



# Essential: Reference & Master Data



<https://api.deepai.org/job-view-file/98f46174-0564-4b06-b93b-b0daca83699b/outputs/output.jpg>



© Copyright 2023 Peter Aiken, PhD Slide # 1

## Peter Aiken, Ph.D.

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



<https://anythingawesome.com>

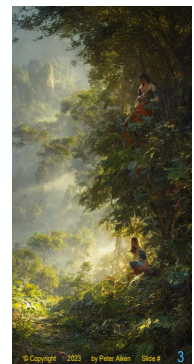


# Program Overview

- 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

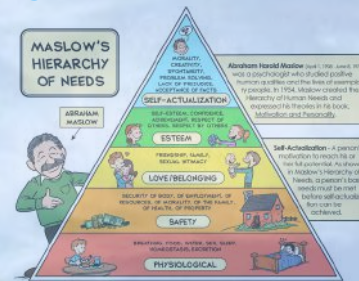
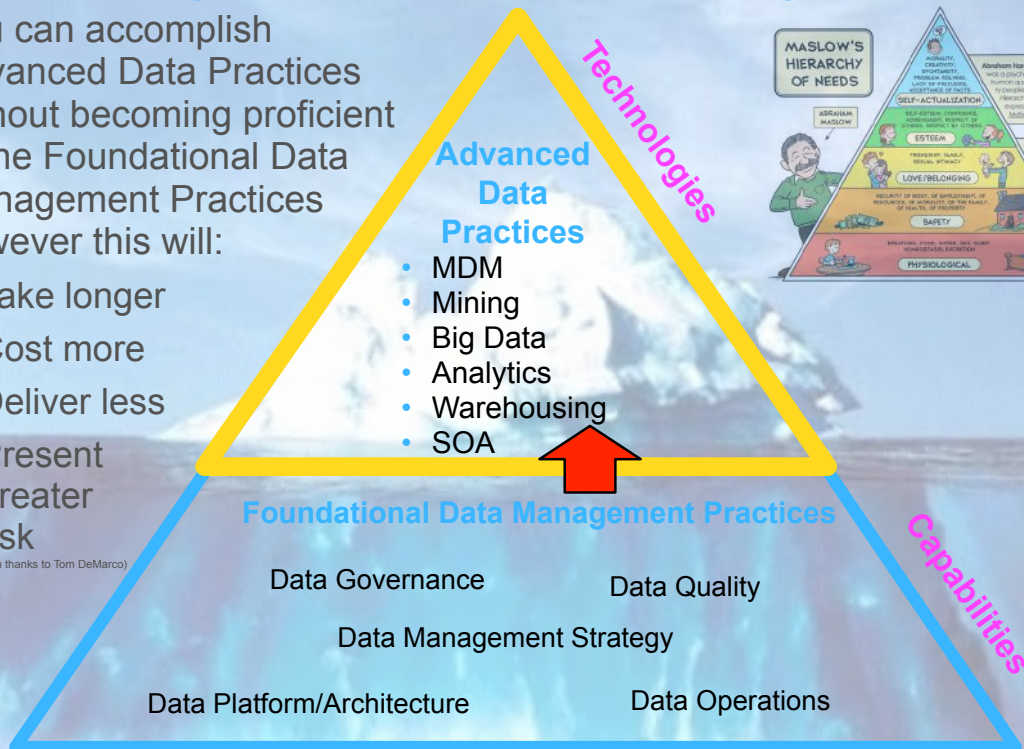


## Data Management Practices Hierarchy

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

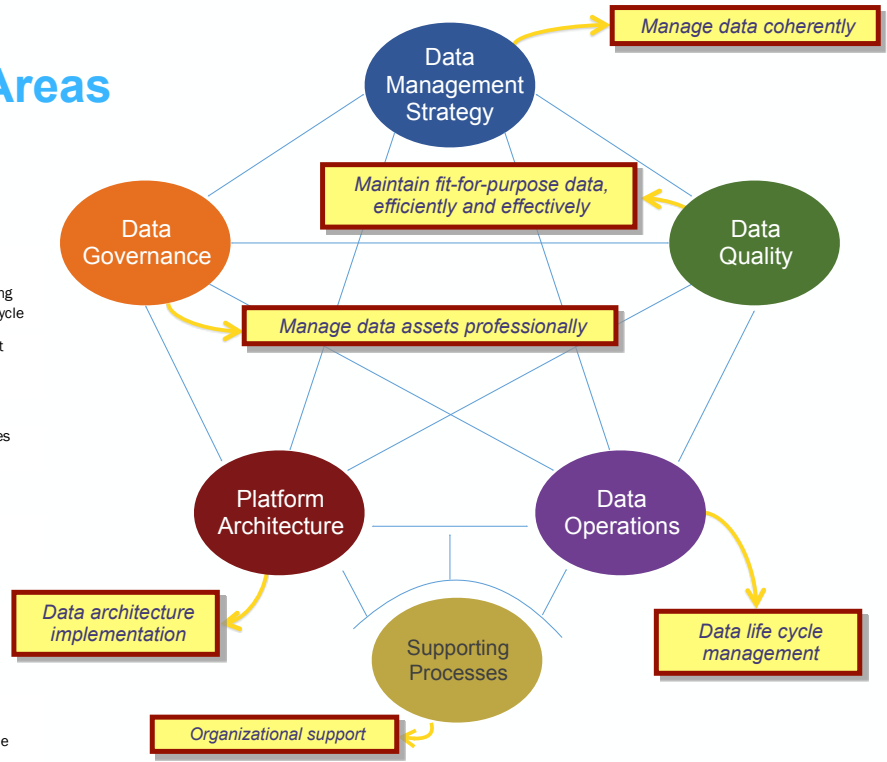
- Take longer
- Cost more
- Deliver less
- Present greater risk

(with thanks to Tom DeMarco)



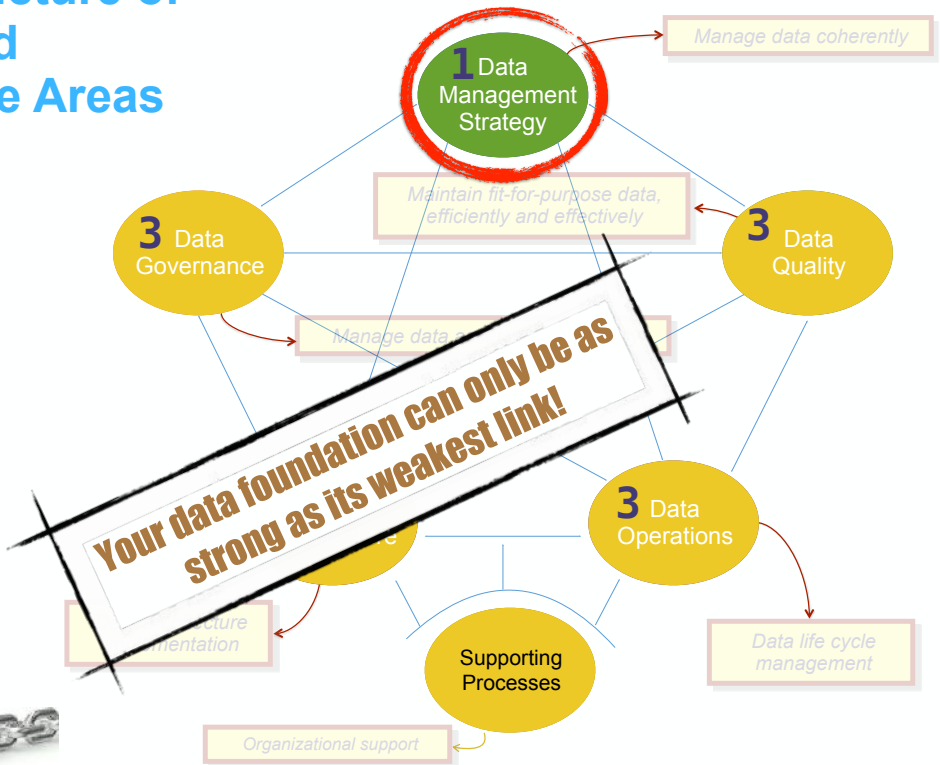
# DMM<sup>SM</sup> Structure of 5 Integrated DM Practice Areas

| Component                | Process Areas  |
|--------------------------|--|
| Data Management Strategy | Data Management Goals<br>Corporate Culture<br>Data Management Funding<br>Data Requirements Lifecycle                     |
| Data Governance          | Governance Management<br>Business Glossary<br>Metadata Management  |
| Data Operations          | Standards and Procedures<br>Data Sourcing  |
| Data Quality             | Data Quality Framework<br>Data Quality Assurance   |
| Platform & Architecture  | Architectural Framework<br>Platforms & Integration   |
| Supporting Processes     | Measurement & Analysis<br>Process Management<br>Process Quality Assurance<br>Risk Management<br>Configuration Management |

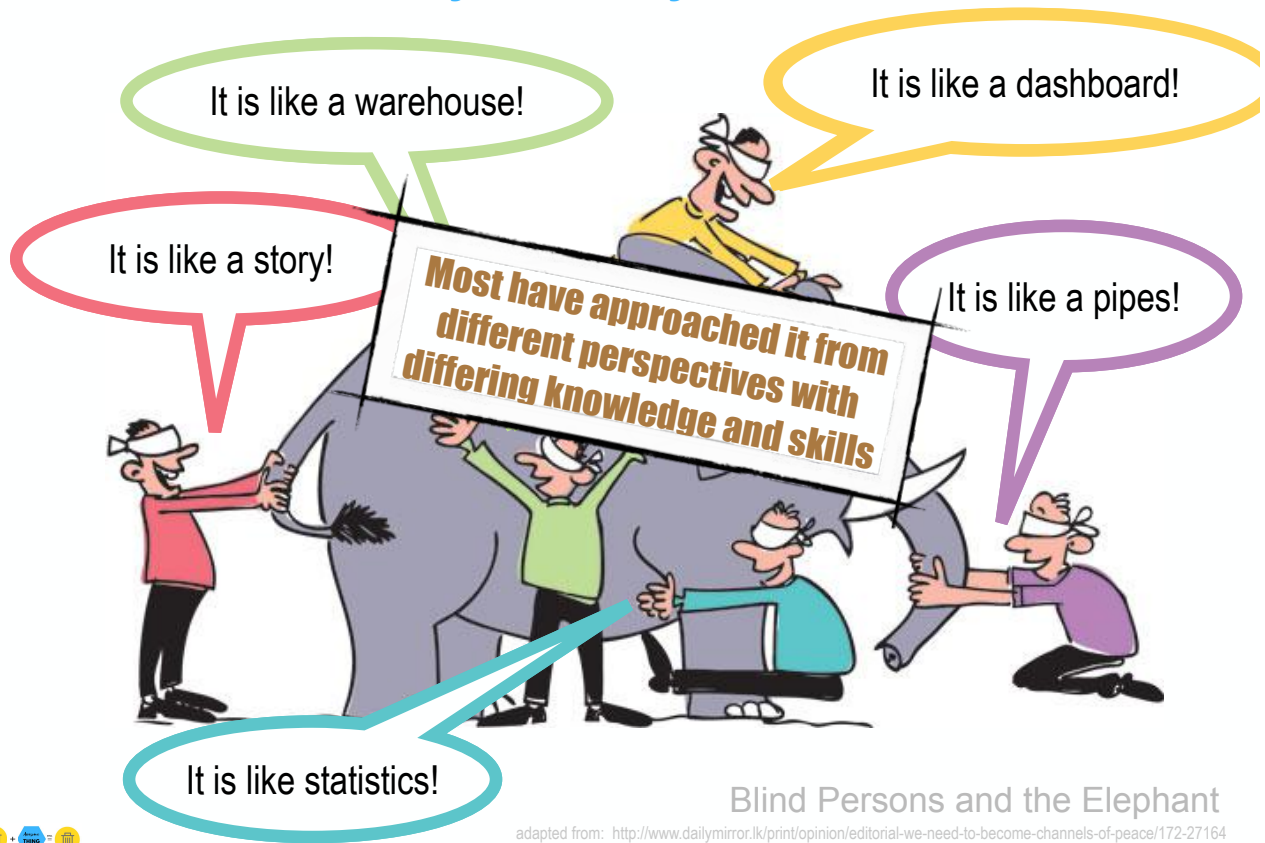


# DMM<sup>SM</sup> Structure of 5 Integrated DM Practice Areas

|   |           |
|---|-----------|
| 5 | Optimized |
| 4 | Measured  |
| 3 | Defined   |
| 2 | Managed   |
| 1 | Initial   |



# Data is not broadly or widely understood



Sources



Data Management

Unrefined  
data management  
definition

Uses



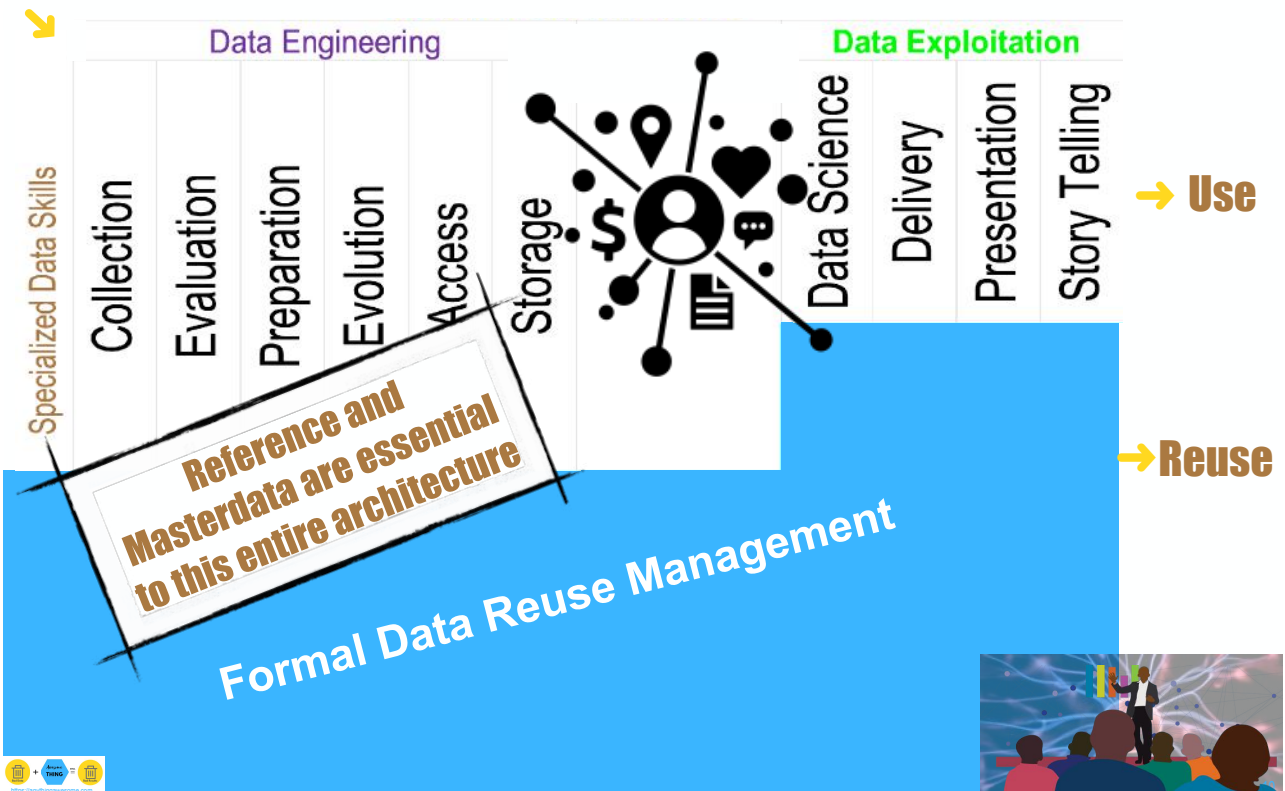
# More refined data management definition

Sources → Data Management → Reuse



# Better still data management definition

## Sources



## Program overview

- Data Management Overview

- What is Reference and MDM?

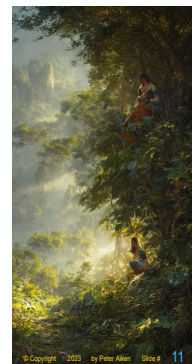
- Why is Reference and MDM important?

- Reference & MDM Building Blocks

- Guiding Principles & Best Practices

- Take Aways, References & Q&A

Essential:  
Reference &  
Master Data

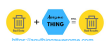


## Data Preparation Tools & Vendor Hype

- CIOs/CDOs feel pressure
- Vendor/project promise auditing
- No understanding of hype cycle



# HYPE



## Who wrote this ... ?

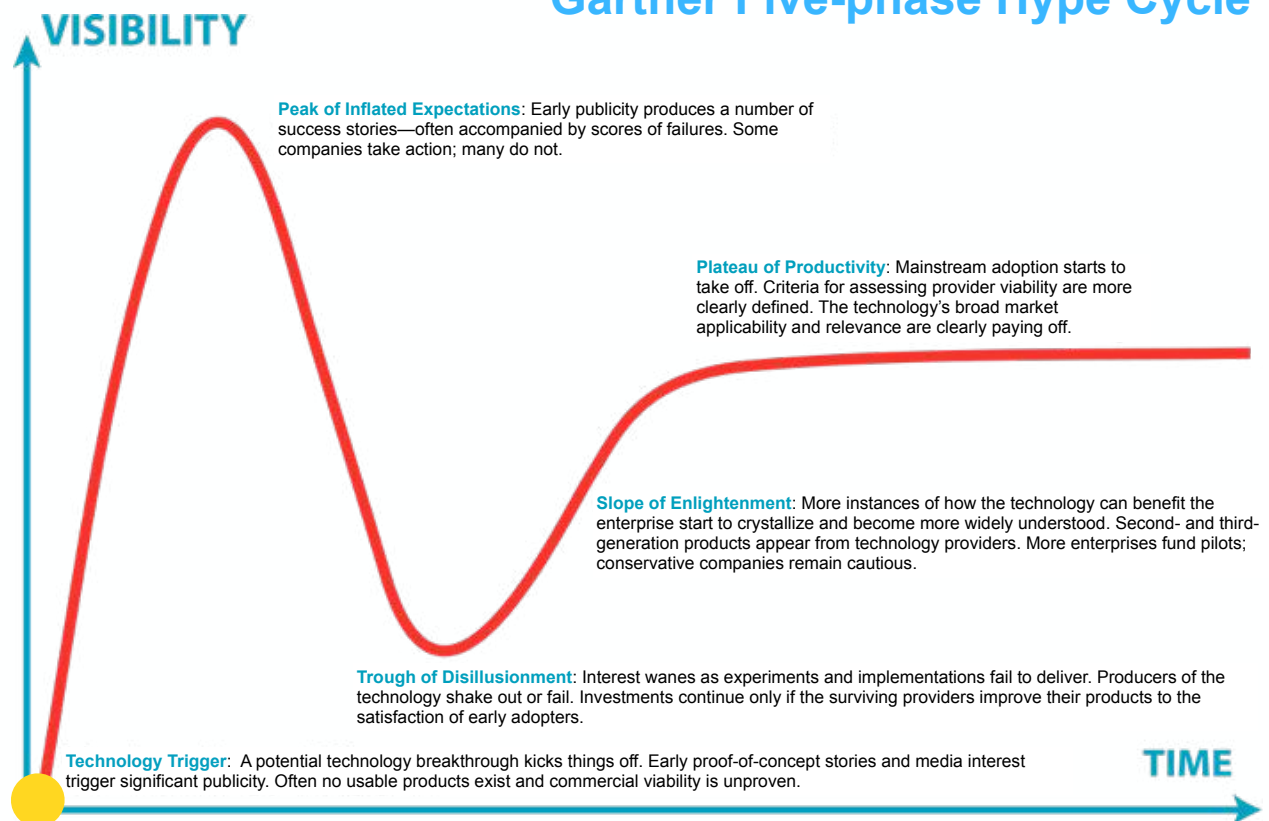
- *In considering any new subject,*
- *there is frequently a tendency first to overrate what we find to be already interesting or remarkable, and*
- *secondly - by a sort of natural reaction - to undervalue the true state of the case.*



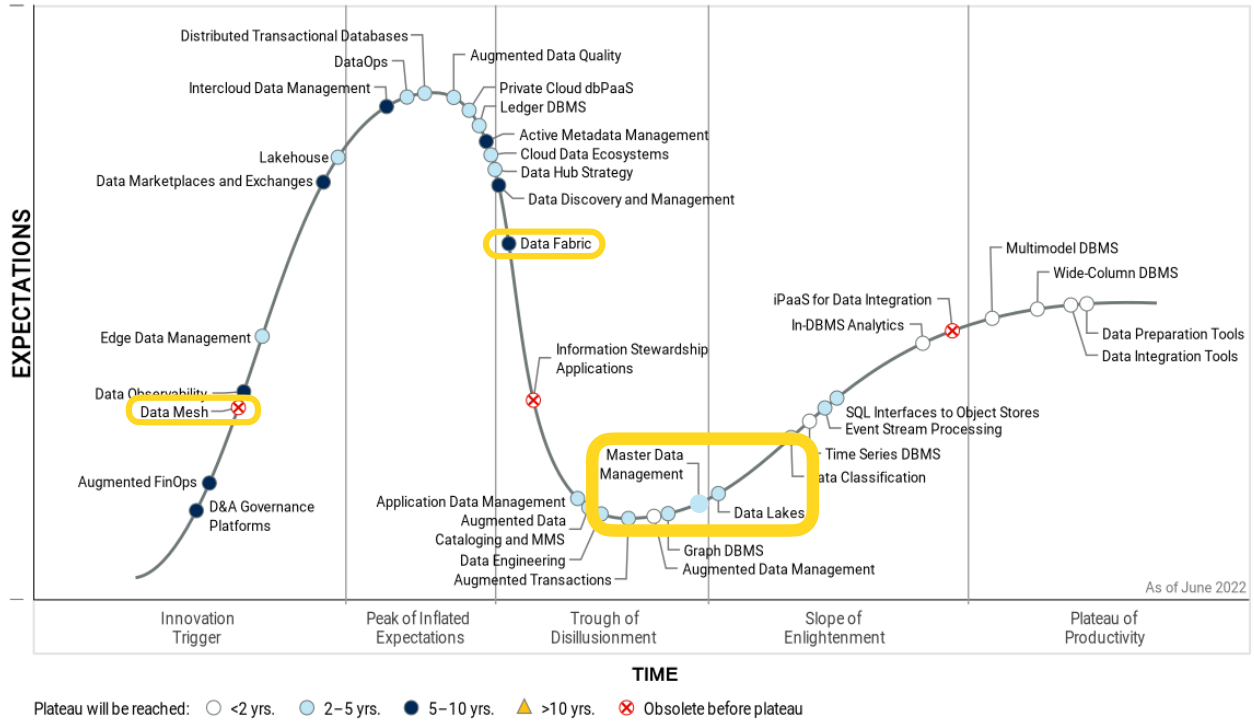
- Lady Augusta Ada King, (1815 – 1852) Countess of Lovelace
- (aka) Ada Lovelace, daughter of Lord Byron
- Publisher of the first computing program



## Gartner Five-phase Hype Cycle



## Hype Cycle for Data Management, 2022



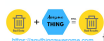
Gartner



## Data Management Body of Knowledge (DM BoK V2)



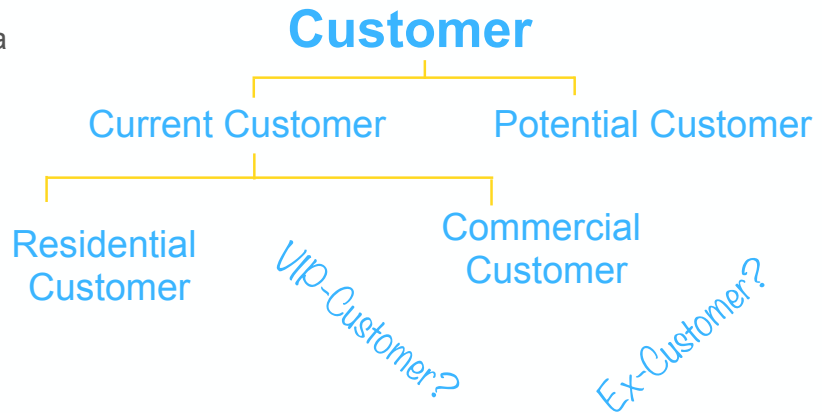
Practice Areas





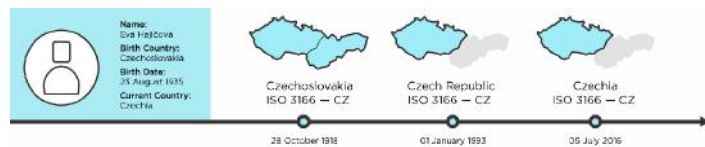
## Definition: Reference Data Management

- 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

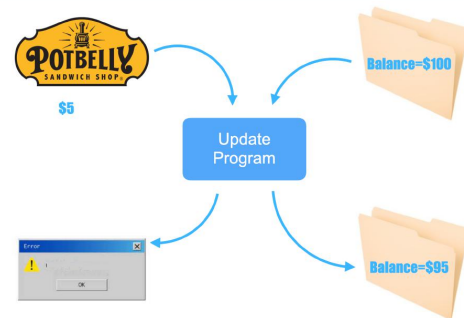
|             |                |
|-------------|----------------|
| US          | United States  |
| GB (not UK) | United Kingdom |



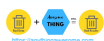
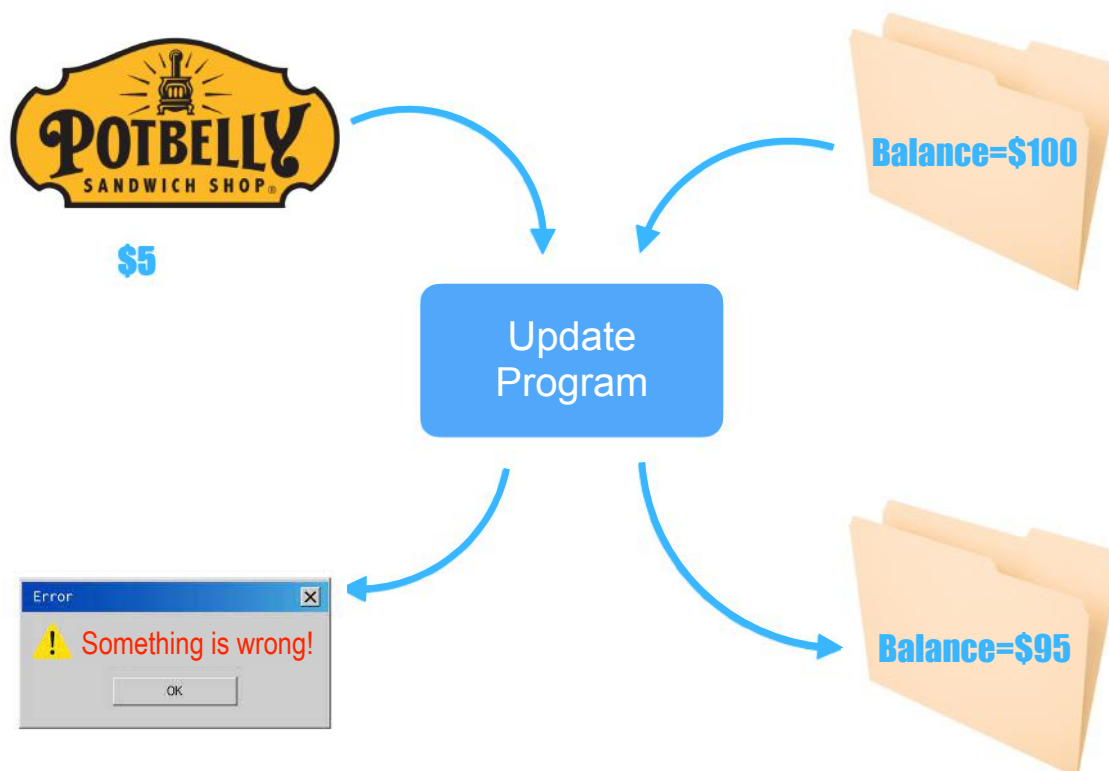
## Master Data



- 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"



## Example Transaction Processing System



# 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"

|   |                             |
|---|-----------------------------|
|  | 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 **mobile device management**
- 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)



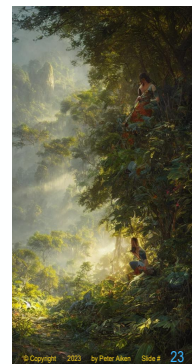
# Program overview

- Data Management Overview
- What is Reference and MDM?

Essential:  
Reference &  
Master Data

- Why is Reference and MDM important?

- Reference & MDM Building Blocks
- Guiding Principles & Best Practices
- Take Aways, References & Q&A



## 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](mailto:chris.bradley@dmadvisors.co.uk)



# What is Strategy?

strat·e·gy

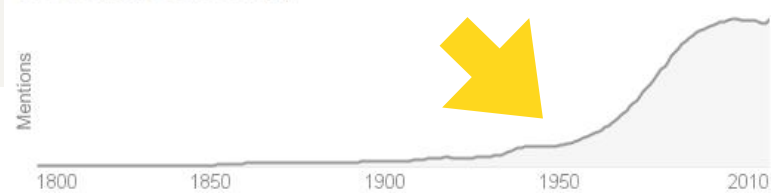
/ˈstrætəjē/

noun

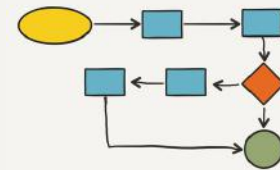
1. a plan of action or policy designed to achieve a major or overall aim.  
"time to develop a coherent economic strategy"  
synonyms: master plan, grand design, game plan, plan (of action), action plan, policy, program; More

A thing

Use over time for: Strategy



- Current use derived from military
  - a pattern in a stream of decisions [Henry Mintzberg]



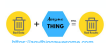
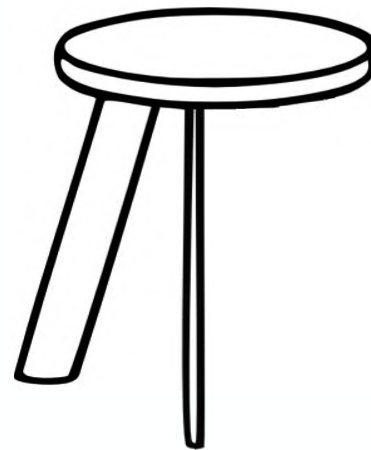
PROCESS



© Copyright 2023 by Peter Allen Slide 25

## + 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



© Copyright 2023 by Peter Allen Slide 26

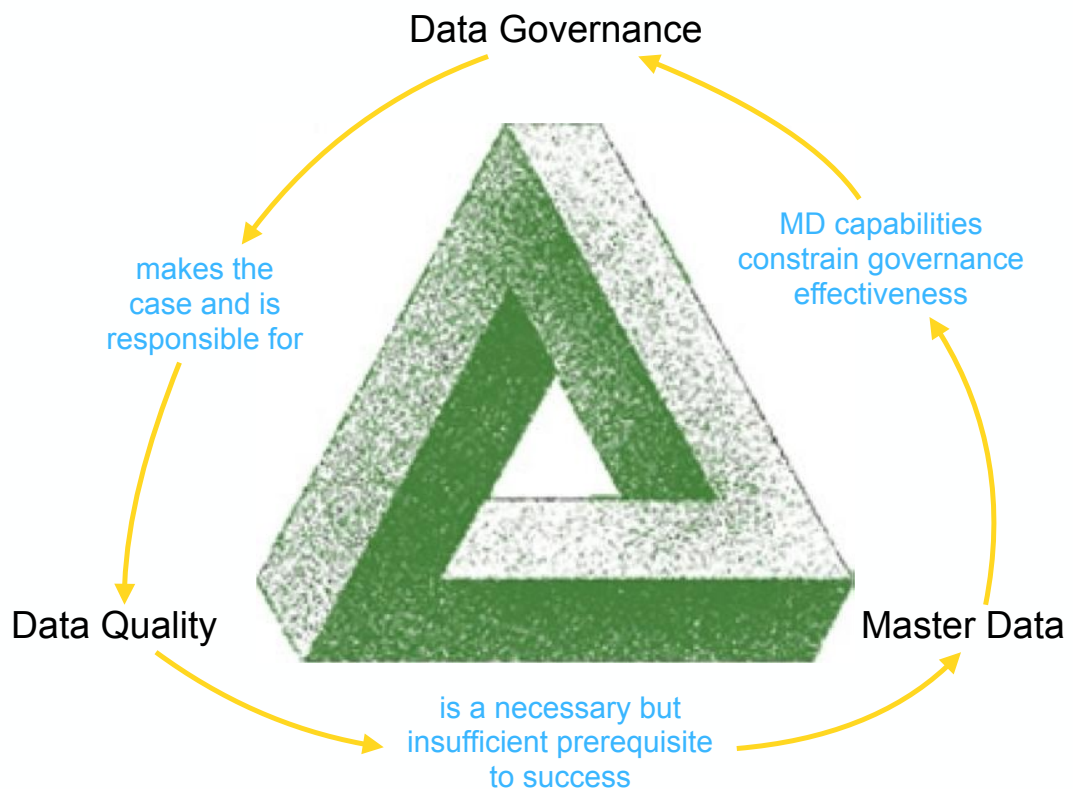
## My most profound lesson! (so far)



# Garbage In → Garbage Out!



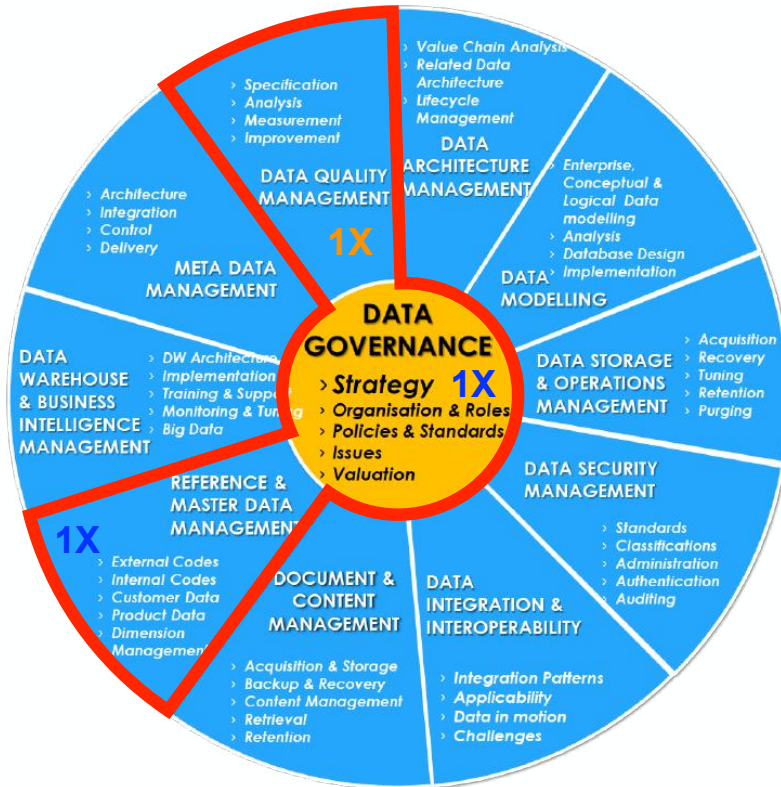
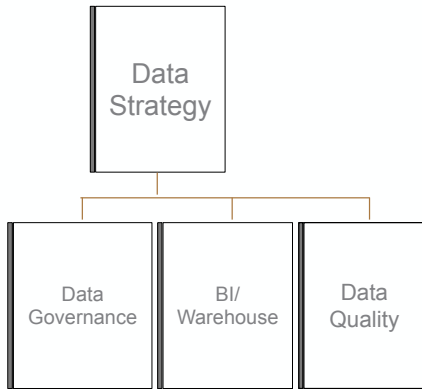
## Interdependencies



# Iteration 1



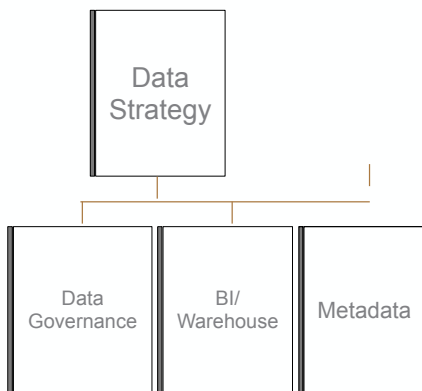
## Perfecting operations in 3 data management practice areas



# Iteration 2



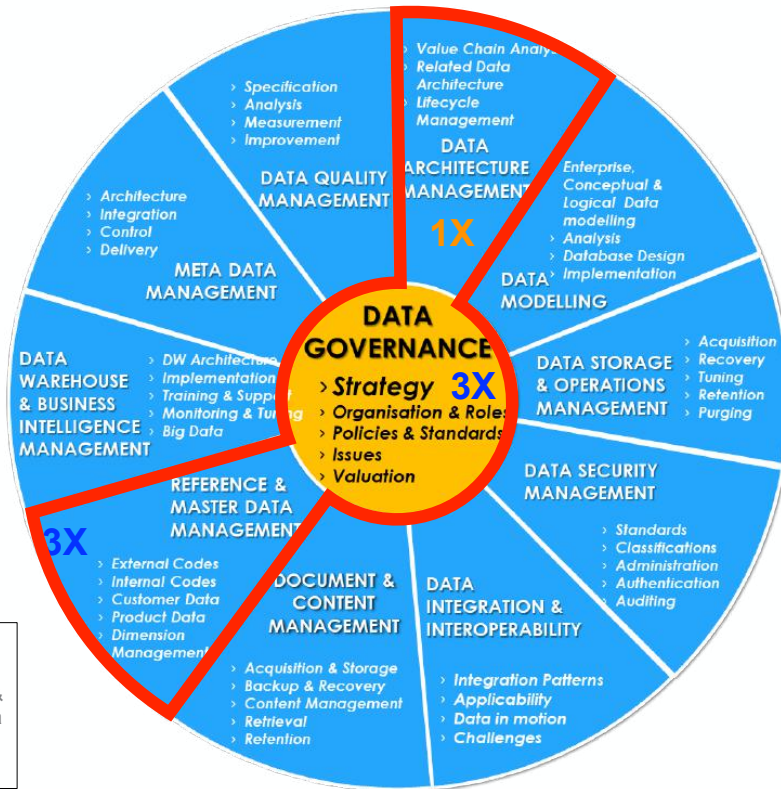
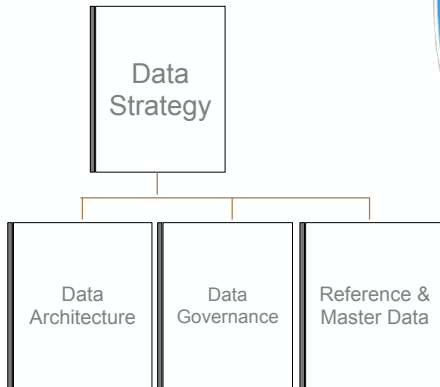
## Perfecting operations in 3 data management practice areas



# Iteration 3



## Perfecting operations in 3 data management practice areas

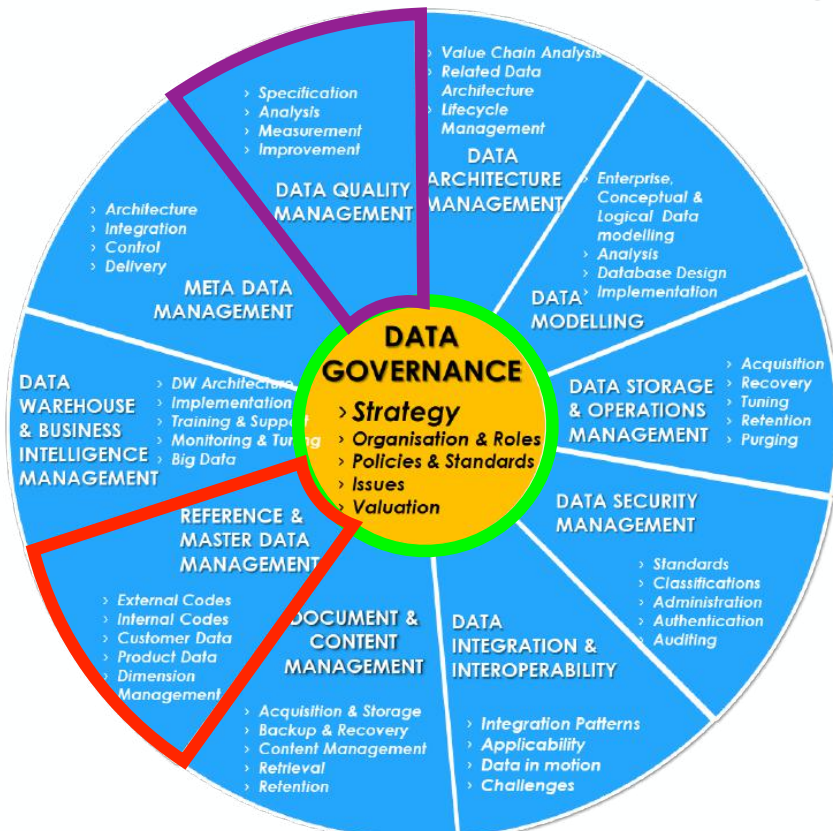


# A realistic way to begin practicing MDM



- Select 3 data management practice areas (for example)

- Reference and Master Data Management
- Data Quality Management
- Data Governance

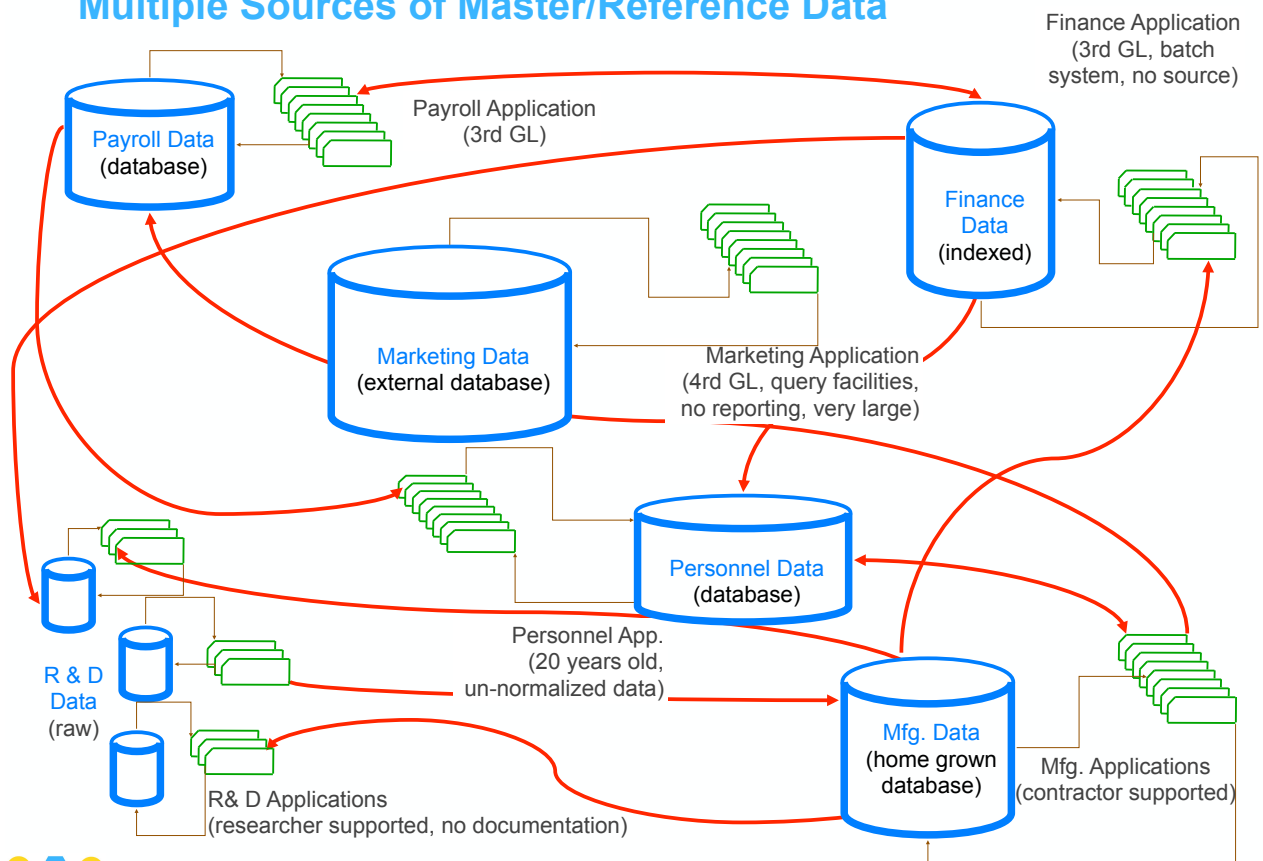




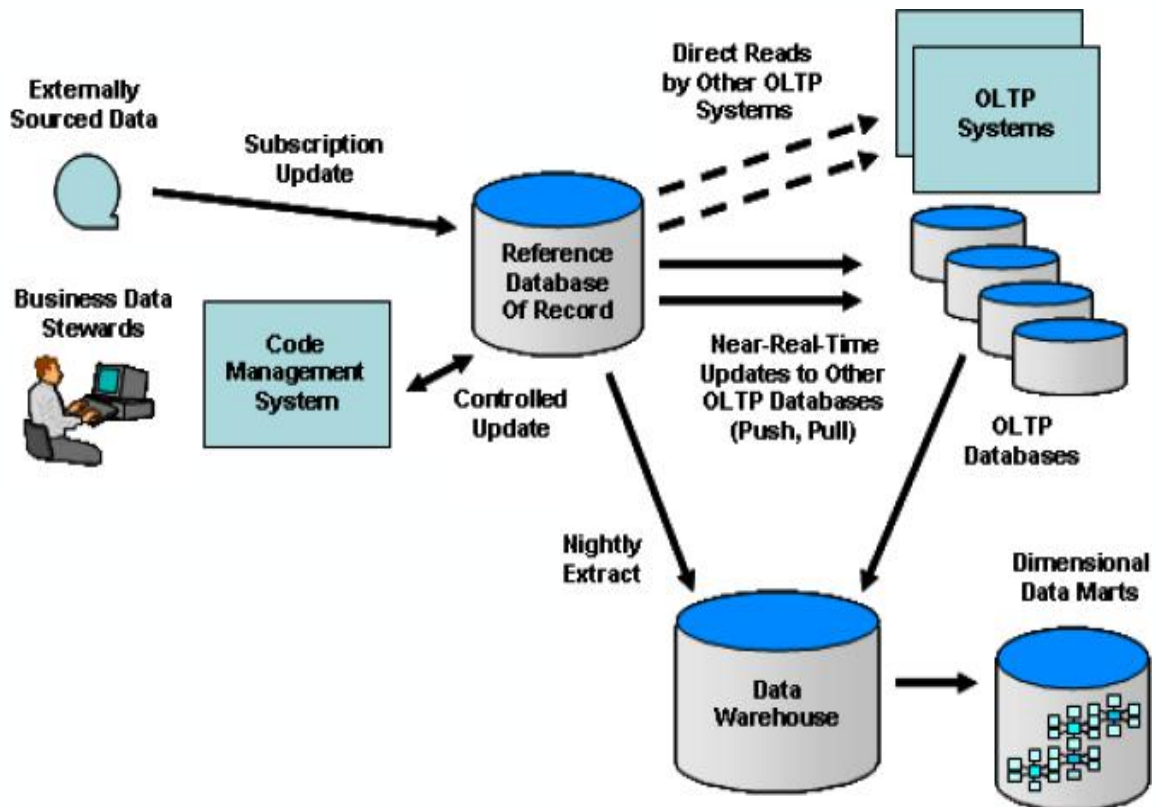
# Vocabulary is Important-Tank, Tanks, Tankers



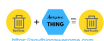
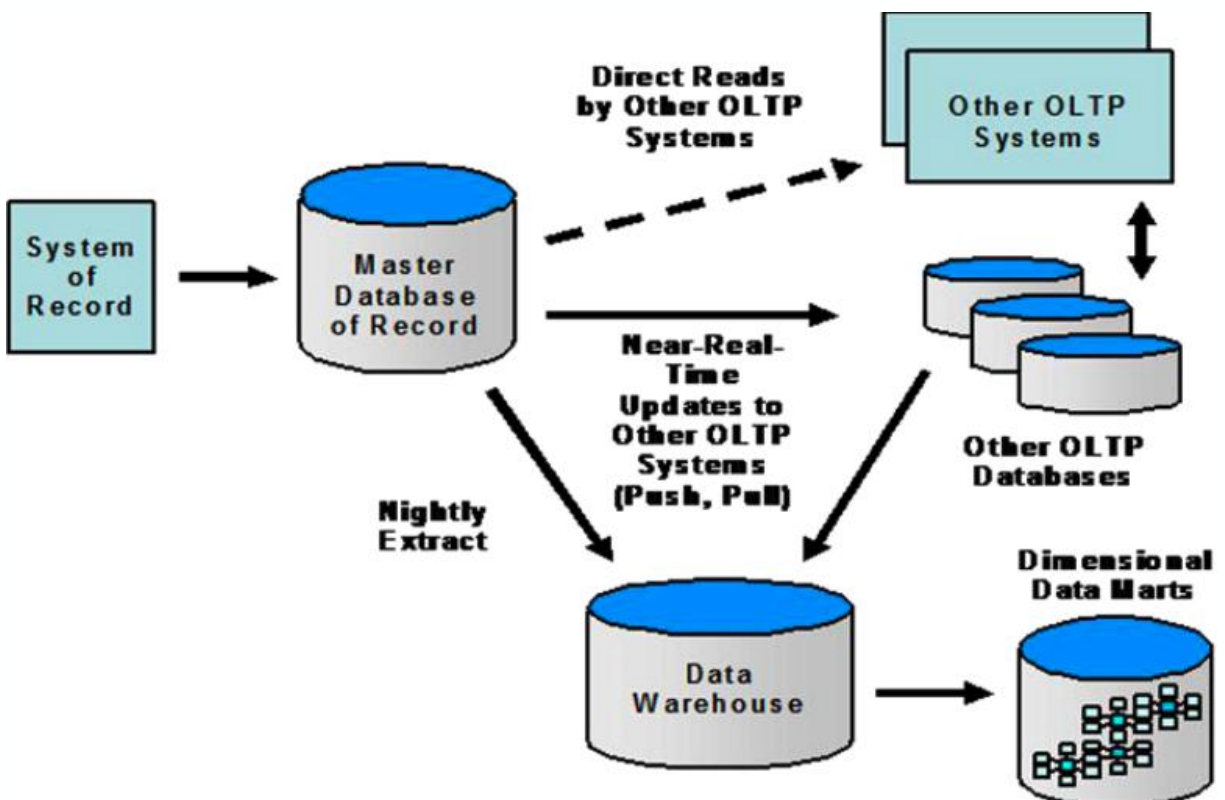
## Multiple Sources of Master/Reference Data



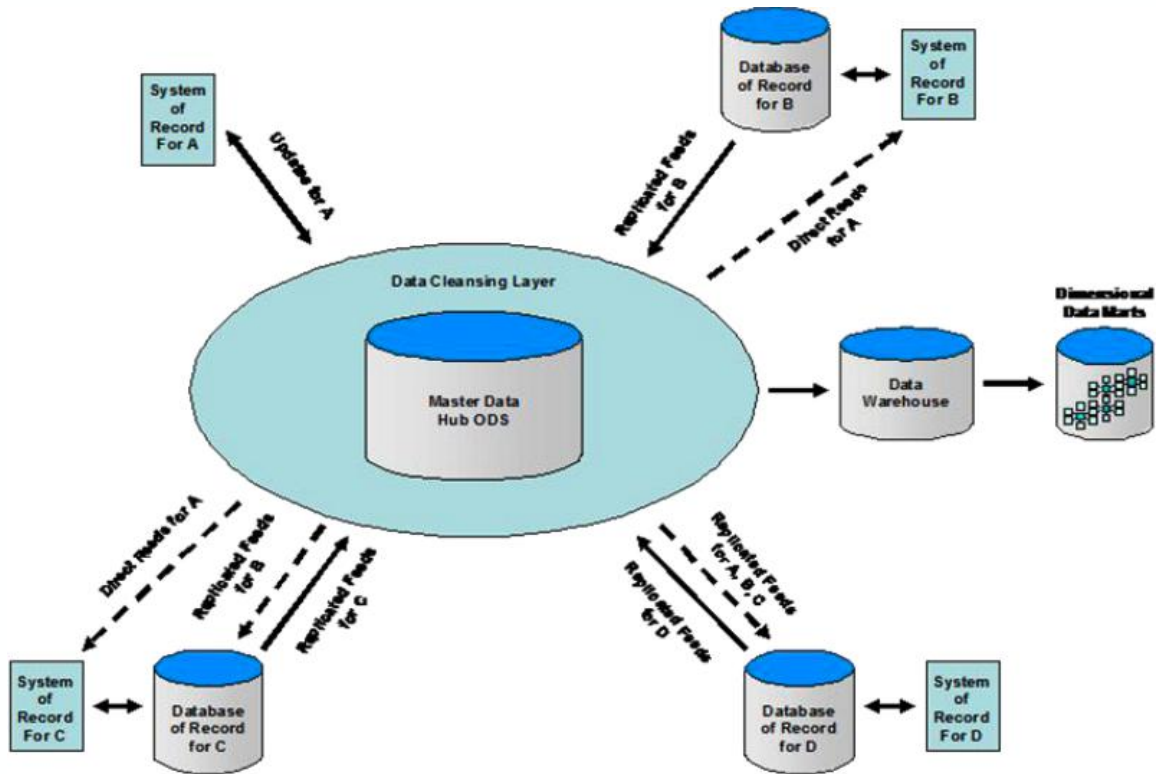
# Reference Data Architecture



# Master Data Architecture



# Combined R/M Data Architecture

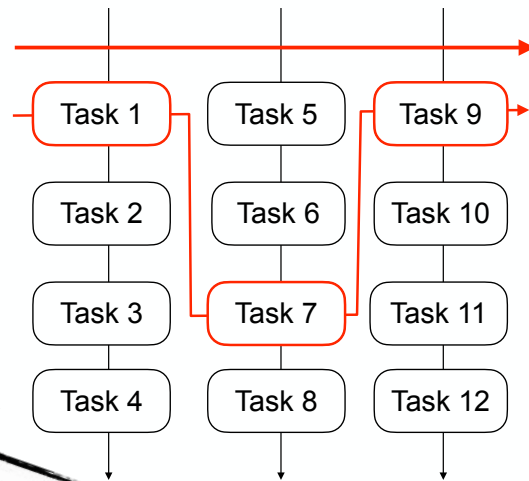


from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International

© Copyright 2023 by Peter Allen Slide 37

## 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

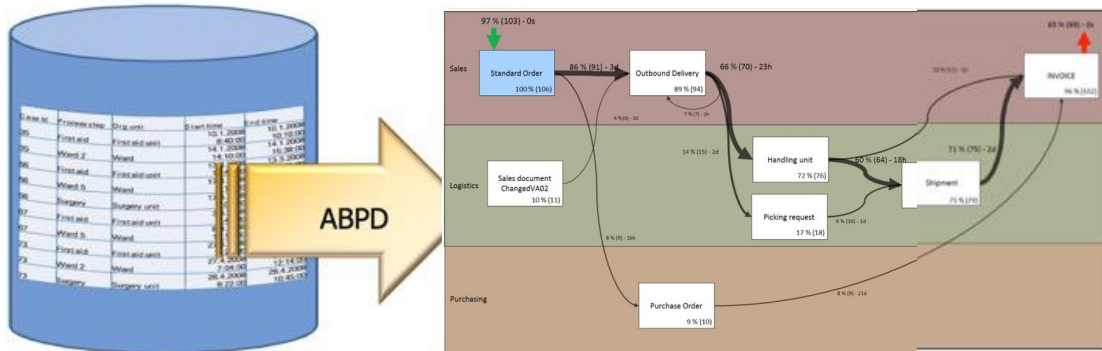


**Source System: all info on 1 screen**  
**Target system: same info spread across 23 screens**



© Copyright 2023 by Peter Allen Slide 38

# Automating Business Process Discovery (qpr.com)



## • Benefits

- Obtain holistic perspective on roles and value creation
- Customers understand and value outputs
- All develop better shared understanding

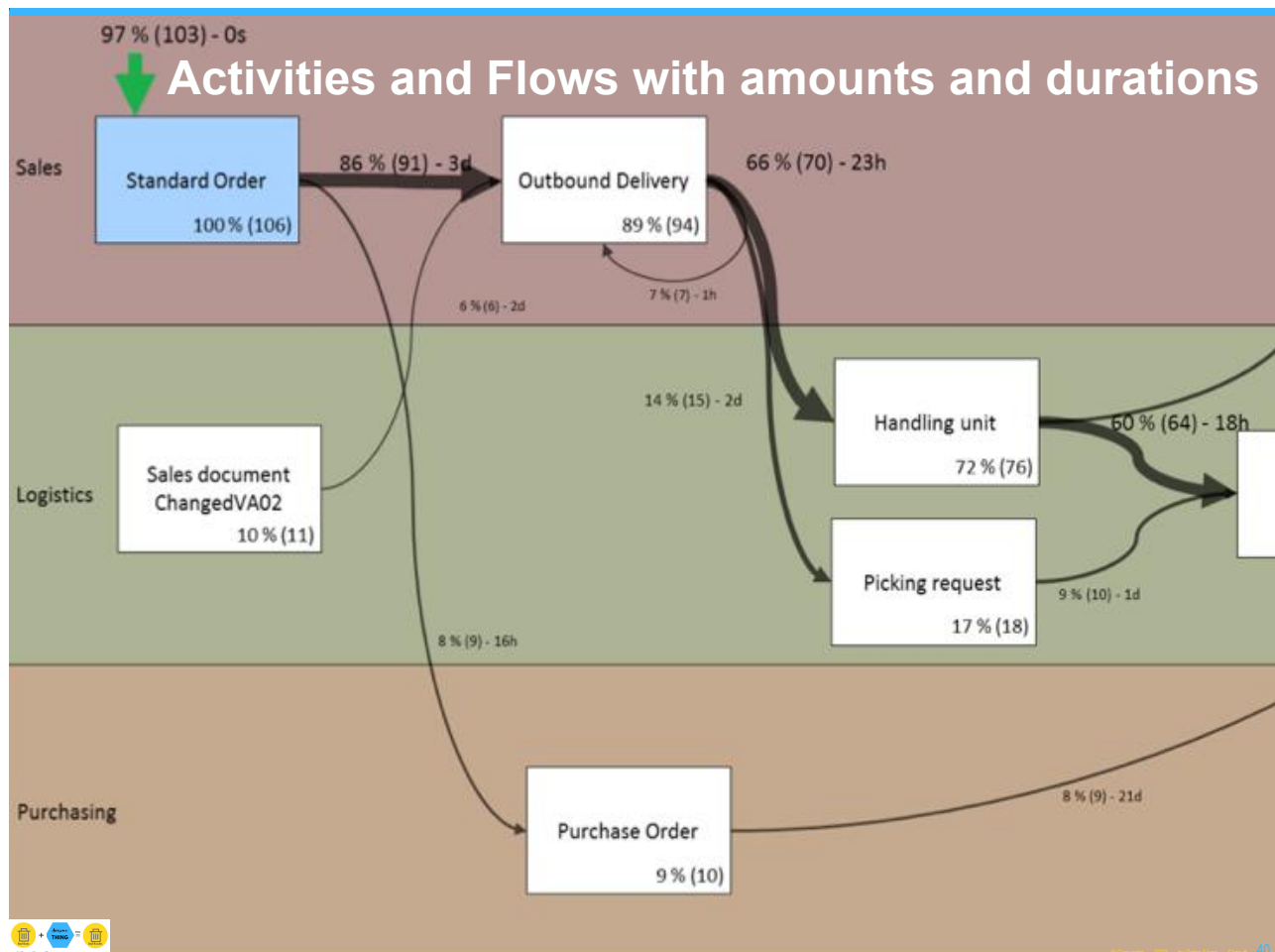
## • Results

- Speed up process
- Cost savings
- Increased compliance
- Increased output
- IT systems documentation



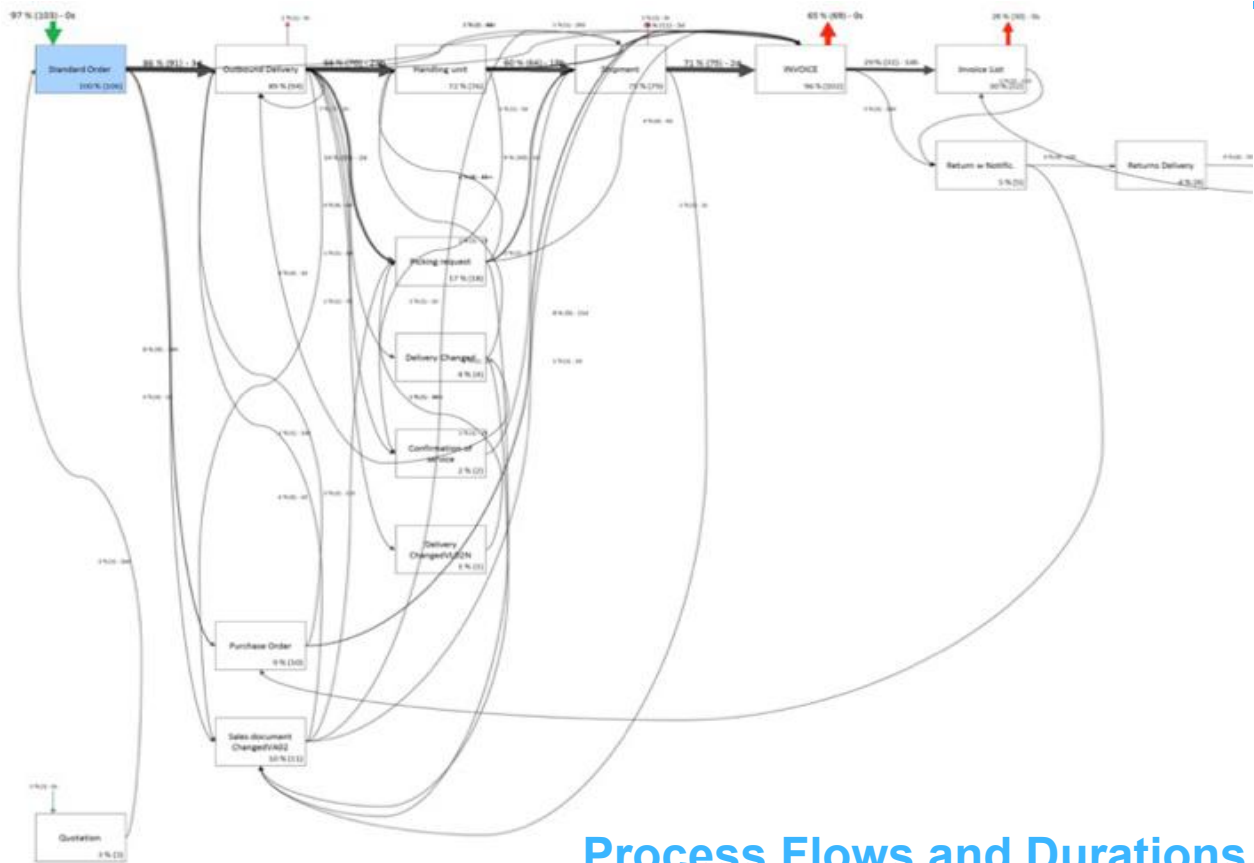
<https://anyting.com>

© Copyright 2023 by Peter Allen Slide 4 39



<https://anyting.com>

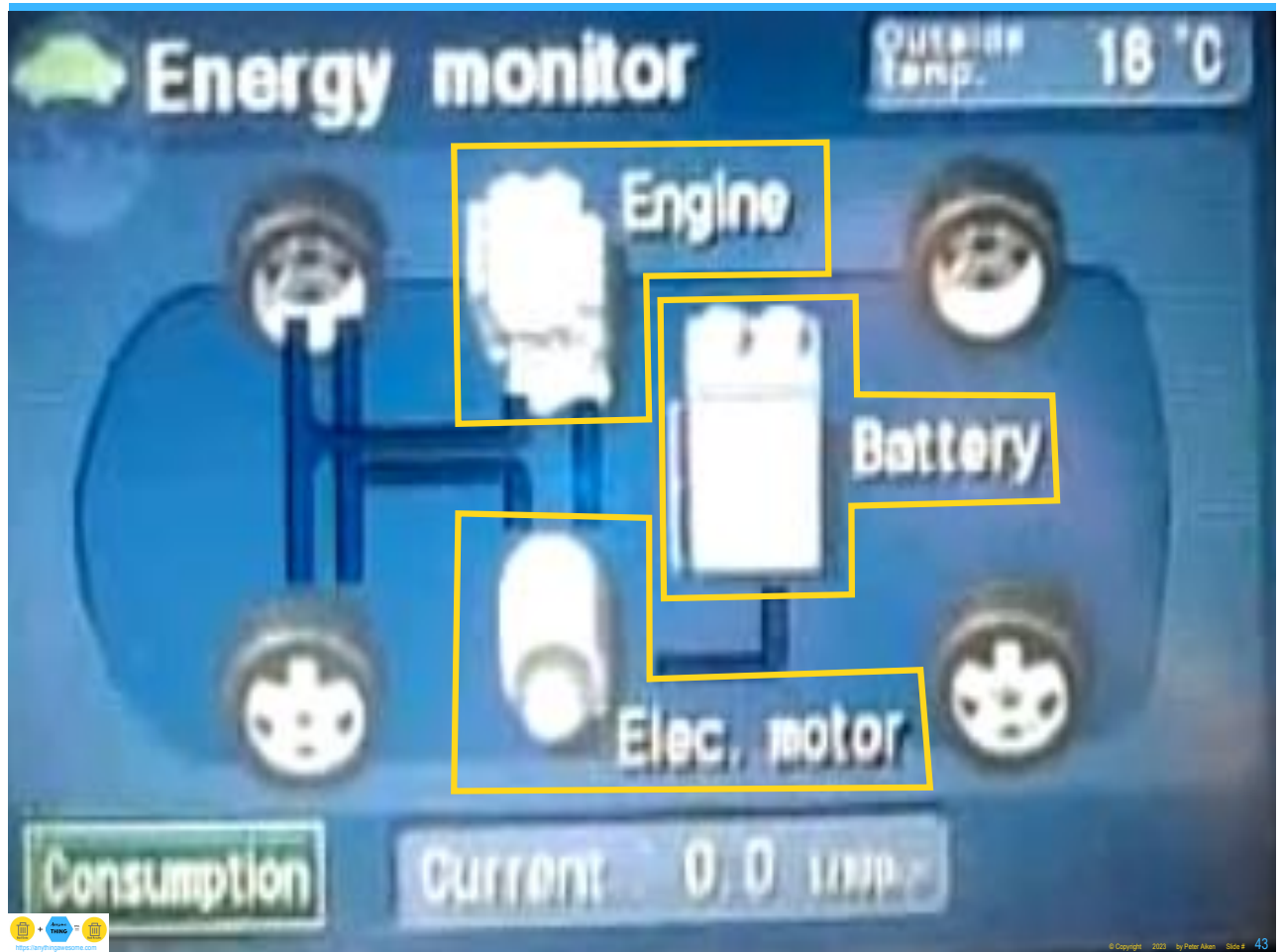
© Copyright 2023 by Peter Allen Slide 4 40



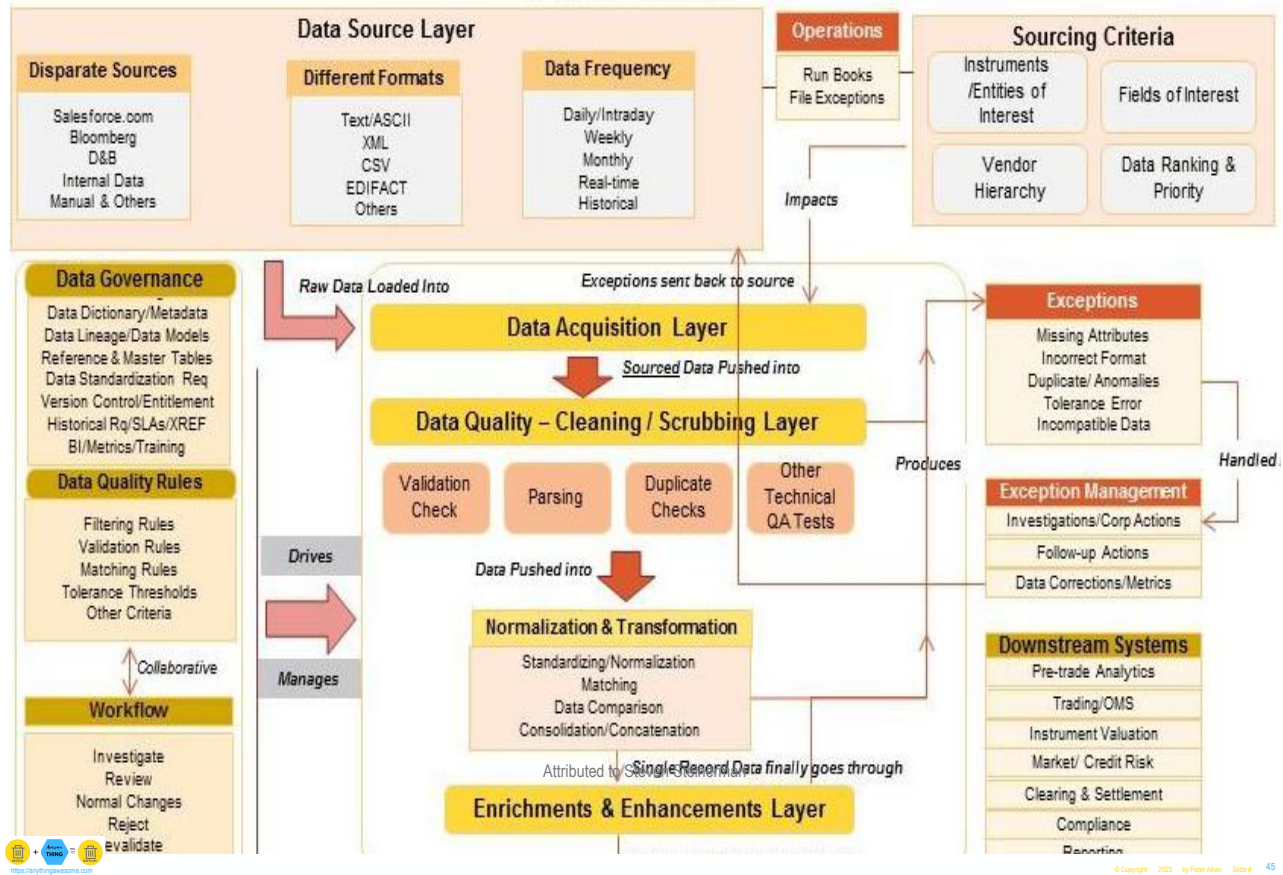
## Process Flows and Durations



## Traditional Engine



# Sample MDM Business Process Overview



## Program Overview

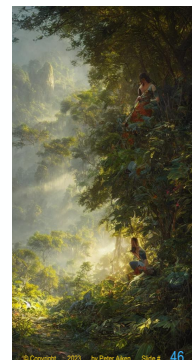
- Data Management Overview
- What is Reference and MDM?
- Why is Reference and MDM important?

Essential:  
Reference &  
Master Data

- Reference & MDM Building Blocks

- Guiding Principles & Best Practices

- Take Aways, References & Q&A



## 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

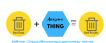
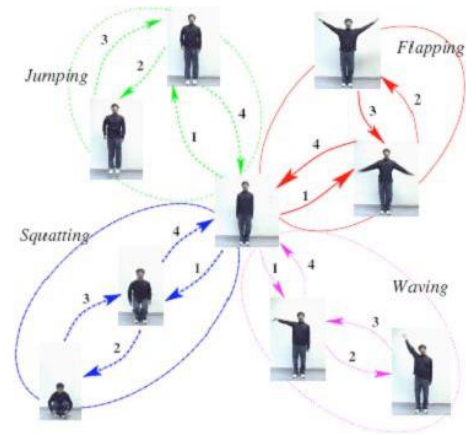


from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

© Copyright 2023 by Peter Allen Slide 47

## 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



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

© Copyright 2023 by Peter Allen Slide 48



---

## 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**



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

© Copyright 2023 by Peter Allen Slide 49

---

## 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

© Copyright 2023 by Peter Allen Slide 50

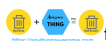
## Roles and Responsibilities

- 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
  - ...

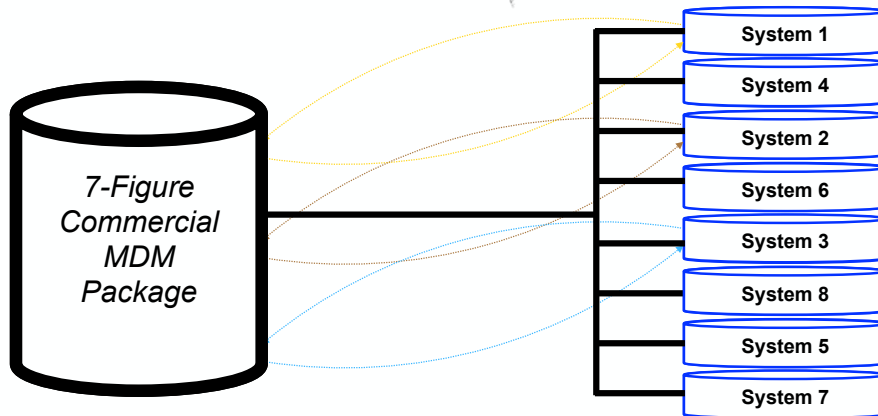


## Technology

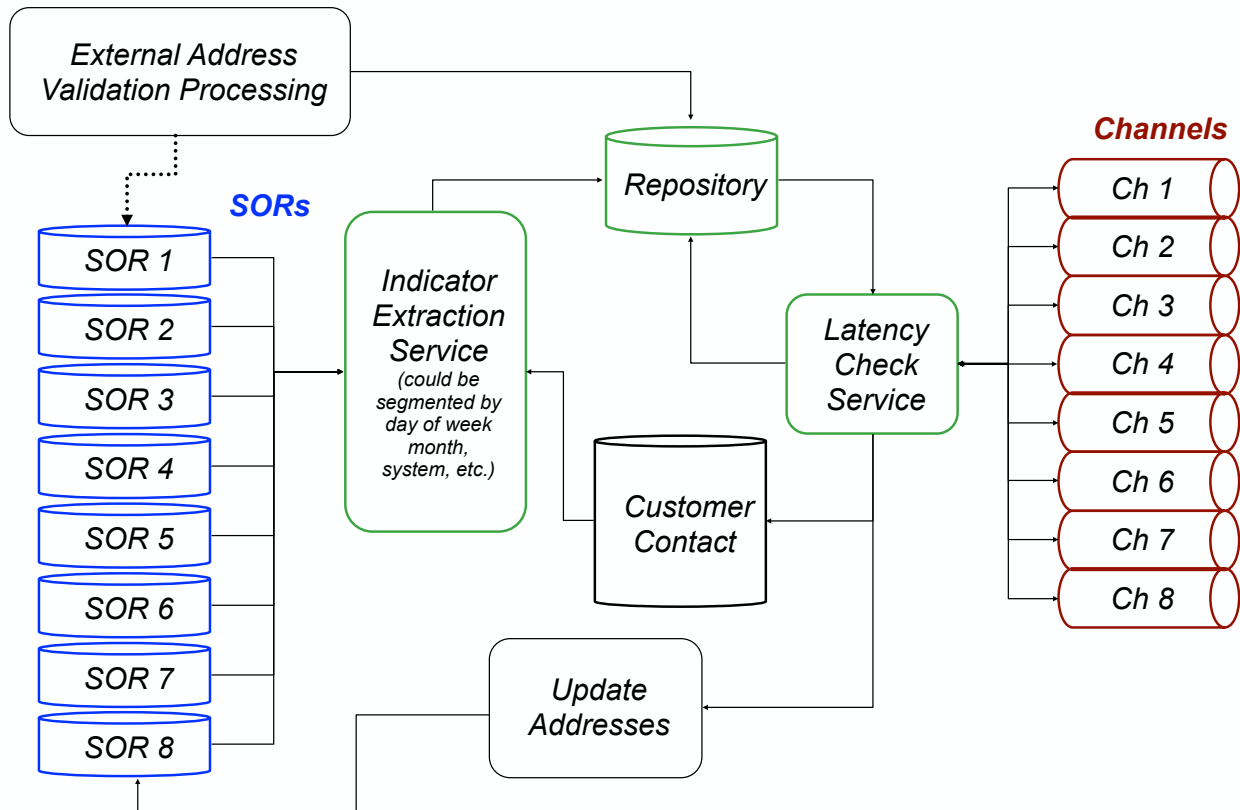
- ETL
- Reference Data Management Applications
- Master Data Management Applications
- Data Modeling Tools
- Process Modeling Tools
- Metadata Repositories
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Business Process and Rule Engines
- Change Management Tools



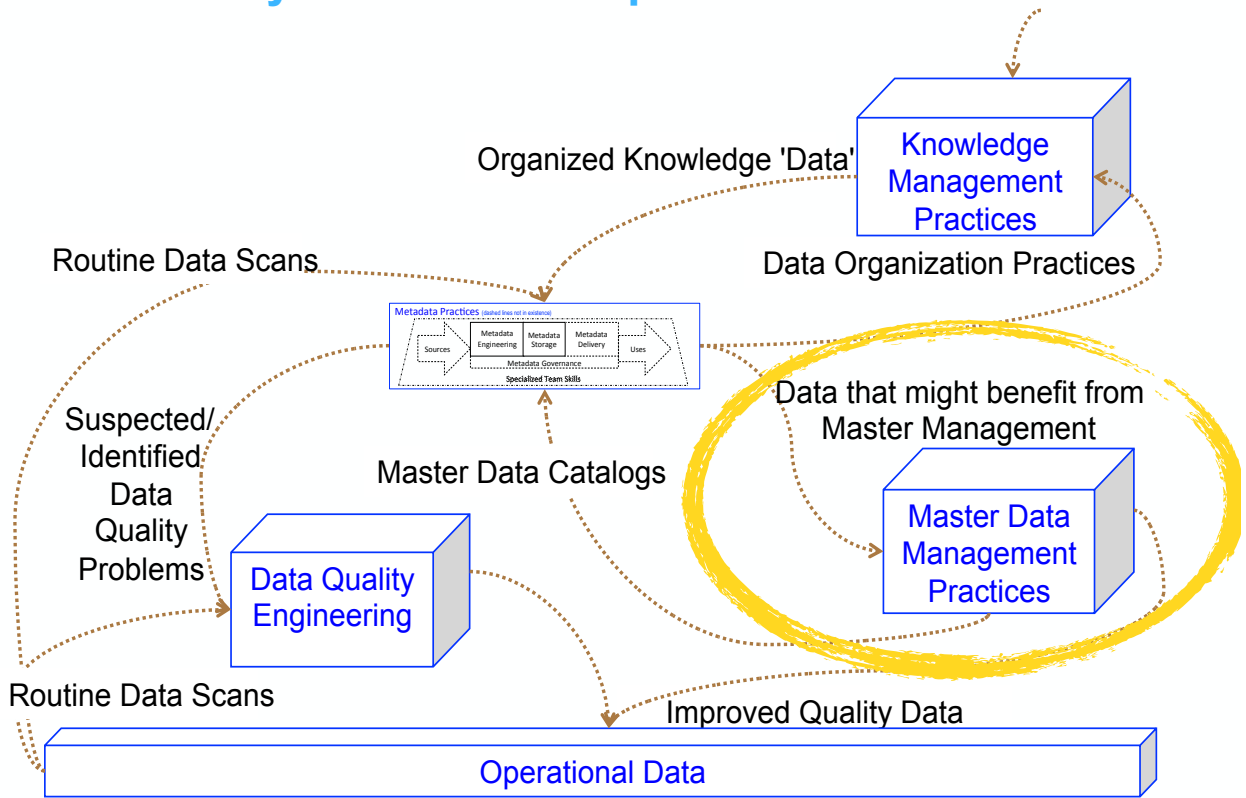
# Build Your First MDM



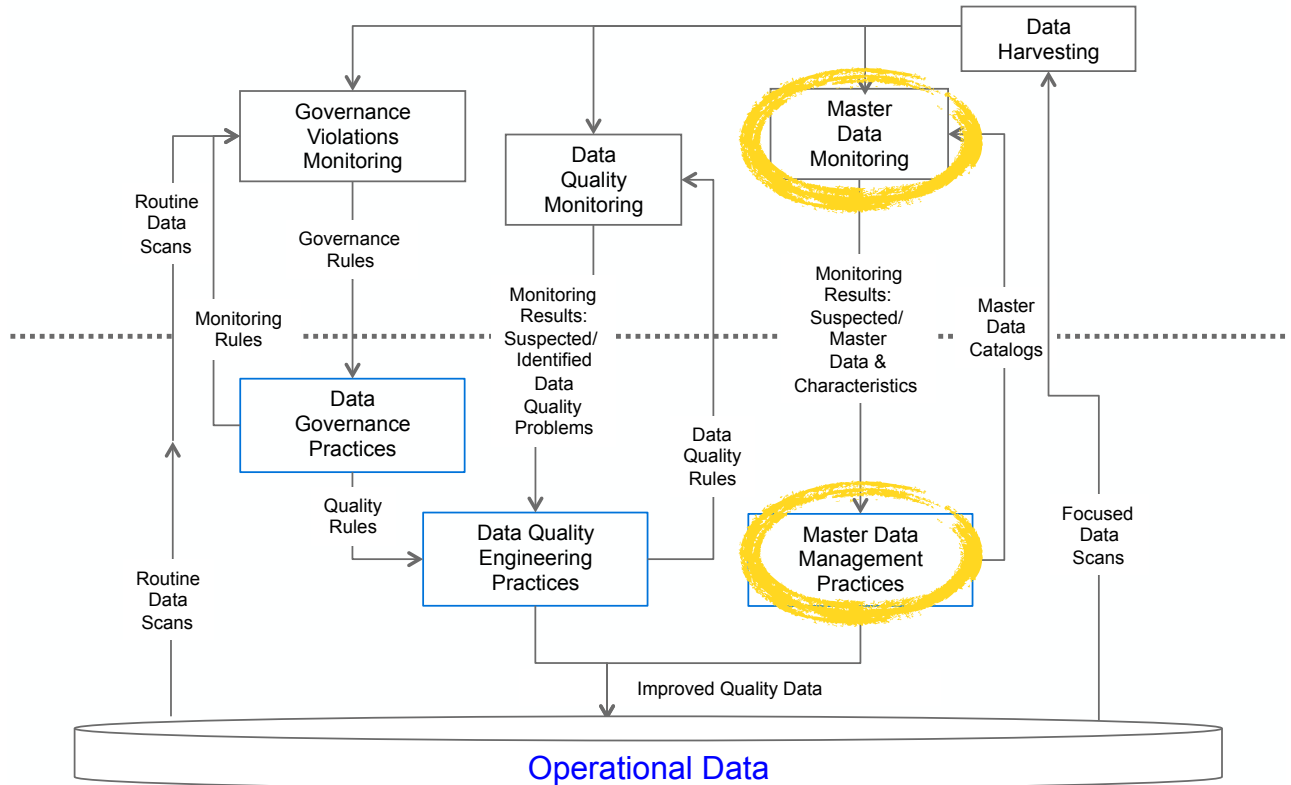
## Sample Solution Framework



# Inextricably intertwined implementations and ...



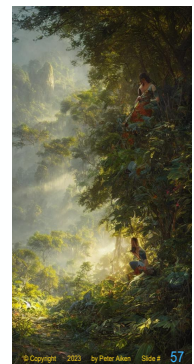
# Interactions



# 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



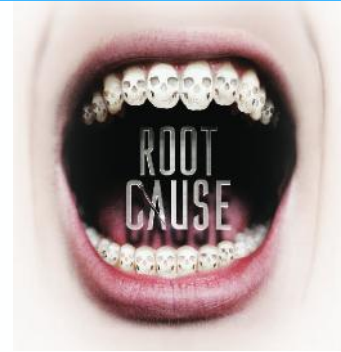
## "180% Failure Rate" Fred Cohen, Patni

- 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.



## 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.



<https://www.ase.org.uk/bestpractice>



Source: <http://www.mdmsource.com/master-data-management-tips-best-practices.html>

© Copyright 2023 by Peter Allen Slide # 61

# 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

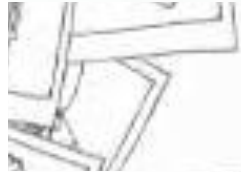
© Copyright 2023 by Peter Allen Slide # 62



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



<https://anythingawesome.com/sevensistersvideo.html>

Thanks to Dave Evans

© Copyright 2023 by Peter Allen Slide 6 63

**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

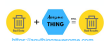
### Consumers:

- 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

© Copyright 2023 by Peter Allen Slide 6 64



---

## 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](http://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](http://www.dataqualityfirst.com/getting_master_data_facts_straight_is_hard.htm)



## Upcoming Events

### Gaining Business & Executive Support for Data Management

9 May 2023



### Key Elements of a Successful Data Governance Program

13 June 2023

**Time: 19:00 UTC (2:00 PM NYC) | Presented by: Peter Aiken, PhD**

### Conceptual vs. Logical vs. Physical Data Modeling

11 July 2023

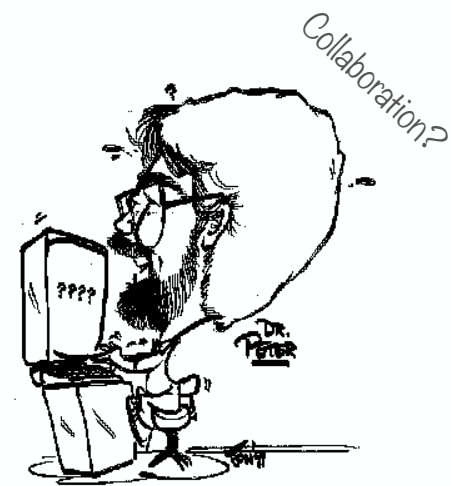
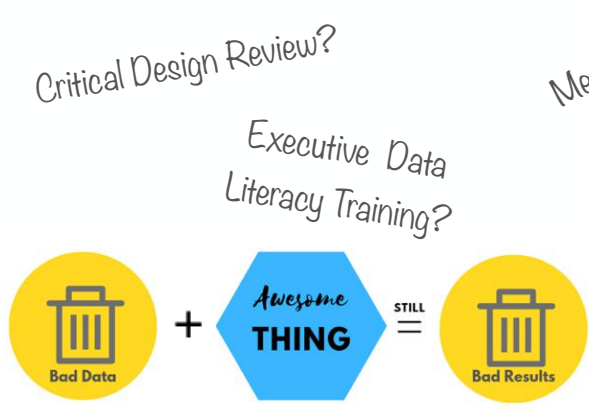
Brought to you by:



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

© Copyright 2023 by Peter Aiken Slide 67

Independent Verification & Validation



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>

