If your MDM needs are very complex and unique, a custom solution might seem appealing, but vendor solutions are increasingly flexible.
- **Budget and Resources**: Vendor solutions have higher upfront costs, but custom solutions have high development and maintenance costs.
- **Time to Value**: Vendor solutions offer a much faster time to value.
- **Internal Expertise**: Do you have the in-house expertise to design, develop, and maintain a robust MDM solution?
- **Long-Term Strategy**: Consider your long-term data management strategy and how the chosen approach will support it.

In many cases, especially for organizations with complex MDM requirements, limited development resources, or a need for rapid deployment, vendor solutions are often the more practical and effective choice. They provide specialized features, reduce development risk, and offer ongoing support.

However, if an organization has very specific needs, strong development capabilities, and a long-term focus, a carefully planned custom solution might be considered.

Ultimately, a thorough evaluation of your organization's specific needs and capabilities is essential to determine the best approach for reference and master data management.

Program **Overview**

- Data Management Overview

- What is Reference and MDM?

- Why is Reference and MDM important?

- Reference & MDM Building Blocks

- Guiding Principles & Best Practices

- Take Aways, References & Q&A

Essential:
Reference &
Master Data

DATA-ED
Online

Ⅲ DATAVERSITY

Definition: Planning, implementation, and control activities to ensure consistency with a "golden version" of contextual data values.

Goals:
1. Provide authoritative source of reconciled, high-quality master and reference data.
2. Lower cost and complexity through reuse and leverage of standards.
3. Support business intelligence and information integration efforts.

Inputs:
- Business Drivers
- Data Requirements
- Policy and Regulations
- Standards
- Code Sets
- Master Data
- Transactional Data

Suppliers:
- Steering Committees
- Business Data Stewards
- Subject Matter Experts
- Data Consumers
- Standards Organizations
- Data Providers

Participants:
- Data Stewards
- Subject Matter Experts
- Data Architects
- Data Analysts
- Application Architects
- Data Governance Council
- Data Providers
- Other IT Professionals

Activities:
1. Understand Reference and Master Data Integration Needs (P)
2. Identify Master and Reference Data Sources and Contributors (P)
3. Define and Maintain the Data Integration Architecture (P)
4. Implement Reference and Master Data Management Solutions (D)
5. Define and Maintain Match Rules (C)
6. Establish "Golden" Records (C)
7. Define and Maintain Hierarchies and Affiliations (C)
8. Plan and Implement Integration of New Data Sources (D)
9. Replicate and Distribute Reference and Master Data (O)
10. Manage Changes to Reference and Master Data (O)

Tools:
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Meta-data Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Business Process and Rule Engines
- Change Management Tools

Primary Deliverables:
- Master and Reference Data Requirements
- Data Models and Documentation
- Reliable Reference and Master Data
- "Golden Record" Data Lineage
- Data Quality Metrics and Reports
- Data Cleansing Services

Consumers:
- Application Users
- BI and Reporting Users
- Application Developers and Architects
- Data Integration Developers and Architects
- BI Developers and Architects
- Vendors, Customers, and Partners

Metrics:
- Reference and Master Data Quality
- Change Activity
- Issues, Costs, Volume
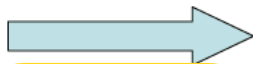- Use and Re-Use
- Availability
- Data Steward Coverage

Activities: (P) – Planning (C) – Control (D) – Development (O) - Operational

Definition: Planning, implementation, and control activities to ensure consistency with a "golden version" of contextual data values.

Goals:
1. Provide authoritative source of reconciled, high-quality master and reference data.
2. Lower cost and complexity through reuse and leverage of standards.
3. Support business intelligence and information integration efforts.

Inputs:
- Business Drivers
- Data Requirements
- Policy and Regulations
- Standards
- Code Sets
- Master Data
- Transactional Data

Suppliers:
- Steering Committees
- Business Data Stewards
- Subject Matter Experts
- Data Consumers
- Standards Organizations
- Data Providers

Participants:
- Data Stewards
- Subject Matter Experts
- Data Architects
- Data Analysts
- Application Architects
- Data Governance Council
- Data Providers
- Other IT Professionals

Activities:
1. Understand Reference and Master Data Integration Needs (P)
2. Identify Master and Reference Data Sources and Contributors (P)
3. Define and Maintain the Data Integration Architecture (P)
4. Implement Reference and Master Data Management Solutions (D)
5. Define and Maintain Match Rules (C)
6. Establish "Golden" Records (C)
7. Define and Maintain Hierarchies and Affiliations (C)
8. Plan and Implement Integration of New Data Sources (D)
9. Replicate and Distribute Reference and Master Data (O)
10. Manage Changes to Reference and Master Data (O)

Tools:
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Meta-data Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Business Process and Rule Engines
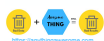- Change Management Tools

Primary Deliverables:
- Master and Reference Data Requirements
- Data Models and Documentation
- Reliable Reference and Master Data
- "Golden Record" Data Lineage
- Data Quality Metrics and Reports
- Data Cleansing Services

Consumers:
- Application Users
- BI and Reporting Users
- Application Developers and Architects
- Data Integration Developers and Architects
- BI Developers and Architects
- Vendors, Customers, and Partners

Metrics:
- Reference and Master Data Quality
- Change Activity
- Issues, Costs, Volume
- Use and Re-Use
- Availability
- Data Steward Coverage

Activities: (P) – Planning (C) – Control (D) – Development (O) - Operational

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Definition: Planning, implementation, and control activities to ensure consistency with a "golden version" of contextual data values.

Goals:
1. Provide authoritative source of reconciled, high-quality master and reference data.
2. Lower cost and complexity through reuse and leverage of standards.
3. Support business intelligence and information integration efforts.

Inputs:
- Business Drivers
- Data Requirements
- Policy and Regulations
- Standards
- Code Sets
- Master Data
- Transactional Data

Suppliers:
- Steering Committees
- Business Data Stewards
- Subject Matter Experts
- Data Consumers
- Standards Organizations
- Data Providers

Participants:
- Data Stewards
- Subject Matter Experts
- Data Architects
- Data Analysts
- Application Architects
- Data Governance Council
- Data Providers
- Other IT Professionals

Activities:
1. Understand Reference and Master Data Integration Needs (P)
2. Identify Master and Reference Data Sources and Contributors (P)
3. Define and Maintain the Data Integration Architecture (P)
4. Implement Reference and Master Data Management Solutions (D)
5. Define and Maintain Match Rules (C)
6. Establish "Golden" Records (C)
7. Define and Maintain Hierarchies and Affiliations (C)
8. Plan and Implement Integration of New Data Sources (D)
9. Replicate and Distribute Reference and Master Data (O)
10. Manage Changes to Reference and Master Data (O)

Tools:
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Meta-data Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Business Process and Rule Engines
- Change Management Tools

Primary Deliverables:
- Master and Reference Data Requirements
- Data Models and Documentation
- Reliable Reference and Master Data
- "Golden Record" Data Lineage
- Data Quality Metrics and Reports
- Data Cleansing Services

Consumers:
- Application Users
- BI and Reporting Users
- Application Developers and Architects
- Data Integration Developers and Architects
- BI Developers and Architects
- Vendors, Customers, and Partners

Metrics:
- Reference and Master Data Quality
- Change Activity
- Issues, Costs, Volume
- Use and Re-Use
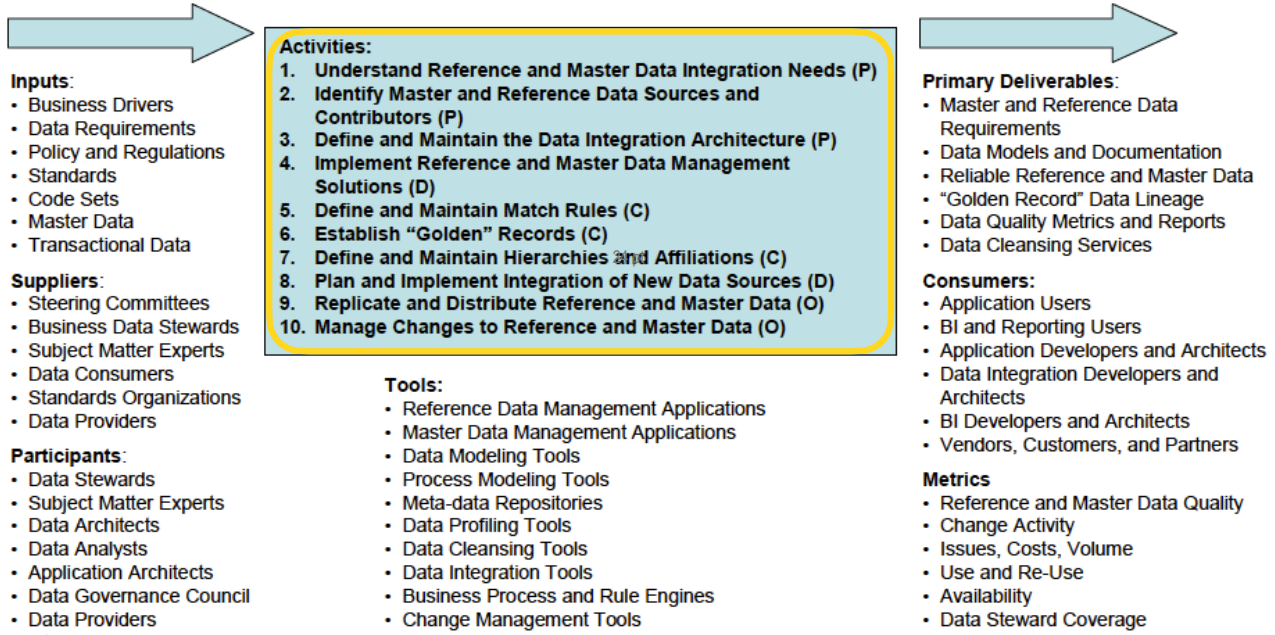- Availability
- Data Steward Coverage

Activities: (P) – Planning (C) – Control (D) – Development (O) - Operational

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

https://anythingawesome.com

Definition: Planning, implementation, and control activities to ensure consistency with a "golden version" of contextual data values.

Goals:
1. Provide authoritative source of reconciled, high-quality master and reference data.
2. Lower cost and complexity through reuse and leverage of standards.
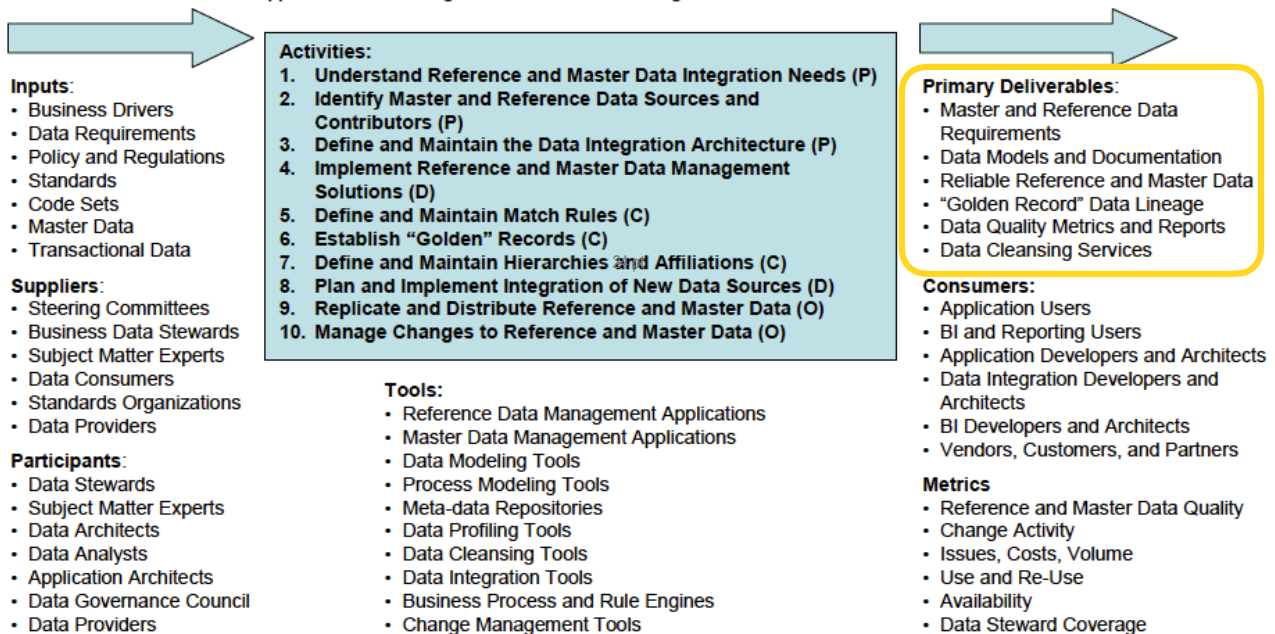3. Support business intelligence and information integration efforts.

Inputs:
- Business Drivers
- Data Requirements
- Policy and Regulations
- Standards
- Code Sets
- Master Data
- Transactional Data

Suppliers:
- Steering Committees
- Business Data Stewards
- Subject Matter Experts
- Data Consumers
- Standards Organizations
- Data Providers

Participants:
- Data Stewards
- Subject Matter Experts
- Data Architects
- Data Analysts
- Application Architects
- Data Governance Council
- Data Providers
- Other IT Professionals

Activities:
1. Understand Reference and Master Data Integration Needs (P)
2. Identify Master and Reference Data Sources and Contributors (P)
3. Define and Maintain the Data Integration Architecture (P)
4. Implement Reference and Master Data Management Solutions (D)
5. Define and Maintain Match Rules (C)
6. Establish "Golden" Records (C)
7. Define and Maintain Hierarchies and Affiliations (C)
8. Plan and Implement Integration of New Data Sources (D)
9. Replicate and Distribute Reference and Master Data (O)
10. Manage Changes to Reference and Master Data (O)

Tools:
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Meta-data Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Business Process and Rule Engines
- Change Management Tools

Primary Deliverables:
- Master and Reference Data Requirements
- Data Models and Documentation
- Reliable Reference and Master Data
- "Golden Record" Data Lineage
- Data Quality Metrics and Reports
- Data Cleansing Services

Consumers:
- Application Users
- BI and Reporting Users
- Application Developers and Architects
- Data Integration Developers and Architects
- BI Developers and Architects
- Vendors, Customers, and Partners

Metrics:
- Reference and Master Data Quality
- Change Activity
- Issues, Costs, Volume
- Use and Re-Use
- Availability
- Data Steward Coverage

Activities: (P) – Planning (C) – Control (D) – Development (O) - Operational

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

---

Definition: Planning, implementation, and control activities to ensure consistency with a "golden version" of contextual data values.

Goals:
1. Provide authoritative source of reconciled, high-quality master and reference data.
2. Lower cost and complexity through reuse and leverage of standards.
3. Support business intelligence and information integration efforts.

Inputs:
- Business Drivers
- Data Requirements
- Policy and Regulations
- Standards
- Code Sets
- Master Data
- Transactional Data

Suppliers:
- Steering Committees
- Business Data Stewards
- Subject Matter Experts
- Data Consumers
- Standards Organizations
- Data Providers

Participants:
- Data Stewards
- Subject Matter Experts
- Data Architects
- Data Analysts
- Application Architects
- Data Governance Council
- Data Providers
- Other IT Professionals

Activities:
1. Understand Reference and Master Data Integration Needs (P)
2. Identify Master and Reference Data Sources and Contributors (P)
3. Define and Maintain the Data Integration Architecture (P)
4. Implement Reference and Master Data Management Solutions (D)
5. Define and Maintain Match Rules (C)
6. Establish "Golden" Records (C)
7. Define and Maintain Hierarchies and Affiliations (C)
8. Plan and Implement Integration of New Data Sources (D)
9. Replicate and Distribute Reference and Master Data (O)
10. Manage Changes to Reference and Master Data (O)

Tools:
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Meta-data Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Business Process and Rule Engines
- Change Management Tools

Primary Deliverables:
- Master and Reference Data Requirements
- Data Models and Documentation
- Reliable Reference and Master Data
- "Golden Record" Data Lineage
- Data Quality Metrics and Reports
- Data Cleansing Services

Consumers:
- Application Users
- BI and Reporting Users
- Application Developers and Architects
- Data Integration Developers and Architects
- BI Developers and Architects
- Vendors, Customers, and Partners

Metrics:
- Reference and Master Data Quality
- Change Activity
- Issues, Costs, Volume
- Use and Re-Use
- Availability
- Data Steward Coverage

Activities: (P) – Planning (C) – Control (D) – Development (O) - Operational

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Definition: Planning, implementation, and control activities to ensure consistency with a "golden version" of contextual data values.

Goals:
1. Provide authoritative source of reconciled, high-quality master and reference data.
2. Lower cost and complexity through reuse and leverage of standards.
3. Support business intelligence and information integration efforts.

Inputs:
- Business Drivers
- Data Requirements
- Policy and Regulations
- Standards
- Code Sets
- Master Data
- Transactional Data

Suppliers:
- Steering Committees
- Business Data Stewards
- Subject Matter Experts
- Data Consumers
- Standards Organizations
- Data Providers

Participants:
- Data Stewards
- Subject Matter Experts
- Data Architects
- Data Analysts
- Application Architects
- Data Governance Council
- Data Providers
- Other IT Professionals

Activities:
1. Understand Reference and Master Data Integration Needs (P)
2. Identify Master and Reference Data Sources and Contributors (P)
3. Define and Maintain the Data Integration Architecture (P)
4. Implement Reference and Master Data Management Solutions (D)
5. Define and Maintain Match Rules (C)
6. Establish "Golden" Records (C)
7. Define and Maintain Hierarchies and Affiliations (C)
8. Plan and Implement Integration of New Data Sources (D)
9. Replicate and Distribute Reference and Master Data (O)
10. Manage Changes to Reference and Master Data (O)

Tools:
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Meta-data Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Business Process and Rule Engines
- Change Management Tools

Primary Deliverables:
- Master and Reference Data Requirements
- Data Models and Documentation
- Reliable Reference and Master Data
- "Golden Record" Data Lineage
- Data Quality Metrics and Reports
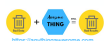- Data Cleansing Services

Consumers:
- Application Users
- BI and Reporting Users
- Application Developers and Architects
- Data Integration Developers and Architects
- BI Developers and Architects
- Vendors, Customers, and Partners

Metrics:
- Reference and Master Data Quality
- Change Activity
- Issues, Costs, Volume
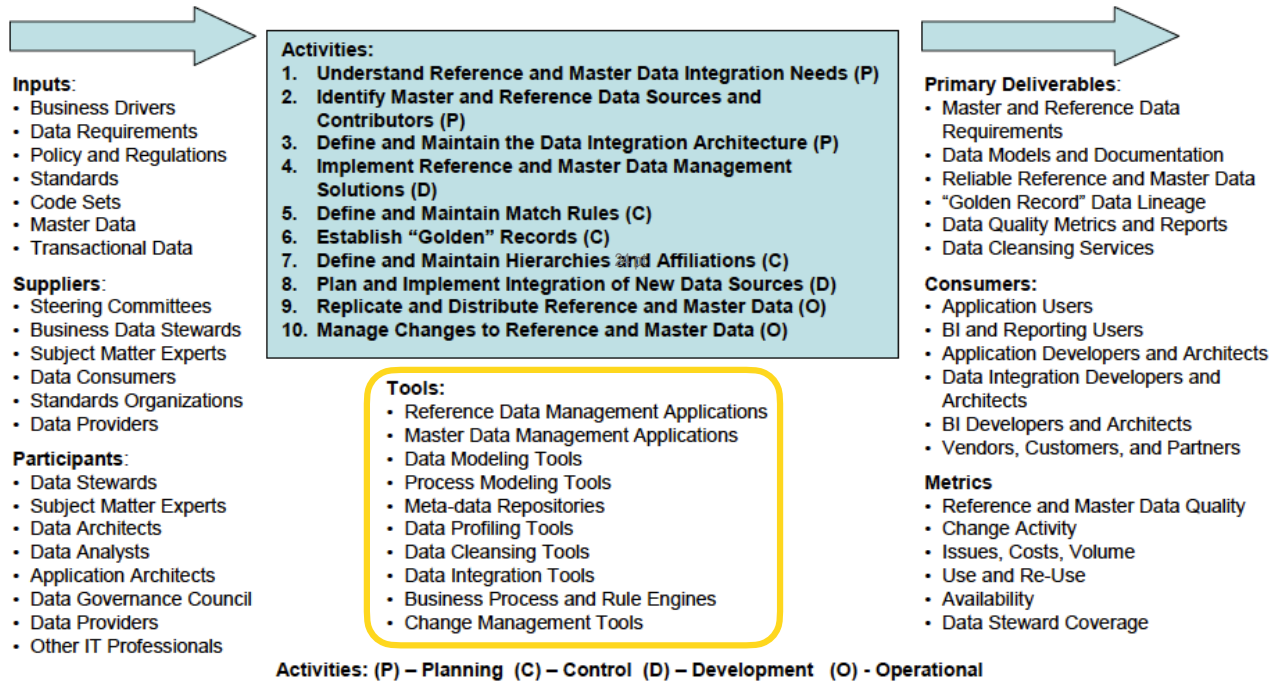- Use and Re-Use
- Availability
- Data Steward Coverage

Activities: (P) – Planning   (C) – Control   (D) – Development   (O) - Operational

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

References

Bean, James. XML for Data Architects: Designing for Reuse and Integration. Morgan Kaufmann, 2003. ISBN 1-558-60907-5. 250 pages.

Berson, Alex and Larry Dubov. Master Data Management and Customer Data Integration for a Global Enterprise. McGraw-Hill, 2007. ISBN 0-072-26349-0. 400 pages.

Brackett, Michael. Data Sharing Using A Common Data Architecture. New York: John Wiley & Sons, 1994. ISBN 0-471-30993-1. 478 pages.

Chisholm, Malcolm. Managing Reference Data in Enterprise Databases: Binding Corporate Data to the Wider World. Morgan Kaufmann, 2000. ISBN 1-558-60697-1. 389 pages.

Dreibelbis, Allen, Eberhard Hechler, Ivan Milman, Martin Oberhofer, Paul van Run, and Dan Wolfson. Enterprise Master Data Management: An SOA Approach to Managing Core Information. IBM Press, 2008. ISBN 978-0-13-236625-0. 617 pages.

Dyche, Jill and Evan Levy. Customer Data Integration: Reaching a Single Version of the Truth. John Wiley & Sons, 2006. ISBN 0-471-91697-8. 320 pages.

Finkelstein, Clive. Enterprise Architecture for Integration: Rapid Delivery Methods and Techniques. Artech House Mobile Communications Library, 2006. ISBN 1-580-53713-8. 546 pages.

Loshin, David. Master Data Management. Morgan Kaufmann, 2008. ISBN 98-0-12-374225-4. 274 pages.

Loshin, David. Enterprise Knowledge Management: The Data Quality Approach. Morgan Kaufmann, 2001. ISBN 0-124-55840-2. 494 pages.

National Information Standards Association (NISO), ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. 2005. 172 pages. www.niso.org

Additional References

- http://www.mdmsource.com/master-data-management-tips-best-practices.html
- http://www.igate.com/22926.aspx
- http://www.itbusinessedge.com/cm/blogs/lawson/just-the-stats-master-data-management/?cs=50349
- http://searchcio-midmarket.techtarget.com/news/2240150296/Smart-grid-systems-expert-devises-business-transformation-template
- http://www.itbusinessedge.com/cm/blogs/lawson/free-report-shows-businesses-fed-up-with-bad-data/?cs=50416
- http://www.itbusinessedge.com/cm/blogs/lawson/whats-ahead-for-master-data-management/?cs=50082
- http://www.itbusinessedge.com/cm/blogs/vizard/master-data-management-reaches-for-the-cloud/?cs=49264
- http://www.information-management.com/channels/master-data-management.html
- http://www.dataversity.net/applying-six-sigma-to-master-data-management-mdm-framework-for-integrating-mdm-into-ea-part-2/
- http://www.dataqualityfirst.com/getting_master_data_facts_straight_is_hard.htm

Event Pricing on Peter's Books
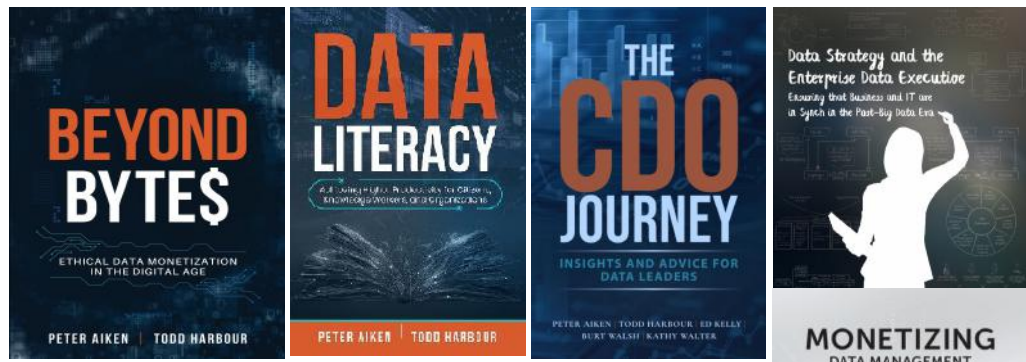
- 25% off directly from the publisher on some titles
- My 'Book Store' @
  https://anythingawesome.com/books
- Enter the code
  "TP25" at the Technics bookstore
  checkout where it says to "Apply Coupon"

TP25   Apply coupon
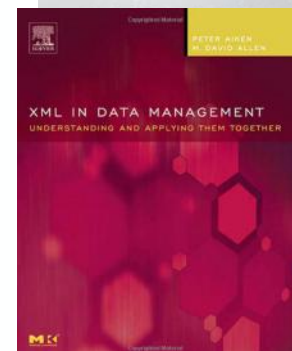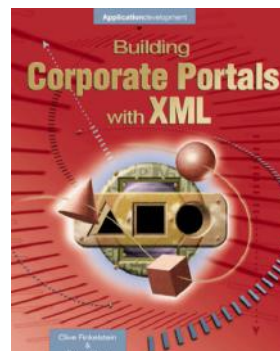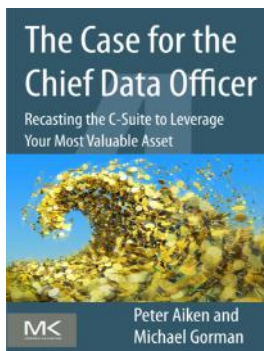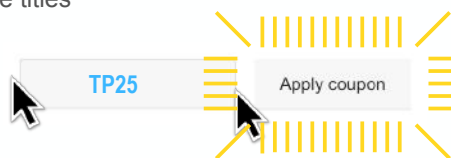
The Importance of Ethics in Data Governance
13 May 2025

Implementing Effective Data Governance: A Practical Guide
10 June 2025

Mastering Data Modeling: Understanding Conceptual, Logical, and Physical Models
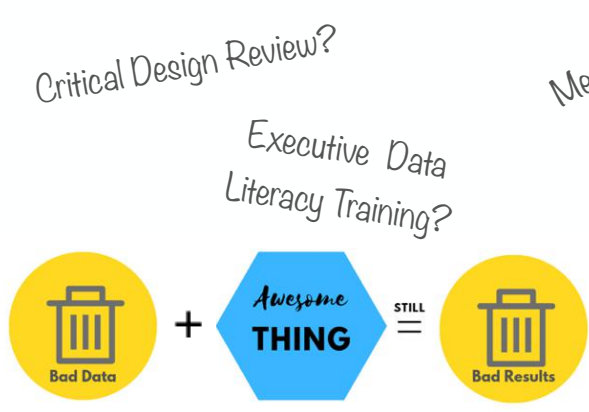*8 July 2025*

Brought to you by:

[Clicking any webinar title will link directly to the registration page]

III DATAVERSITY

© Copyright 2025 by Peter Aiken Slide # 79

Critical Design Review?

Mentoring?

Collaboration?

Executive Data Literacy Training?

Independent Verification & Validation

Bad Data + Awesome THING STILL = Bad Results

Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Reverse Engineering Expertise?

Hiring Assistance?

Thank You!

Use your data more strategically?

Tool/automation evaluation?

Book a call with Peter to discuss anything - https://anythingawesome.com/OfficeHours.html